LEANDRO FREDERICO FERRAZ MEYER

THE COMMONS DILEMMA REVISITED: TOWARD AN INTEGRAL APPROACH

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia Aplicada, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2006

Ficha catalográfica preparada pela Seção de Catalogação e Classificação da Biblioteca Central da UFV

T

M612c
2008

Meyer, Leandro Frederico Ferraz, 1968-
The commons dilemma revisited: toward na integral approach / Leandro Frederico Ferraz Meyer. – Viçosa, MG, 2008.
xix, 533f.: il. (algumas col.) ; 29cm.

Inclui apêndice.
Orientador: Marcelo José Braga.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 483-519.

1. Graves, Clara. 2. Existencialismo. 3. Valores. 4. Sistema intrapessoal. 5. Psicologia existencial. 6. Comportamento humano. I. Universidade Federal de Viçosa. II.Título.

CDD 22.ed. 142.78

LEANDRO FREDERICO FERRAZ MEYER

THE COMMONS DILEMMA REVISITED:
AN APPLICATION OF THE INTEGRAL APPROACH

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia Aplicada, para obtenção do título de *Doctor Scientiae*.

APROVADA: 21 de dezembro de 2006.

Arnoldo José de Hoyos Guevara

Newton Paulo Bueno

João Eustáquio de Lima

Erly Cardoso Teixeira
(Co-orientador)

Marcelo José Braga
(Orientador)

ACKNOWLEDGEMENTS

Works of this nature are always the result of a collective effort. I have many people to thank for their guidance and support along the journey, plenty of twists and turns, that took me to the thesis I present here. Marcelo José Braga, my supervisor, provided the vision that guided this research towards its main theme, namely: the "puzzle of cooperation". It was Marcelo's idea to inquiry on agents' different dispositions to cooperate. Starting from quite different perspectives on this issue, Professor Marcelo gave me the freedom to pursue my own reading and research path; one that eventually took us to the Integral approach, as a way to conciliate them.

I am also in debt and sincerely grateful to Professor Marcelo Lobato, from the Department of Physics (UFV), who received and provided invaluable assistance in developing my first attempt to research the commons dilemma using agent-based modeling; later deferred in favor of the experimental method.

I am thankful to Professor Elinor Ostrom, with whom I first made contact at Indiana University, in Bloomington, and who opened the doors of the Center for the Study of Institutions, Population, and Environmental Change (CIPEC) and the Anthropological Center for Training and Research on Global Environmental Change (ACT). It was during my staying at the ACT as a visiting doctoral student that I conceived how the constructivist developmental paradigm

could fit into the Institutional Analysis and Development (IAD) framework to animate institutional analysis and explain the diversity of the individuals' responses to similar incentive structures in collective-action settings. Thanks also to Professors Eduardo Brondízio, my supervisor during my presence at the ACT, and Rick Wilk, who introduced me to the field of Anthropological Economics and postmodern deconstructionism.

The "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) and the Federal Rural University of Amazon (UFRA) supported my work financially. I am also grateful to Chris Cowan and Natasha Todorovic, for their tuition and substantial price cut that allowed me to obtain the needed qualification to work with Clare Graves's theory and to develop the Portuguese version of the assessment tool without which this research would not be possible.

I cannot pass up my debt with Mestre Irineu Serra and Padrinho Sebastião, together with all the continuators of the Santo Daime Doctrine, for the utterly precious opportunity to open the path of my spiritual awakening and character transformation, along which I now persist under the guidance, protection, and continuous encouragement of my Master, Li Hongzhi.

I am especially thankful to Ken Wilber's extraordinary intellectual endeavor (Integral Theory) and his insightful recognition of the significance of meditative practices, which holds up the arduous task of reframing my mind and body to unlock the fruition of the radiant spirit. Thank you very much to the Integral Institute for granting a full scholarship and to the Integral Sustainability Seminar people for their enthusiasm and commitment in offering us all their precious experience and qualified training.

My sincere gratitude to all those mentioned above, Professors Arnoldo José de Hoyos Guevara, Newton Paulo Bueno, João Eustáquio de Lima, and Erly Cardoso Teixeira, members of the dissertation committee, and to the so many other indirect contributors referred to along the text, who compound with me, and you, this circle of togetherness, as Wilber puts it, "without which the heart of the *Kosmos* could never beat a single beat, and would have no reason to do so, even if it could".

BIOGRAPHY

Born on June the first, 1968, in São Paulo, Brazil, the author obtained his B.A in Animal Production Science and Technology (Zootecnia), at the São Paulo State University Júlio of Mesquita Filho (1992), and Master's degree in Rural Economy in the Federal University of Viçosa (1997). Out from author's concern with the development of sustainable production systems and technologies, his research interests gradually migrated from the study of agrarian policy and agricultural modernization in Brazil and later toward a more fundamental inquiry on the origin of agents' preferences and intentional states, in connection with the development of new models of the human actor in economic theory. Along this pathway, the author followed the thread of social and developmental psychology until he came across with Ken Wilber's Integral Psychology and Clare Graves's Emergent-Cyclical Levels of Existence Theory of adult *biopsychosocial* development, which together provide the distinctive framework with which the author approaches the broader puzzle of sustainable development under the leitmotiv theme and focuses on collective action dilemmas. The author is currently Professor of Rural Administration and Agricultural Development in the Federal Rural University of the Amazon (UFRA), in Belém-PA, Brazil.

CONTENT

Page

RESUMO xiii

ABSTRACT xvii

1. REFRAMING THE INQUIRY INTO COLLECTIVE ACTION: TOWARD AN INTEGRAL APPROACH …….. 1

1.1. Introduction …… 1

1.2. Overview of Graves's ECLET, working hypothesis, and the thesis of the study ……. 12

1.3. Collective action as moral dilemma: beyond moral noncognitivism …….. 20

1.4. Research strategy and the Integral Methodological Pluralism (IMP) ……. 28

1.5. Organization and limitations of the study …….. 31

2. THE INTEGRAL VIEW OF KEN WILBER AND THE INTEGRAL OPERATING SYSTEM …….. 34

2.1. Introduction …… 34

Page

2.2. The "collapse of the Kosmos" and the "paradox of damage": the epistemological puzzle under the present ecological crisis and the emergence of integral theory ……... 37

2.3. The pattern that connects and the validity claim of mysticism: toward the reenchantment of the world ……. 50

2.4. The IOS and the IMP: upgrading our minds for integrally processing the Kosmos …….. 73

2.5. The implications of the integral approach for the investigation of the commons dilemma ……... 86

3. UNCOVERING PARTIALNESS ……. 92

3.1. Introduction …… 92

3.2. An integral map to come across partialness ……... 94

3.3. The neoclassical economic theory of collective action ……..... 97

3.4. The formal model of the appropriation dilemma: an illustration ... 102

3.5. Rethinking the model of individual in economics: relaxing assumptions and approaching complexity ……. 107

3.6. A half step toward freedom: the emancipatory and the automating faces of the IAD framework ……... 114

3.7. A correction of perspective on the structure of the action arena ... 120

3.8. The general schema of the IAD framework ……... 123

3.9. Administrating nihilism beyond panaceas: a growing consensus toward eight design principles of successful institutions ….... 126

3.10. The outer face of cooperation: the IAD framework's move toward a behavioral theory linking trust, reciprocity, and reputation ……. 133

3.11. On the partialness of the evolutionary approach to cooperation . 145

Page

3.12. On the difficulties of the immediate apprehension of reality, the elusiveness of its communication, and the validity claim of mysticism …… 163

4. THE DEVELOPMENTAL STRUCTURE OF VALUE SYSTEMS AND WORLDVIEWS: CLARE GRAVES AND THE SPIRAL DYNAMICS ……. 179

4.1. Entering ZONE # 2: the meaning of structure and the importance of post/structuralism …….. 180

4.2. Basic steps in the paradigm of adequate structuralism …….. 184

4.3. The phases of Graves's research throughout the steps of adequate structuralism …….. 188

4.4. The emergent-cyclical, double-helix, open-ended and two-tiered aspects of Clare Graves's model …….. 194

4.5. The nested or holarchical aspect of Clare Graves's model …… 200

4.6. The wave-like or multilevel aspect of Graves's model and the change dynamics …….. 202

4.7. The waves of agency and communion in the spiral dynamics: characterizing Graves's constructs …… 208

5. INTEGRATING PARTIALNESS …… 216

5.1. Introduction …… 216

5.2. The integral view of the human holon ……... 218

5.3. The pattern with connects and the sociocultural evolution …… 221

5.4. The many streams of a same river: limitations and possibilities of a generalization based on the ECLET conception ……. 227

5.5. Decentering of the self, moral skepticism, and the farther reaches of human potential ……. 233

5.6. On the occasion of omniscience and other arcane subjects ……... 242

Page

5.7. Communication in the sacred mirror: integrating rationality ……. 244

5.8. The twofold dimension of mutual understanding: toward our proximate framework and empirical strategy …… 257

5.9. Rescuing depth from flatland: toward and integral approach of the commons dilemma ……... 258

5.10. Buddhas, demons and ordinary people in the action arena ……. 259

6. EMPIRICAL STRATEGY AND THEORETICAL EXPECTATIONS 273

6.1. Accessing the Gravesian psychosocial levels of existence …… 273

6.1.1. Participants and constitution of the sample ……. 279

6.2. Producing behavioral observations …… 281

6.2.1. The common-pool resource experiment …….. 284

6.2.1.1. Experiment summary ……. 285

6.2.1.2. Participants, experimental setting and procedures ……. 287

6.2.1.3. Hypothetic-deductive model of "rational egoists" ……. 289

6.2.2. The public good experiment …… 304

6.2.2.1. Participants and experimental setting …… 304

6.2.2.2. Experiment summary ……. 305

6.2.3. The ultimatum game experiment ……. 306

6.2.3.1. Experiment summary and equilibrium analysis …… 307

6.2.3.2. Results from prior experimental research …….. 309

6.3. Statistical procedures for hypothesis testing …….. 311

6.3.1. Factor analysis ……. 311

6.3.2. Statistical tests ……. 316

Page

6.3.3. Working hypotheses …… 318

6.3.3.1. The levels of existence within the appropriation dilemma 320

6.3.3.2. The levels of existence within the provision dilemma ….. 323

6.3.3.3. The levels of existence and the ultimatum game ……... 326

6.4. Concluding remarks ……... 326

7. RESULTS AND DISCUSSION ……... 328

7.1. Some patterns in the great river of existence ……. 328

7.2. Three principal waves of existence …… 331

7.3. A new look at the sample ……... 340

7.4. A first look at the groups characteristics …… 342

7.5. The levels of existence within the appropriation dilemma …… 343

7.5.1. Aggregate effects of communication and sanctioning ……. 349

7.5.2. Value systems, rule following and economic performance …. 355

7.5.3. Disentangling the connections ……. 398

7.5.4. Bivariate analysis ……. 400

7.5.4.1. The ER system reveals its go-getting character in the unregulated and nonbinding communication settings ….. 401

7.5.4.2. Egocentrics defect significantly more while can be put in line through sanctioning ……... 404

7.5.4.3. Sociocentrics cooperate spontaneously and don't like to impose fines ……. 405

7.5.4.4. No significant relationships with absolutism (Factor 2) but GT strongly reciprocates …… 406

7.5.5. Multivariate analyses ……... 407

Page

7.5.5.1. The higher the score in sociocentrism, the lower the investment after communication (evidence for hypothesis HF2-CPR) …….. 409

7.5.5.2. Egocentrics defect more and sociocentrics defect less (evidence for hypothesis HF2.1-CPR) ……. 409

7.5.5.3. Egocentrics cheat more, sociocentrics cheat less (evidence for hypothesis HF2.2-CPR) ……. 411

7.5.5.4. Sociocentrics support the collective strategy regardless the behavior of others (evidence for hypothesis HF2.3-CPR) 412

7.5.5.5. Sociocentrics pay less fines (evidence for hypothesis HF3-CPR) …… 414

7.5.5.6. The punishers were not sociocentrics but DQ and GT (evidence for hypothesis H4-CPR) …….. 415

7.6. Levels of existence, motivation and the provision dilemma …….. 419

7.6.1. Reframing the working hypotheses ……. 427

7.6.2. Bivariate analysis ……. 429

7.6.2.1. Different waves of existence form different expectations regarding the contribution of others (hypothesis H1-PG and HF1-PG) …… 429

7.6.2.2. Sacrifice-self systems contribute more, express-self systems contribute less (hypothesis H2-PG, HF2-PG, H4-PG and HF4-PG) …… 432

7.6.2.3. Evidences showing the underlining motivations agree with the psychosocial characterizations as set forth by the ECLET conception (hypothesis H3.1-PG, H3.2-PG, HF3.1-PG, and HF3.2-PG) …….. 434

7.6.3. Multivariate analysis …… 438

Page

7.6.3.1. Sociocentrics are more optimist regarding the contribution of others for the low contribution level (hypotheses HF1-PG) ……... 438

7.6.3.2. Sociocentrics are more supportive of the provision of public good (hypotheses HF2-PG and HF4-PG) ……. 439

7.6.3.3. Sociocentrics are motivated by solidarity and sense of duty and not by fear or greed (hypotheses HF3.1-PG and HF3.2-PG) …… 440

7.7. How different levels of existence play the ultimatum game ……. 442

7.7.1. Reframing the working hypotheses ……. 444

7.7.2. Bivariate analysis ……. 445

7.7.2.1. Egocentrics make lower offerings (evidence for hypothesis HF1-UG) …… 445

7.7.2.2. Absolutists reject low offerings (evidence for hypothesis HF2-UG) …….. 447

7.7.3. Multivariate analysis …… 448

7.7.3.1. More on egocentric greediness (evidence for hypothesis HF2-UG) …….. 448

7.7.3.2. More on absolutists' taste for punishing (evidence for hypothesis HF2-UG) …… 449

7.8. Discussion and limitations ……. 450

7.8.1. Our findings impart strengthened cross-cultural reliability to Graves's ECLET conception …….. 452

7.8.2. The experimental observations verified the cyclical turn of the agency versus communion capacities described by the ECLET conception ……. 453

Page

7.8.3. Our experimental observations show that decreasing egocentrism and increasing willingness to cooperate underline the psychosocial development of adult human beings …….. 458

7.8.4. The evidences support that psychosocial development is a separate factor in solving collective action problems ……. 461

7.8.5. Nothing new under the sun: the validity claim of mysticism .. 464

7.8.6. Limitations ……... 469

8. CONCLUSIONS AND IMPLICATIONS …… 473

REFERENCES …….. 485

APPENDIXES ……... 523

APPENDIX A ……..……. 524

APPENDIX B …… 525

APPENDIX C …… 527

APPENDIX D ……... 528

APPENDIX E …… 533

APPENDIX F …… 536

RESUMO

MEYER, Leandro Frederico Ferraz, D.Sc., Universidade Federal de Viçosa, dezembro de 2006. **The commons dilemma revisited: toward an integral approach**. Orientador: Marcelo José Braga. Co-orientadores: Erly Cardoso Teixeira e Marcelo Lobato Martins.

Neste estudo, eu exploro as implicações da Teoria de Desenvolvimento em Estágios Emergentes da Personalidade de Humanos Adultos (Clare W. Graves), em problemas de ação coletiva. A teoria de Graves oferece um modelo empiricamente fundamentado de desenvolvimento *bio-psico-social* de humanos adultos no qual a tensão permanente entre as capacidades de auto-afirmação (*agency*) e auto-adaptação (*communion*) do self origina uma sucessão de modelos mentais que emergem de forma hierarquicamente estruturada – conectando crenças, sistemas de valores, idéias de moralidade e visões de mundo – como resultado do inter-relacionamento entre as variáveis ambientais e a ativação das capacidades potenciais do organismo. As características distintivas de cada estágio de desenvolvimento propostas por Graves (oito níveis existenciais) fornecem a base para a formulação de hipóteses específicas, relativas ao comportamento dos atores humanos em situações de ação coletiva. Em

particular, a superposição cíclica das capacidades de auto-afirmação e auto-adaptação, ao longo dos estágios de desenvolvimento bio-psico-social descritos por Graves, delineiam um padrão de *egocentrismo decrescente*, acompanhado por aumento da autonomia comportamental e ampliação do círculo de participantes (seres humanos ou não) que são incluídos nas considerações morais do agente tomador de decisão. Eu sugiro que estas características apresentam evidentes e importantes implicações na nossa compreensão da dinâmica das interações sociais em situações de ação coletiva. Ao posicionar a concepção de Graves no centro da análise destas situações, eu enfatizo temas na pesquisa da subjetividade e desenvolvimento da consciência que têm recebido atenção inadequada nas principais abordagens sobre ação coletiva e o chamado "dilema da cooperação". Neste sentido, ao ser correlacionada com uma ampla gama de observações empíricas e modelos teóricos de desenvolvimento da consciência, devidas a Ken Wilber, a concepção de Graves fornece a tese substantiva que eu submeto a teste experimental nesta revisão do "dilema dos comuns", qual seja: a diretriz fundamental para promoção de mudanças institucional e comportamental direcionadas a resolver ou atenuar problemas de ação coletiva deve levar em conta, e até mesmo focalizar, os potenciais e as leis próprias da transformação interior e do despertar da consciência humana. Isso porque, contrariamente ao pressuposto pelo modelo unitário de racionalidade (otimização da utilidade esperada), a resposta comportamental à estrutura de incentivos não pode ser adequadamente prevista pela simples consideração das variáveis situacionais. Em vez disso, e para que se obtenham os resultados esperados, aquilo que de fato constituirá tal estrutura precisa ajustar-se ao estágio de desenvolvimento subjetivo dos participantes. Mais ainda, dadas tanto a natureza qualitativa das transformações interiores, delineada acima, quanto o fato de que estas transformações expressam o despertar de potencialidades humanas universais, conforme Ken Wilber, a consideração séria destes processos abre perspectivas totalmente novas sobre como atuar em problemas de ação coletiva e para a promoção das mudanças institucionais e comportamentais sustentáveis. Reciprocamente, a concepção de estruturas de incentivo que assumem um

modelo único e representativo de todos os indivíduo, normalmente fundamentada nos pressupostos cognitivos estáticos que caracterizam o chamado "homem econômico", pode produzir resultados inesperados e desastrosos. Para investigar as implicações e a robustez intercultural da concepção Graves nas situações de ação coletiva, as hipóteses comportamentais derivadas do seu modelo foram testadas em condições experimentais variadas, conforme estabelecidas em protocolos testados em outras aplicações relaciondas à gestão de recursos de propriedade comum e à provisão de bens públicos. Eu começo identificando três principais estágios teóricos de desenvolvimento psico-social presentes numa amostra de 322 sujeitos brasileiros por meio de uma aplicação confirmatória da Análise Fatorial sobre os perfis individuais dos participantes. Estes foram em seguida expostos a três situações experimentais, como segue: (1) um "dilema de apropriação" envolvendo três condições diferentes (sem oportunidade de regulação; duas oportunidades de comunicação direta, sem custo, e sem instrumento de fiscalização; repetidas oportunidades de aplicação de multas em valores variáveis, mediante pagamento de taxa específica); (2) um "dilema de provisão" de um bem público envolvendo nove níveis de contribuições mínima requerida; e (3) um "dilema de eqüidade/risco" conforme representado pelo Jogo do Ultimato. Eu exploro as relações estatísticas (correlações bivariadas e análises de regressão linear multivariada) entre os escores fatoriais nos principais estágios de desenvolvimento identificados previamente e as respostas comportamentais produzidas pelos participantes, nas várias situações experimentais. Os resultados sugerem que as caracterizações dos estágios de desenvolvimento descritas no modelo de Graves são culturalmente robustas e que elas provêem uma base segura para prever o comportamento e discriminar a motivação subjetiva destes comportamentos dentro das situações experimentais. Eu insiro a concepção de Graves no contexto da pesquisa institucional em ação coletiva usando a Teoria Integral de Ken Wilber. A Teoria Integral fornece o mapa conceitual para um tratamento mais amplo e equilibrado dos diversos domínios existenciais e epistemológicos. Nesta aplicação, ela provê a base teórica para *integrar* as dimensões comportamental, sistêmica, psicológica e cultural relacionadas aos

problemas de ação coletiva. A contribuição de Wilber fortalece e estende a visão de Graves, sugerindo que as características de redução do egocentrismo e ampliação do escopo da ação moral descrevem uma tendência universal, divisada em virtualmente todos os modelos de desenvolvimento humano, como expressão básica do despertar da consciência integral (Kósmica). Eu utilizo estes achados para apoiar a generalização dos resultados experimentais produzidos nesta pesquisa e o significado profundamente espiritual envolvido na tese substantiva deste estudo.

ABSTRACT

MEYER, Leandro Frederico Ferraz, D.Sc., Universidade Federal de Viçosa, December, 2006. **The commons dilemma revisited: toward an integral approach**. Adviser: Marcelo José Braga. Co-advisers: Erly Cardoso Teixeira and Marcelo Lobato Martins.

In this study, I explore the implications of Clare Graves' Emergent-Cyclical Double-Helix Levels of Human Existence (ECLET) conception of adult personality and cultural institutions on collective action problems. Graves' theory offers an empirically grounded model of adult *biopsychosocial* development in which both the *agency* and the *communion* capacities of the self (the double-helix) underpin a developmental sequence of evolving mental models – connecting values and belief systems, worldviews and ideas of morality – that unfold as a result of the interplay between the changing life conditions and the neuronal coping systems of biologically mature human beings. The defining characteristics of each of Graves' constructs (eight hierarchical levels or waves of existence) support the advancement of specific hypotheses concerning the behavior of the human actors in collective action situations. Graves' model of cyclical and intermingling fluctuations of the agency and communion capacities of the self suggests a pattern of *decreasing egocentrism* and *increasing moral*

embrace along the path of human psychosocial development. This trait presents clear and useful ways to increase our understanding of the dynamics in collective action situations. By bringing Graves' ECLET conception to the forefront, I emphasize issues of subjectivity that have received inadequate attention in mainstream approaches to collective action. The ECLET conception thus imparts the substantial thesis we hold in this reassessment of the commons dilemma, namely: *the prime directive for institutional change or intervention, if proper, should focus upon and foster the awakening of human potential (i.e. interior growth or consciousness development). In order to bring about the expected outcomes, the structure and type of incentives must match the subjective level of existence of the individuals and communities, which evolves along a definite sequence, according to Graves' model. Conversely, the conception of incentive structures that assume a representative model of individual, usually grounded on the unitary and static cognitive foundation that characterizes the so-called "economic man", may produce unexpected and misleading results.* To check the value and implications of Graves' ECLET conception on collective action, the behavioral hypotheses derived from it were tested in experimental conditions following well established experimental protocols found in the literature of common-pool resource management and the provision of public-goods. I begin by assessing the psychosocial systems of 322 subjects' (mostly Brazilian undergraduate and graduate students) and submitting the data to Factor Analysis in order to extract three principal levels or waves of existence, consistent with Graves' theory. Participants were then exposed to three experimental situations, as follows: (1) a common-pool appropriation dilemma under three different institutional settings (no communication, costless communication, and costly sanctioning), (2) a public-good provision dilemma repeated under nine requisite contributions levels, and (3) an equity dilemma as represented in the well-known Ultimatum game. I then explore the statistical relationships (bivariate correlations and multivariate regression analysis) between Graves' constructs and the behavioral responses produced in the experimental situations. The results show that Graves' constructs are cross-culturally robust and that they provide a

reliable basis for predicting both behavior and motivation within the experimental situations, thus supporting the prime directive sketched above. I connect Graves' approach within the scope of collective action research using Ken Wilber's Integral Theory. Centered in the concept of *holon* and the Whiteheadian transcend-and-include conception, Integral Theory provides a path-braking map of the *Kosmos* and evolutionary history around four elementary domains of reality: subjective, objective, intersubjective and interobjective. In this application, Integral Theory provides the framework to integrate the behavioral (objective), the systemic (interobjective), the psychological (subjective) and the cultural (intersubjective) aspects of collective action problems. Moreover, Wilber's contribution strengthens and extends Graves' theory by combining it with extensive research on the evolution of consciousness using both Eastern and Western and ranging from premodern to modern to postmodern sources. In particular, Wilber shows that the developmental trait defined by decreasing egocentrism and increasing moral embrace, present in Graves' conception, is a deep-rooted trend found in virtually all developmental systems. I utilize these findings to support the interpretation of our experimental results and evaluate the significance of the prime directive in the context of collective action problems.

1. REFRAMING THE INQUIRY ON COLLECTIVE ACTION: TOWARD AN INTEGRAL APPROACH

1.1. Introduction

How we frame the inquiry into any subject has decisive implications. In a very broad definition, a framework is a basic conceptual scaffolding, used to solve a complex issue. In software development, for instance, a framework is a defined support structure in which another software project can be organized and developed. In programs of scientific research, a sufficiently general framework provides a metatheoretical language for thinking about diverse theories and models and investigating their potential usefulness in addressing questions of relevance to the analyst (OSTROM et al., 1994).

However, as a quality inherent to its very appeal, frameworks outline *limited* courses of action and suggest *preferred* approaches to a given subject. These basic definitions develop from some initial consensus of their protagonists, whence core propositions, assumptions, or premises emerge as a set of methodological decisions that configure the very conditions of possibility of that research program or discipline. In this way, together with the selected array of observations or data they disclose, alternative frameworks not only set forth different guidelines to address a certain problem but also bring about

fundamentally distinct viewpoints on a same phenomenon. Although different perspectives on a complex issue should in principle compound a better picture, the diverse foundations of alternative theoretical structures commonly engender solid barriers to transdisciplinary communication.

The consequences of approaching collective action problems from different premises (both philosophical and methodological) are particularly significant due to the archetypical conflict between the individual and the collective benefits that defines the situation as a social dilemma. Putting aside the increasing sterile debate about the merits of "micro" versus "macro" explanations and the status of "methodological individualism",[1] the inquiry on collective action problems is especially tricky, from a scientific point of view, because the opposition between the individual and the collective interests actually generates an essentially *moral dilemma*, whose scientific assessment depends not only on what is assumed about the nature of the human actor and the circumstances of action, but also on rather challenging epistemological issues involving the *cognitive status* of values judgments, moral stances, and questions of justice.

Beginning with the models of individual and the nature of the social interactions, Heath (2001) remarks that a conspicuous feature of the social sciences is that different disciplines are dominated by different theories about how people act, i.e., what psychological states govern their decisions, what incentives they respond to, and how the actions of multiple agents are coordinated. Also Gintis (2003) notices that, while each of the behavioral disciplines contributes strongly to human behavioral science, taken separately and at face value, they offer partial, conflicting, and incompatible models of human behavior.[2]

---

[1] Methodological individualism, strictly specking, is just the view that explanations of social phenomena are *incomplete* until some account of their action-theoretic underpinnings has been supplied. According to Heath (2001, note 3, p. 318), "It becomes controversial only when paired with the further claim that these action-theoretic explanations must be instrumental in form". Some controversy may also arise from attempts to derive the methodological individualism directly from an *ontological* individualism, such as "to *be* is to be in the mind of the *individuals*". As Blaug (1980) notices, the point in this regard is that such ontology has no necessary relationship with the manner how we are supposed to investigate collective phenomena, that is, with the matter of *methodological* individualism (cf. BLAUG, p. 85-91).

[2] Gintis (2003) indicates that the human behavioral sciences include economics, human biology, anthropology, sociology, political science, and behavioral psychology (by which he means social

These observations suggest, first, that the inquiry on social dilemmas has no obvious disciplinary home, and then that the conventional economic theory of "externalities", grounded as it is in a rather particular and unitary model of individual (i.e., self-interested or norm-free maximizers of the individual utility) and a limited form of social interactions (the interactions mediated through the markets), tells us just one version of the story – one that proved powerful in predicting outcomes as well as in deriving policy prescriptions to cope with the overexploitation of "common-pool resources"[3] and the deficient provision of "public goods"[4] in settings resembling highly competitive markets but that have potentially seriously misleading implications when applied outside this sphere.

Indeed, concerning the methodological bond between what is assumed about the human actor and the circumstances of action, Persky (1995, p. 230) observes on behalf of Stuart Mill's original abstraction that "the message to derive from Mill's *homo economicus* is not that humans are greedy, nor that man is rational, but that social science works best when it ruthlessly limits its range". However, Manski (2000) notices that economists have long been ambivalent about whether the discipline should focus on the analysis of markets or should be concerned with social interactions more generally. "From the vantage of today", he says, "it is clear that the narrowing of economics ended by the 1970s. Since then a new phase has been underway, in which the discipline seeks to broaden its scope while maintaining the rigor that has become emblematic of economic analysis" (p. 1). In characterizing this process in microeconomics, Manski points to the adoption of noncooperative game theory as a language and set of tools for the study of market and other interactions as possibly the defining event of the late twentieth century. He explains that game theory encouraged economists to see all interactions as games, markets being special cases. As a result, he says, "economic theorists have in recent years studied phenomena as far from traditional economic concerns as the evolution of social norms" (p. 2).

---

psychology and psychological decision theory). By "human biology" Gintis means the application of biological techniques to modeling human behavior.

[3] A common-pool resource (CPR) is a resource system where it is costly to exclude potential beneficiaries, but one person's use subtracts resource units from those available to others.

[4] A public good (PG) is defined as a nonexcludable and nonrival good.

Clearly, the said "evolution of social norms" is a central theme in addressing self-organization in collection action settings, and this study is heading for shedding some light precisely on this matter. However, instead of approaching the issue right from the evolutionary perspective – and thus rely on a sociobiological framework that has been pointed out as empirically inappropriate to address norm-governed behavior (more on this issue bellow) – I explore the *developmental orientation*, in order to examine the implications of the *hierarchical* emergence of a series of empirically grounded personality systems on the behavior of concrete individuals embroiled in experimental social dilemmas under variable institutional settings.

The importance of inquiring on the diverse interior dispositions of the individuals engaged in social dilemmas becomes clear as one realizes that, as far as the selfishness axiom of the postulate of (economic) rationality is retained, the game theoretic account of collective action problems just restates the conventional theory and prediction, so that no social arrangement (rule or norm) shall ever emerge from the commons dilemma situation. As Ostrom summarizes this issue with respect to the governance of common-pool resources (CPRs), the theory-laden dilemma is that organizing so as to create rules that specify rights and duties of users of a CPR creates a public good for those involved. That is, anyone who is included in the community of users benefits from this public good, whether they contribute or not. Given the standard assumption of self-interested behavior (which subtly identifies egoism and rationality) getting "out of the trap" develops into a second-level dilemma. Further, investing in monitoring and sanctioning activities in order to increase the likelihood that participants follow some agreement (norm or rule of interaction) they have eventually made also generates a public good. The provision of such investments, therefore, represents a third-level dilemma. But "since much of the initial problem exists because the individuals are stuck in a setting where they generate negative externalities for each other, it is not consistent with the conventional theory that they solve a second- and third-level dilemma in order to address the first-level dilemma under analysis" (OSTROM, 1999, p. 3).

Therefore, as suggested, the standard game-theoretic account of this kind of situation just restated the convention interpretation. In Ostrom's words, "[while] these models formalize the problem differently, [they] do not change any of the basic theoretical assumptions about the finite and predictable supply of resource units, complete information, homogeneity of users, their maximization of expected profits, and their lack of [direct] interaction with one another or capacity to change their institutions. Nor is the prediction changed" (ibid, p. 2). Hence, just like the conventional theory, the standard game-theoretic analysis represents individuals trying to provide public goods or sustain common-pool resources as trapped in Prisoner's dilemma-like situations with no way out on their own means – whence the classic policy recommendations involving compulsory contribution and the specification of property rights, traditionally opposing the advocates of privatization (DEMSETZ, 1967; POSNER, 1977; SIMMONS et al., 1996) to those of government ownership and control (OPHULS, 1973).

Of course, were the reliability of the standard predictions and the adequacy of the corresponding policy measures not challenged by the evidences, one would have no good reasons (other than the philosophical quest for true explanations, and some concern vis-à-vis the continual scholarly inculcation of self-interest as if it was the same as rationality) to contend the instrumental quality of the conventional theory just on account of the unrealism of its presuppositions.[5] However, as Ostrom (1999) indicates, the growing evidence from many field studies of CPRs – including irrigation systems (TANG, 1992), fisheries (SCHLAGER, 1990), rangelands (BLOMQUIST, 1992), pastures (AGRAWAL, 1999) and forests (FORTMANN; BRUCE, 1988; FOX, 1993; FAIRHEAD; LEACH, 1996) – have challenged the applicability of the standard theory by showing how the users in many locations have organized themselves to

[5] The proposition, due to Hempel and Oppenheim (1965), that the operation called *explanation* involves the same rules of logical inference that the operation called *prevision* is known as the "thesis of symmetry". Under the thesis of symmetry, successful predictions based in the conventional model of individual might appear as evidence of the realism of its premises (as source of good explanations). The symmetry thesis has been increasingly discredited, though (see Popper, 1959, p. 61; Popper, 1972, p. 191-195). While it is doubtful the one may have good explanation without prediction, the idea that true explanations require more than good predictions is generally accepted (see Blaug, 1980, p. 38-46).

vigorously protect and, in some cases, enhance recourse conservation by successfully regulating the collective appropriation. Similarly, experimental researches using games consistently show that cooperation between participants arises quickly in finitely repeated Prisoners' Dilemma situations (BOONE et al., 1999; see also PRUITT; KIMMEL, 1977; RAIFFA, 1982; OSTROM et al., 1994, among others), and that the opportunities to communicate facilitates significantly the adoption of cooperative strategies, thus improving the efficiency of both the collective appropriation of common-pool resources and the provision of public goods, *even when no effective enforcement mechanism is available* (cf. e.g. ISAAC; WALKER, 1991; HACKETT et al., 1994; ROTH et al., 1992; OSTROM et al., 1994; FEHR; GÄCHTER, 2000).

In view of the vast empirical research on collective action problems of the last several decades, Frohlich and Oppenheimer (2001, p. 22) reflect that the so-called "anomalies" which have been identified in light of the standard rational choice model "are broad and diffuse and they are likely to require more theoretical superstructure for their explanation". More specifically, Ostrom indicates that while the conventional theory is successful in predicting outcomes in settings where users are alienated from each other or cannot communicate effectively, it does not provide an explanation for settings where users are able to create and sustain agreements to avoid serious problems of overappropriation. Nor does it predict well when government ownership will perform appropriately or how privatization will improve outcomes. "Often", she says, "problems in forest management have emerged when local self-organization has not been recognized by policy makers, and the autonomy of forest users to continue their own forest use practices has been threatened" (OSTROM, 1999, p. 3; see also ARNOLD; CAMPBELL, 1986; GILLIS, 1988; ARNOLD; STEWART, 1999; BOWLES; GINTIS, 2001; OSTROM, 2005 and 2007).

According to Ostrom's (2005) view, in spite of a number of puzzles still to be solved, the conventional theory of "externalities" has now been shown to be a special case of a more general theoretical structure. In particular, concerning the assumptions about the behavior of the individuals – including the *valuations*

that they assign to actions and outcomes; the way how they acquire, process, represent, retain, and use *information*; and the *choice processes* that they use for selecting actions in light of their resources – Ostrom suggests that in cutting through the complexity inherent to social dilemmas we are likely to make more progress if we do *not* try to develop a single model of human behavior that can be used to predict behavior in all market and nonmarket action situations. According to her, while assuming that all preferences are self-regarding without exploring how individuals reach *utility judgments* has been a satisfactory strategy to model behavior in highly competitive markets (where utility is implicitly equated with the material payoffs), it is not so when one turns to the large realm of social dilemmas because these situations frequently evoke internal values that are *not* monotonically related to the objective payoffs involved. Accordingly, depending on the context of the situation, previously socialized individuals may add *diverse* (positive or negative) normative payoffs to their preference function. As a consequence, contrary to what is assumed in the standard game theoretic account, collective action problems bring about a situation where one normally *cannot* suppose that all individuals who face the same objective choice structure evaluate decisions in the same way. Moreover, she adds, "even if the objective circumstances configures a relatively simple social dilemma, the situation is one of incomplete, rather than, complete information because other participants cannot know exactly how an individual is valuing these actions and outcomes" (ibid).

Indeed, Hurwics's theory of compatible incentives demonstrates the implication of both the normative orientations of the individuals and the means they may apply to influence the behavior of each other on the equilibria of games representing the "genuine implementation" of institutions (HURWICS, 1998). As he puts it, "the equilibria of the true game [i.e. the prevalence of the rule of the norm and its enforcement] depend on what is feasible and the actual consequences, as well as on the preferences (sometimes called attitudes or values) of the players" (p. 13). Even if discounting on Plato's reflection, in *The Republic*, on the ridiculousness of a guardian in need of being guarded himself,

Hurwics explains that, if we think of an external authority or intervener, he or she "must have preferences that rule out his/her own illegal behavior (especially corruption) and harbor a dislike for others' illegal behavior". Akin to Ostrom's focus on self-organization, Hurwics points that we do not have to rely on the presence of external authorities, for there are other structures conducive to successful enforcement. In particular, he presents a circular model where everyone is a guardian but also guarded, in order to suggest that enforcement is possible of being modeled in the context of lager games with incentive constraints.[6] However, all the same, in order that the norm or law may be actually implemented, each guardian of the guardians must have powers, assets, charisma, or perhaps rational arguments enabling him or her to *influence others* in the desired direction, whilst both the chosen action and its actual influence upon others are again reliant on the values and attitudes (sometimes called preferences) of the individuals in interaction. Consequently, "depending on the participants in the game (i.e., the composition of the society) *and their preferences/values*, a given desired mechanism may or may not be implementable within that society" (HURWICS, op. cit., p. 14, my italics)[7].

So far so good. Yet, as abovementioned, while inquiring on how the individuals reach utility judgments may be a secondary issue in inferring behavior in the market setting, "how can we understand a person's choices if, when confronting a given environment, she can have multiple representations, each of which is associated with different values?" (FROHLICH; OPPENHEIMER 2001, p. 6). In particular, how can we explain normative agreement and commitment in collective action situations as we admit that, on

[6] Hurwics suggests that such circular model is an essential aspect of democracy, where the voters work as "guardians" of top guardians. He adds that this type of structure is also closely related to (but not identical with) the notion of separation of powers.

[7] Myerson (2006) reviews Hurwics's general theory of how institutions can be enforced in larger games, suggesting curb sets and the development of group identity as an alternative enforcement theory that admits focal-point effects. He also considers specific problems of leadership and trust in establishing political institutions, where officials can be deterred from abuse of power only by promises of large future rewards, which a leader must be credibly committed to fulfill. I note that Myerson's alternatives do not change the fact that enforcement continue conditional to the individuals' attitudes or values in face of the so-called "focal points" and what counts as rewards.

the basis of disparate internal values and normative orientations, the individuals may evaluate any given incentive structure differently?

In view of these puzzling difficulties, Ostrom (2005) suggests that one of the main theoretical challenges in explaining successful and unsuccessful efforts to engage in collective action to provide public goods and sustain common-pool resources consists in developing "an appropriate family of assumptions to make about the intrinsic values individuals place on actions and outcomes – particularly outcomes obtained by others". According to her, "without further progress in developing our theories and models of human valuation in social dilemma situations, those convinced that all human behavior can be explained using 'rational egoist' models will continue to recommend Leviathan like remedies for overcoming all social dilemmas". Further, she points that we need address how to focus on the role of norms and other regarding preferences, without falling into the trap that all that is needed is to assume that individuals learn and use norms. "We then need to dig into the analysis of institutions so that we can understand how individuals adopt norms as well as rules to overcome social dilemmas". In this context, she draws particular attention to the importance of deepening our understanding of the role of communication in fostering cooperative strategies in collective action settings (see also BANKS; CALVERT, 1992a and 1992b; OSTROM, 2003; OSTROM et al., 1994).

Similarly, in providing a broad overview of the interplay among cognition, belief systems, and institutions, and how they affect economic performance, Mantzavinos et al. (2004) highlight the importance of a *theory of learning* as a prerequisite for explaining societal change. As they put it, "the ability to learn is the main reason for the observed plasticity of human behavior, and the interaction of learning individuals gives rise to change in society, polity, economy, and organizations" (p. 75). Accordingly, the authors suggest that the real issue in understanding human choices regarding institutional change is not whether agents are perfectly or boundedly rational, but rather how human beings *actually* reason and choose, individually and in collective settings. Given that learning is the main object of inquiry in cognitive science, the authors argue that

cognitive science "should be the *starting point* for any serious discussion of societal change", and that "only a dogmatic attitude would prevent social scientists interested in phenomena related to change from paying appropriate attention to its findings" (see also DENZAU; NORTH, 2000; NORTH, 2005).

In this study, I present Clare Graves's Emergent-Cyclical Levels of Existence Theory (ECLET) of personality systems development and cultural institutions (GRAVES [1971] 2005) as both the starting point for understanding how individuals reason and choose, particularly in collective action settings, and the source of empirically grounded information to generate the family of assumptions about the intrinsic values individuals place on actions and outcomes, particularly outcomes obtained by other, as requested by Mantzavinos et al. (2004) and Ostrom (2005), respectively.

As we shall see, the point of the developmental orientation is that it confers *restrictions* to the range of behavioral responses that are open to individuals centered at different levels or stages of psychosocial development. Such restrictions in turn provide the basis for predicting those behavioral responses in face of alternative incentive structures, as well as to evaluate the adequacy of those structures, for the purpose of overcoming social dilemmas, depending on the psychosocial profile of the participants in the situation. In addition, by allowing us to take into account the *hermeneutic dimension* of mutual understanding, the developmental orientation sheds new light – beyond the exclusionary instrumental account – on the limits and potentials of communication opportunities to foster normative agreement and commitment to overcome social dilemmas.

The specific purpose of this study is then to test experimentally the prediction power associated to Graves's constructs, so as to better inform policy analysis vis-à-vis the implications and potentials resulting from taking the developmental orientation into account when inquiring on both the consequences of "externally" imposed institutions and the likelihood of self-organization in collective action settings.

However, as suggested in the beginning, in order to explain the coherence and adaptability of norm-governed systems by appealing to cognitive recourses that agents themselves can deploy in managing their interactions (particularly in the case of self-organization in social dilemma situations), we must get through a rather challenging epistemological barrier, as construed by the *correspondence theory of truth* and the *foundationalist model of justification*, and articulated in empiricist objections to cognitivist approaches in moral theory. This effort is required because, located in the hermeneutic dimension, the internal connection between norms and justifying grounds settles outside the restricted concepts of knowledge, rationality, and truth that are oriented to the modern empirical sciences (cf. HABERMAS, 1993; see also WILBER, 2006, excerpt B).

Consequently, the appreciation of the theoretical perspective underpinning this study, and from which its central thesis is launched, requires to a great extent a framework quite broader than any of those that can be construed under the canon of the instrumental conception of rationality. I found out such framework in Ken Wilber's Integral Approach (WILBER, 2000a, 2000b and 2001), to be outlined in the next chapter. As we shall see, Wilber's Integral view brings forth a path-breaking epistemological integration that, as far as I can tell, has never really happened before in the history of humankind.

While the significance of the Integral Approach for the investigation of collective action problems shall become entirely clear only after we have looked at its theoretical foundation and structure, I will introduce in this chapter the reasons that render it appealing to address those situations by carrying out a little further the discussion of the epistemological hindrance embodied in the exclusively instrumental conception of rationality. Before addressing this intricate issue, however, I offer in the next section a brief outline of the main characteristics of Graves's theory, so as to expound first on the significance of the developmental orientation for the inquiry on collective action dilemmas. In doing so, I will at the same time introduce the preliminary working hypotheses and submit the central thesis of the study.

1.2. Overview of Graves's ECLET, working hypothesis, and the thesis of the study

As the inquiry on collective action moves away from the cognitive foundation of *Homo economicus*, the lack of a theory of mind that specifies agents' relevant psychological features affecting their choices in face of alternative institutional settings renders the prediction of behavior virtually impossible. Although the growing information on how different personality types and cultural traits explain the behavior of the participants in collective action experiments and Prisoner's Dilemma games certainly helps to advance more effective approaches to overcome social dilemmas, these undertakings remain frequently limited to isolated qualities such as ostracism (HIRSHLEIFER; RASMUSEN, 1989), locus of control (PHARES, 1976; LESTER, 1992; COOK; CHI, 1984), guilt-proneness (PRICE, 2005), time urgency (CURRY, 2005) that, even when considered in combination (e.g. BARON, 1989; BOONE et al., 1999), nonetheless lack the metatheoretical scaffolding that would be required in order to arrange such diversity of subjective qualities into a coherent fashion.

The distinctive feature of the developmental orientation is that it recognizes into the complexity of human subjective experience both *structure* and *internal dynamics* that not only allows the identification of diverse intentional dispositions that assists predicting behavior, but also suggests in which direction the behavioral changes are more likely to occur in response to different institutional incentives. Actually, a central tenet in developmental psychology is that, in order to produce the expected results, the type of incentive should be tuned to the characteristics of each psychosocial centralization stage.

As "a marriage of the cognitive-developmental and existential systems of thought" (GRAVES, 2005, p. 29), and following about a decade of careful empirical research, Graves conceptualized eight waves, stages, or *Levels of Human Existence* which provide a description of states of *biopsychosocial* equilibrium, comprising a perception of the environment, a reciprocal state of neurochemical balance, reflected in a social construction that then influences those mental states of equilibrium, as part of the environment perceived

(VAREY, 2006). In this vein, rather than purely a psychological study, Graves's orientation integrates "bio-", "psycho-", and "socio-", and postulates that the biopsychosocial development of human beings arises from the interaction of a double-helix complex of two sets of determining forces: (a) the *environmental social determinants* (the existential problems of living), and (b) the *neuropsychological equipment* of the organism for living (the biological coping system). In Graves's words:

> The psychology of the adult human being is an unfolding, ever-emergent process marked by subordination of older behavior systems to newer, higher order systems. The mature person tends to change his psychology continuously as the conditions of his existence change. Each successive stage or level of existence is a state through which people may pass on the way to other states of equilibrium. When a person is centralized in one of the states of equilibrium, he has a psychology which is particular to that state. His emotions, ethics and values, biochemistry, state of neurological activation, learning-systems, preference for education, management and psychotherapy are all appropriate to that state. If he were centralized in some other state he would think, feel and be motivated in manners appropriate to that state. He would have biochemical characteristics and state of neurological activation particular to it. When in a certain state, he would have opened only certain systems for coping and learning. Thus, he would respond most positively do education, management, and therapy which are congruent with that state. And he would have to respond negatively to forms of education, management and therapy not appropriate to the state of his centralization (GRAVES [1981] 2005, p. 29-30)[8].

Beginning with the recognition that human interiority comprises multiple *lines* or *streams* of consciousness unfolding (e.g. cognition, social perspective taken, interpersonal intelligence, morals, value systems, and so on) that develop in a *relatively* independent fashion, the theoretical constructs put forward by Graves should *not* be taken as rigid levels or stages but rather as "flowing waves, with much overlap and interweaving, resulting in a meshwork or dynamic spiral of consciousness unfold" (WILBER, 2001, p. 7, see also BECK; COWAN, 1996). This flowing nature is what renders human behavior so hard to explain and model. Nonetheless, as indicated, due to unique motivational needs that differ within each stage, the distinctive traits associated with different states of

[8] While distinctively open-ended, Graves's model admits that the actual development frequently become arrested, in different stages, due to several factors, including different biological potential and the varied attributes of the sociocultural environment. We shall return to this issue in the chapter three.

equilibrium provide the basis for predicting *different* behavioral responses to otherwise presumed unequivocal incentive structures.

As a theoretical contribution to phenomenological, existential, humanistic and cognitive emergent stage theoretical psychologies, as Graves (2005, p. 7) puts it, the ECLET deals with at least five of the major criticisms generally directed at them. First, it offers an explanation of how the constructs develop. Second, it presents a picture of what the process of development is like. Third, it hypothesizes what factors determine the (nested) hierarchical order of the constructs. Fourth, it explains what determines the particular characteristics of constructs. And, fifth, it suggests how the constructs operate.

Furthermore, as Graves's methodology establishes an internal connection between the stages of sociocognitive development and individuals' openness to revise their conceptions of adult healthy personality and social competency in view of both authority and peer argumentations, Graves's model presents a number of important characteristics that renders it particularly useful for inquiring on the role of communication opportunities in fostering normative commitment, particularly in situations where the individual and the collective benefits conflict.

According to Graves's scheme, adult development entails a cyclical alternation between a focus upon the external world and attempts to change it and a focus upon the inner-world and how to come to peace with it. As indicated, though each system must be viewed with a different premise, out of their own specific aims and means, the outward focus brings forth a family of *express-self systems*, where the locus of control is within the individual, while the inward focus underlines a family of *sacrifice-self systems*, where the locus of control is outside the individual. Hence, the theory puts forward that, along the path of consciousness awakening, people tend to oscillate back and forth between these two fundamental stances, much like the relative position of a pendulum in its arc between "me" (agency) and "we" (communion) orientations (COWAN; TODOROVIC, 2005).

Clearly, even if as a still crude inference, to be refined latter, the initial suggestion here is that:

> In the context of collective action dilemmas, choices exhibiting concern for the collective good are more likely to come about from some of the sacrifice-self systems, while the family of express-self systems are expected to bring forth more self-regarding choices.

The critical point, however, is that we are talking about *universal stages of development* or *consciousness unfolding*, not of mere *types* of personality systems resulting from either the contingent features of our traditions or individual histories. Furthermore, still according to Graves's model, the developmental sequence involves a series of *cumulative gains*, so that Graves's scheme presents a *nested hierarchical* configuration. In this fashion, any given personality system arises when certain basic competencies defining former states of equilibrium are put to use in the production of more complex behavioral structures. Thus, each Level of Existence does not simply replace the prior ones as the preferred form of coping with the challenges of the environment, but rather sublates them. This means that lower or earlier existential structures are still important developmental prerequisites of any given higher level system. As a result, the developmental order brings forth a trend defined by *increasing behavioral freedom* and *autonomy*, which runs beneath the cyclical pattern just described.

On the one hand, this nested hierarchical organization explains human behavioral flexibility and adaptability, as it has been recognized as a basic feature of complex systems (cf. e.g. SIMON, 1962; KOESTLER, 1973; ARTHUR et al., 1997; CILLIERS, 2005). On the other hand, as the principle of hierarchical inclusiveness entails an increasing conciliation of the fundamental tension between the agency and communion dispositions, this pattern imparts a rather important *qualitative change* in human sociocognitive and emotional capacity to engage in relations of *reciprocal recognition and respect*. This trait is what underpins Wilber's (2001, p. 18) central claim that *human interior development is a process involving, for the most part, a continuing decline in egocentrism and*

increasing ability to take other people, places, and things into account when making decisions.

Wilber's assertion in backed by an extensive analysis of coherence comprising over one hundred developmental models, including cognitive development, self-related stages of morals and perspectives, stages of spirituality, and sociocultural development (cf. WILBER, 2000b). The same trend was verified in Graves's own analysis of the consistence between his model and those of Maslow, Kohlberg, Loevinger, Erikson, Perry, Selman, and Fromm, among many others (cf. GRAVES, 2005). Finally, the development of the decentered stance toward the world and considerate attitude toward the interest of others is a fundamental quality also in Habermas's phylogenetic account of communicative action and "discourse ethics", which he presents in the form of an interpretation of work by Emile Durkheim and Herbert Mead (HEATH, 2001, p. 35; cf. HABERMAS, 1984/1987, fifth chapter). Furthermore, Habermas sees in Kohlberg's theory of the development of moral consciousness – one that concurs and "stands in confirmation" of the Levels of Existence theory, according to Graves (2005, p. 453-454) – an ontological recapitulation (and an indirect validation) of the genesis of the form of social interaction represented by communicative action. More specifically, as Heath puts it, Habermas tries to show that the stages in Kohlberg's model are *isomorphic* to the stages that occur in his own reconstruction of the development of communicative action (ibid; cf. HABERMAS, 1990).

The implication is that if one adopts "criteria of coherence to govern the *division of labor* between *philosophical ethics* and a *developmental psychology* designed to rationally reconstruct the pretheoretical knowledge of competently judging subjects" (HABERMAS, 1990, p. 118, emphasis in the original), then one can establish, as Habermas does, a bond between developmental psychology and hermeneutics so at to advance, beyond the instrumental account, our understanding of the role of communication in fostering normative commitment in collective action settings. In this sense, i.e., by assuming the existence of such internal connection between the stages of psychosocial centralization and the

likelihood that a group of people might achieve mutual understanding on normative questions, I raise the additional preliminary hypothesis that:

> The genuine implementation or stability of nonenforceable normative agreement (and then the economic performance) in collective action dilemma situations is more likely when agreement is settled among individuals centered at some of the sacrifice-self systems, as well as among individuals centered at the highest stage of psychosocial development, as described in Graves's scheme.

Actually, if human interior growth is really marked by a continuing decline in egocentrism and increasing ability to take the interests of other into account, then the potential implication of the developmental framework on the analysis of social dilemmas seems just quite obvious as remarkable. In a sense, it patches up with empirical evidences the century long philosophical debate about whether good government depends upon better people or better institutions – a debate that raged anew among the 18th century political economists, whose work provides a foundation for new institutional economics. In this context, as one abandons both the *aprioristic* notion that preferences are self-regarding and that resource scarcity unmistakably renders opportunist behavior rational, the cognitivist-developmental viewpoint on values and morals allows us to specify in which sense, more explicitly, the composition of the group of people in interaction and their preferences/values affect the likelihood that a norm may be "genuinely implemented", as Hurwics (1998) would put it.

Insofar as such sense is actually *developmental*, meaning that human interior dispositions (underlying different attitudes, preference, or values) present its own stages of growth and laws of transformation, being neither an unequivocal function of the circumstances of action nor a mere contingent feature of our traditions or individual histories, then the **central thesis** of this study just restates Graves's suggestion that "*the prime goal of any societies' governing figures should be to promote human movement up the levels of human existence*" (GRAVES, 2005, p. 483). More specifically, in the context of the analysis and development of institutions to overcome collective action dilemmas, I am aimed at defending the idea that:

> The criterion to judge the desirability of alternative incentive arrangements, together with the quality of the efforts dedicated to analyze and develop them, should combine the ability to guarantee the efficiency and sustainability of resource use with the capability to foster human upwardly move along the stages of interior growth.

Bearing in mind the nature of Stuart Mill's revision of classic utilitarianism[9], and his aspiration of grounding political economy in a general Science of Ethology – one that would disclose the *general laws of mind and the formation of the character*[10] (cf. MILL, 1843) – it seems noteworthy that Graves's suggestion and the thesis of this study are quite in line with the "twofold division of merit" according to which, in Mill's view, the political intuitions should be considered. In his *Considerations on Representative Government*, for example, Mill's put forward that the merit of any set of political institutions:

> consists partly of the degree in which they promote the general mental advancements of the community, including under that phrase advancements in intellect, in virtue, and in practical activity and efficiency; and partly of the degree of perfection with which they organize the moral, intellectual and active worth already existing, so as to operate with the greatest effort on practical affairs.

Seeing as Mill's political recommendation may appear an "expression of the hopes and fears of an eminent Vitorian", as Gray (1998, p. xxviii) puts it, while the moral substance embedded in the developmental path may likewise appear to some as a particular expression of perfectionism, one should notice that Graves's statement does *not* imply that one style of being or form of human existence is, inevitably and in all circumstances, superior or better than another form of human existence. Actually, Graves's ECLET provides no ethic of good life, nor does it offer the ethological basis in which one might want to found, as

[9] According to Gray's (1998) report, Mill's attempt to move away from the Benthamite value-subjectivism and utilitarianism of pleasure and painful mental states undertook three major changes, leaning not so much toward any straightforward sort of preference-utilitarianism, but toward a more Aristotelian ethical theory, in which value is connected with self-realization.

[10] Placed logically subsequent to elementary psychology, according to Mill's view, human ethology was that science that "determines the kind of character produced in conformity to those general laws [of psychology], by any set of circumstances, physical and moral" (MILL, 1843, p. 869, my italics). Working from the base of such general ethology, Mill argued that "economic ethology" would then construct "a theory of the causes which determine the type of character belonging to a people or to an age" (MILL, 1843, p. 905).

Mill intended, a political economy based on the principles of utilitarianism and individual liberty. Indeed, as Wilber (2001, p. 56) remarks, the developmental feature actually suggests that even if every society on earth were somehow established fully at highest level thus far, nonetheless every infant born in every society still has to start at the stage of sensorimotor instincts and perceptions, and then must grow and evolve all through the other levels or stages. Furthermore, the nested hierarchical structure means that all stages have important tasks and functions, so that none of them can be bypassed, and none of them can be demeaned without grave consequences to self and society. In this way, according to Wilber, "*The health of the entire spiral [of consciousness unfolding] is the prime directive, not preferential treatment for any one level*" (ibid).

Therefore, insofar as we can take the commons dilemma as a metaphor for the current ecological crisis, the thesis of this study can also be seen as a specification of Wilber's striking statement that:

> The ecological crisis – or Gaia's main problem – is not pollution, toxic dumping, ozone depletion, or any such. Gaia's main problem is that not enough human beings have developed to the postconventional, worldcentric, global levels of consciousness, wherein they will automatically be moved to care for the global commons. And human beings develop to those postconventional levels, nothing by learning system theories but by going through at least a half-dozen major interior transformations, ranging from egocentric to ethnocentric to worldcentric, at which point, and not before, they can awaken to a deep and authentic concern for Gaia. The primary cure for the ecological crisis is not learning that Gaia is a Web of Life, however true that may be, but learning ways to foster many arduous waves of interior growth, none of which are addressed by most of the new-paradigm approaches (WILBER, 2000b, p. 137-138).

However, seeing as empiricist objections to cognitivist approaches in moral theory undermine the meaning of Wilber's statement, as well as the very significance of thesis of this study, I am now required to address the epistemological matter mentioned previously. In doing so, I am at the same time introducing and justifying the need of the chosen theoretical scaffolding and of the methodological stance and strategy I adopt in this study.

1.3. Collective action as a moral dilemma: beyond moral noncognitivism

That collective action problems manifestly bring out an essentially moral dilemma is just what has been implied all along our opening discussion regarding the limits of the canonical assumption of self-interest to address this sort of social interaction situation. This standpoint is similarly manifest in the way as Garrett Hardin opens his influential *The Tragedy of the Commons* by situating the commons dilemma as typical member a class of problems whose way out would require "a change in human values and ideas of morality" (HARDIN [1968], 1993, p. 11).

On the other hand, that the restricted concepts of knowledge, rationality, and truth that are oriented to the modern empirical sciences undermine out ability to address a moral dilemma from a (noninstrumental) cognitivist perspective can be detected in the way as the same Hardin lines up with Flecher's (1966) relativistic proposition that "the morality of an act in a function of the state of the system at the time it is performed" (ibid, p. 14), and then sees in the moral reasoning a mere rhetoric character, "designed to produce feelings of guilt in noncooperators". Hence, Hardin claims that, in the absence of substantial sanctions, "responsibility is a verbal counterfeit for a substantial quid pro quo. It is an attempt to get something for noting" (p. 16). On this basis, Hardin outlaws any attempt to conjure up a moral conscience in others as "tempting to anyone who wishes to extend his control beyond the legal limits", and goes as far as drawing on Batson's "double bind" situation – hypothesized as being in the genesis of the schizophrenia – in order to denounce "serious pathogenic effects" resulting from any appeal to the moral conscience of the individuals in order to regulate the commons (ibid)[11].

[11] To clarify Hardin's point, as far as communication is concerned, Batson's "double bind' theory for the origin of schizophrenia involves the same sort of contradiction among a verbal and a non verbal communication that Hardin applies in the Tragedy of the Commons to imply an essential skepticism and maliciousness involved in any plea to "conscience" for the matter of regulating the resource use in the absence of substantial sanctions. However, Batson's "double bind" situation involves a strong emotional and *unidirectional* message from a mother toward her infant, who cannot dispel the anxiety brought about by the said contradiction through further communication (TARNAS, 1999, p. 445, cf. BATESON, 1972 and 1979). Clearly, this is not the case of linguistic communication among adult individuals aimed at overcoming their social dilemmas, so that Hardin's view in fact presumes that moral claims are essentially unbinding.

It is not by chance, therefore, that being one of the first to bring a *nonfoundationalist* conception of justification and rationality to the task of understanding the logic of social action, the situation when agents attempt to adopt a rule to govern their interactions configure precisely the sort of case that Habermas has in mind when he claims that instrumental models does not provide a sufficient basis for a *general* theory of rational action. According to him, a general theory of rational action should have the capacity to explain how all sorts of intentional states come about, or at least how they acquire their *content*, *without abandoning the action frame of reference* (like the evolutionary framework does by appealing to the principle of functional fit).

To briefly clarify the epistemological matter, the foundationalist model of justification stands for a possible response to the core problem of epistemology, namely, that any attempt to justify a given statement inferentially gives rise to a *trilemma*, where each new argument that we introduce in support of our conclusions will contain premises that themselves stand in need of justification.

As Heath (2001) explains in bringing about a rather important dialog between Habermas's theory of communicative action and game theory, the only way to break out of this regress is either to use one of the conclusions as a premise (i.e., reason in circle), or simply break off the chain of reasons by making an undefended assumption. The foundationalist model of justification instantiates the second strategy, which is defended on the strength of the proposition that some "basic" beliefs are intrinsically (noninferentially) justified by virtue of their content. Together with the correspondence theory of truth and the ontology of positivism[12], the substantive thesis of foundationalism rests then on the idea that our ordinary *empirical* beliefs escape from the regress of justification (thus working as undefended premises without introducing an element of arbitrariness) by virtue of the fact that they are ultimately grounded in

[12] It should be noticed that the foundationalist model assumes a positivist ontology, according to which "to *be* is to be *perceived*". An alternative ontology would be that of hermeneutics, which puts forward the notion that "to *be* is to be *meaningful*". While the epistemology of positivism is that of the *correspondence* theory of truth, as indicated in the text, the one of hermeneutics is the *coherence* theory of truth (cf. TAILOR, 1985, p. 14-16).

perception – regarded as an inherently veridical form of experience (HEATH, 2001).

In this vein, under the positivist paradigm, cognition in Western psychology came to have a very narrow meaning, referring to the *apprehension of the exterior objects*. As Wilber (2000b, p. 20) puts it, "if the content of consciousness were not some sort of objective-empirical object (a rock, a tree, a car, an organism), then that consciousness was said not to posses cognitive validity". Wilber adds that in the hands of such as Piaget, the meaning of cognition was narrowed even further, to types of mathematical operations, which were claimed to underlie all other developmental lines in all other domains of consciousness (such as emotions, value systems, ideas of morality, self-identity, perspective and role taking, among others.). "At that point", he says, "consciousness as 'cognition' had been reduced to perceiving nothing but the flat and faded surfaces of empirical objects", so that "any awareness that saw something other than the world of scientific materialism was not a true awareness, was not a 'true' cognition" (ibid).

Heath reports that the foundationalist model of justification and conception of rationality has become increasingly discredited in philosophical circles, largely as a consequence of the so-called "linguist turn". However, while the impact of this change has been felt in debates over scientific methodology, it has still had little impact on the working theories of social scientists. According to him, even though many social scientists have abandoned positivism as a methodological stance, the idea that "value judgments" or "moral questions" are rationally undecidable – a position that Habermas denotes as *moral noncognitivism* – is still widely accepted. The hindrance then is that as moral arguments (because of their prescriptive character) cannot be ultimately grounded in any direct experience of the physical world, the foundationalist model severs the internal connection between norms and justifying grounds that constitutes what Habermas brings up as the *rational foundation of normative validity* (HABERMAS, 1993, p. 41). In other words, foundationalism just denies that it is possible to justify normative claims. "As a consequence", Heath says,

"most social theorists simply assume that any agent who acts on the basis of a moral principle, or a social norm, is not rationally justified in doing so" (ibid, p. 2). Also, according to him, this is what underlines the widespread tendency among social scientist to assume that instrumental action is the *only* form of rational action and that norm-governed action must have some kind of nonrational source, such as conditioning, socialization, or habit.

With reference to self-organization in collective action settings, the problem is that as contemporary behavioral economics abandons the *aprioristic* notion that preferences are self-regarding, moral noncognitivism develops into a rather limitative perspective, particularly for the understanding of the role of communication in fostering social arrangements among individuals holding different normative orientations. As already indicated, the point is that foundationalism underpins a purely relativist perspective on morality. In consequence, the likelihood of normative commitment resulting from communication opportunities will depend entirely upon the extent to which the internal values that serve as "input" into moral arguments are shared. These values may vary from pure self-interest to widely shared cultural standards, but the point is that argumentation itself will be incapable of producing any greater level of agreement than that which agents already bring with them (HEATH, 2001, p. 198). Heath notices that such standpoint renders the coherence, rationality, and adaptability of norm-governed systems extremely difficult to explain, since these qualities could not be achieved through any specifically intentional process. As a result, "this makes it tempting to abandon the action frame of reference and supply purely functionalist explanations of these norms" (ibid, p. 2).

The functionalist explanation configures evidently the standpoint of the contemporary evolutionary approach to human sociability. According to this view, rather than a matter open to rational argumentation and decision, agents' interior disposition to stick on social norms "emerge" as an adaptive byproduct of selective pressures, where emotional responses such as fear, shame, and guilt connect rules of genetic and cultural (conformist) transmission (see BOWLES;

GINTIS, 1998, 2001 and 2003; AXELROD, 1986; BOYD; RICHERSON, 1985 and 1990; RICHERSON et al., 2003; HENRICH et al., 2003). This is also clearly the position from where Hardin ([1968] 1993) ascribes to "education" the remarkable task (once reserved to religion) of revealing "to all" the necessity of changing (or in that case preserve) our values and moral conceptions, so as to protect civilization, our planet, of some common-pool resource from "unfitting" conceptions of freedom. The correspondence theory of truth and the foundationalist model of justification provide likewise the epistemological basis under the renewed attention to leaning, as in the new so-called "cognitive institutionalism" (DENZAU; NORTH, 2000; MANTZAVINOS et al., 2004; NORTH, 2005).

With reference to the game-theoretic evolutionary approach to norms, Heath (2007) argues that it assumes (usually tacitly) a sociobiological framework that is *empirically* inappropriate for the analysis of human social behavior. He explains that, differently from the biological version, there is not any satisfactory criterion to read "fitness" in the cultural evolutionary interpretation. As a result, "what counts as 'fitness' in models of cultural reproduction remains a black box; which contains dozens of different factors, all of them interacting in unknown and highly unpredictable ways" (HEATH, 2007, p. 346)[13]. Indeed, in view of the additional work that needs to be done to show that the underlying assumptions of the evolutionary models (both the cultural evolutionary dynamics and the initial conditions) are empirically supported, Alexander (2003) suggests that "one may wonder what has been gained by the evolutionary model". According to him, it would seem that the contributions made by evolutionary game theory in the explanation of social phenomena simply are a proper part of the parent social science – sociology, anthropology, economics, and so on. "If so", he says, "then there is nothing *particular* about evolutionary game theory employed in the explanation, and this means that, contrary to appearances, evolutionary game

[13] Heath indicates that the basic assumption that the *morality meme* is unfit because of the (supposedly low) utility payoffs it generates for its host overlooks the point that utility for the host is not the same as "enhanced fitness" for the *selfish meme*. Yet, such supposition is what justifies the use of the same replicator dynamics to analyze the conditions for the emergence and reproduction of cooperation among humans (as used in biology to analyze population dynamics).

theory is really irrelevant to the given explanation" (ibid, emphasis in the original).

Likewise, by tacitly treating justification as an essentially "monological" process, that is, one involving only the agents' cognitive states and the objects of representations, the foundationalist epistemology has the effect of reducing *all* public practices of justification to either secondary or derivative phenomena (HEATH, 2001, p. 201). In this way, foundationalism basically disregards the hermeneutic dimension, and in turn basically reduces the cognitivist approach to leaning to the method of the empirical sciences. As we shall in the next chapter, the consideration that differences in architectures, goals and resources of different cognitive systems matter at the normative level suggests that "science is not a good model of individual cognition, and individual cognition is not a good model of science" (CUMMINS et al., 2004, p. 320). In addition, by passing also over the developmental ordering of cognitive systems, foundationalism renders mysterious not only the rationality of deontic constraints but also "the persistence throughout history of dogmas, myths, superstitions, and ideologies" based on "flawed belief systems", as Mantzavinos et al. (2004, p. 76) put it. According to them, this phenomenon "calls us to pay as much attention to learning that produces such beliefs as we do to learning that appears to *interpret correctly* the problems confronting humans" (ibid, my italics). As indicated, however, the suggestion here is that we possibly cannot adequately accomplish this task from the foundationalist conception of justification and rationality; and this is particularly so when institutional change and therefore normative agreement and commitment are both in the center of the analysis.

In any case, by overlooking the hermeneutic dimension, i.e., the sphere of "dialogical" justification, the foundationalist view on learning first reduces linguistic communication (which is possibly the most fundamental of human institutions and source of social integration and development) into a "mere" vehicle of information exchange, and then subordinates the intralinguistic objectives of speech acts (i.e., producing *mutual understanding*) to agent's *private* projects or strategies. Accordingly, under the correspondence theory of

truth, communication is hopefully expected to foster the convergence of agents' more of less imperfect mental representations of the circumstances of action, including knowledge about action alternatives, action-outcome linkages, the preference/values of others, and so on (see e.g. MANTZAVINOS et al., 2004; DENZAU; NORTH, 2000; FROHLICH; OPPENHEIMER, 2001; OSTROM, 2003 and 2005). Akin to Hardin's view on the role of "education", such convergence of "mental models" would be then the "ultimate" source of action coordination and social integration, while communication itself is supposed to exercise no effective constraint on the range of action alternatives available to agents. Hence, no matter how much self-absorbed or other-regarding agents' preferences might be, the point is that foundationalism implies that there is no connecting point between cognition and deontic constraints, save perhaps for the mentioned "flawed belief systems", underlying dogmas, myths, superstitions, and ideologies. Hence, it really seems that *it is the very epistemological foundation of the empirical sciences that imparts the "unbinding" character to moral reasoning and verbal promises in models of strategic interaction,* not really the postulation that all individuals are norm-free maximizers of material gains.

On the other hand, having embraced the "linguistic turn" and adopted speech act theory as basis for developing a thoroughly *intersubjectivist* theoretical framework that avoids the pitfalls of both objectivism and subjectivism, Habermas (1993, p. 39) argues that "Only cognitivist basic assumptions can to justice the phenomena and experience of a posttraditional morality that has detached itself from the religious and metaphysical contexts from which it arose". In this way, Habermas (ibid, p. 40) first indicates that the cognitivist interpretation of the mode of validity of normative statements is opposed (a) to the empiricist notion that the force of the moral commands rests on the mere feelings of obligation originating in the internalization of threatened sanctions, and (b) to the noncognitivist notion that the reciprocal interest in the observance of norm can ultimately be traced back to an interest in self-respect.

According him, the empiricist notion that norms obligate only to the extent that they are backed up by well-founded expectations of sanctions neglects

the fundamental intuition that the *noncoercive binging force* of recognized norms is transferred from the validity of a valid norm (as agreed upon through the public practices of "dialogical" justification) to the duty and the act of feeling obligated. "Sanctions", he says, "however much they are internalized, are not constitutive of normative validity; they are *symptoms* of an already felt, and thus antecedent, violation of normatively regulated context of life" (HABERMAS, 1993, p. 42, emphasis in the original). Thus, according to the cognitivist view, "We do not adhere to recognized norms from a sense of duty because they are *imposed* upon us by the threat of sanctions but because we *give* them to ourselves". Also, "It is not because the recognized norms are *certified* by custom and tradition that we observe them from a sense of duty but because we take them to be *justified*" (ibid, emphasis in the original).

Therefore, insofar as one *actually* adheres to a social norm or observe an interaction rule from a sense of duty, the "prosocial emotions" that Bowles and Gintis (2003) talk about must be viewed as affective reactions to violations of *legitimate* expectations, meaning that they *presuppose the recognition of the validity of underlying norms*. In this sense, they are neither "ultimately" emotional responses, to be explained in terms of mindless evolutionary processes or selective pressures, nor can be unmistakably elicited by some idealized apprehension of the circumstances of action – one that appears to interpret correctly the problems confronting humans.

Of course, sense of duty is not the only reason why one may feel compelled to stick on the rule. However, to the extent that the cognitivist interpretation of the moral point of view restores the possibility of a rational foundation of normative statements – this way rescuing "preferences" on deontic constraints, utility or value judgments from pure relativism and subjectivism – it becomes possible to examine norm-governed behavior and normative commitment, in particular, by inquiring not only on the objective circumstances of action, but also on the sociocognitive capacities of the agents in interaction.

To carry out such examination, from the perspective of Graves's Levels of Existence theory, is precisely the meaning of the second working hypothesis –

the one connecting Graves's two family of psychosocial systems to the likelihood that communication might lead to normative commitment in collective action settings, as presented in the earlier section.

To reinstate, the practical significance of this effort proceeds from the fact that the developmental framework imposes behavioral restrictions and describes a rather significant qualitative pattern (i.e., decreasing egocentrism or increasing decentering) that can work as guide to analyze the adequacy of existing and support the development of new institutional arrangements to cope with social dilemmas. Furthermore, while I am *not* going to engage into a *direct* examination of linguistic communication, the proposed framework implies, as Habermas (1990, p. 33) puts it, that "social sciences can become aware of the hermeneutic dimension while remaining faithful to the task of generating theoretical knowledge".

1.4. Research strategy and the Integral Methodological Pluralism (IMP)

As already indicated, this study advances an empirical investigation respecting the usefulness of Graves's constructs to predict the behavior of competently judging subjects embroiled in collective action problems. This is a cross-cultural test in nature, since the test instrument was originally expected to be valid only to speakers of standard American English, and for persons with the equivalent of high school American education (HURLBUT, 1979).

Using an authorized Portuguese translation of the testing instrument (Spiral Dynamics® Discovery), I begin by accessing the Gravesian psychosocial profiles of 322 Brazilian subjects and submitting the collected data to a *confirmatory* Factor Analysis intended to verify the theoretical soundness, and therefore the delineation of Graves's constructs in the sample. Subjects then take part in three well-established experimental protocols designed to the inquiry on collective action problems.

The first experiment, adapted from Ostrom et al. (1994), sets up an appropriation dilemma (common-pool resource) including three institutional

conditions, namelly: (a) open access (baseline), (b) a nonenforceable self-organized joint strategy, resulting from two face-to-face communication opportunities (communication), and (c) the optional use a costly sanctioning mechanism to reinforce the joint strategy (sanctioning). While information about individual appropriation is provided all along the thirty rounds of the experiment, confidentiality is preserved by means of a numeric code. The second experiment is a provision dilemma (public good), based on Poppe and Zwikker (1996) design, where four basic intentional states (fear, greed, solidarity, and sense of duty) are revealed by contrasting participants' expectations regarding the contributions of others with their own decisions whether or not to contribute for the common-pool in nine levels of required contribution. No communication is allowed in the public good experiment. Information about whether the public good shall or not be actually provided in each contribution level is given only at the end of the experiment. The last experiment is the well-known Ultimatum Game in its traditional form (one shot with anonymous coupling).

The prediction power of Graves's constructs is evaluated by means of the statistical associations (binary correlations and multivariate regression analysis) involving both individual and group scores on Graves's Levels of Existence, on the one hand, and the corresponding behavioral responses produced in the three experiments, on the other hand.

Insofar as the instrumental action is concerned, the conceptual scaffolding I apply in this study consist basically in placing Graves's model in the center of the analytical scheme provided by the Institutional Analysis and Development (IAD) framework (OSTROM, 2005) and the so-called "cognitive institutionalism" (DENZAU; NORTH, 2000; MANTZAVINOS et al., 2004; NORTH, 2005). Furthermore, as Graves's detailed analysis of dynamic of change shows that stage transition is in general an arduous and slow process – a feature that underlies in fact the very concept of *stages* that characterizes the paradigm of psychological structuralism – I do not really expect that the participants in the first experiment shall change their interior dispositions or

intentional states in response to moral argumentation during the communication opportunities.

This means that, except for the *developmental* feature attached to the such subjective predispositions, everything happens as in the pure relativistic interpretation, that is, the likelihood of normative commitment resulting from communication opportunities should still depend mostly upon the extent to which the internal values of the participants are shared. The only difference is that such concurrence is now seen as a function of participants' level of psychosocial development – not a function of some "horizontal typology" or incommensurable idiosyncrasies. Therefore, it is precisely in the sense of establishing the validity of the developmental interpretation that the epistemological issue discussed previously matters most.

Accordingly, in the place of the methodological monism brought about by the foundationalist model of justification, a move toward an Integral Approach (of any problem situation) requires a methodological attitude orientated by Integral Methodological Pluralism (IMP). Wilber explains that IMP consists on a set of paradigms, behavioral injunctions, and social practices in which the component terms are understood thus:

> "Integral", in that the pluralism is not a mere eclecticism or grab bag of unrelated paradigms, but a meta-paradigm that weaves together its many threads into an integral tapestry, a unity-in-diversity that slights neither the unity nor the diversity. "Methodological", in that this is a real paradigm or set of actual practices and behavioral injunctions to bring forth an integral territory, not merely a new holistic theory or maps without any territory. And "pluralism" in that there is no one overriding or privileged injunction (other than to be radically all-inclusive) (WILBER, 2006, excerpt B).

The theoretical foundation of Wilber's IMP will be discussed in the next chapter. For now, the introductory point is just to indicate that the IMP conjoins pluralism with integralism, while avoiding at the same time the pitfalls of both objectivism and subjectivism by mean of three heuristic principles: *nonexclusion*, *enactment*, and *enfoldment*. In Wilber's words:

> *Nonexclusion* means that "Everybody is right" – or more technically, that the experiences brought forth by one paradigm cannot legitimately be used to criticize, negate, or exclude the experiences brought forth by other paradigms. The reason that "everybody is right" is called *enactment*, which means that no

> experience is innocent and pregiven, but rather is brought forth or enacted in part by the activity of the subject doing the experiencing. Thus, one activity (or paradigm) will bring forth a particular set of experiences – experiences that are not themselves innocent reflections of the one, true, real, and pregiven world, but rather are cocreated and co-enacted by the paradigm or activity itself, and, accordingly, one paradigm does not give "the correct view" of the world and therefore it cannot be used (as if it did) in order to negate, criticize, or exclude other experiences brought forth by other paradigms [i.e. "Paradigm" refers to the methodologies of enacting new phenomena, not merely to the theories that attempt to explain them.]
> However, if one practice or paradigm includes the essentials of another and then adds further practices – such that it "enfolds" or includes the other – then that paradigm can legitimately be claimed to be more integral, which is the *enfoldment* principle (WILBER, 2006, except D).

This is the general outline of the methodological stance underpinning this move toward an Integral approach of the commons dilemma. Next, I offer an overview of how the study is organized, and address some of its limitations. Further implications of the Integral Theory for the inquiry on collective action problems are discussed in the next chapter.

1.5. Organization and limitations of the study

The study comprises seven chapters, in addition to this Introduction. Chapter two presents the basic elements of the Integral Map, or Integral Operating System (IOS), as often called in analogy with the language of informatics. The chapter outlines the Twenty Tenets of evolution, which configure at one and the same time the ontological and epistemological foundation of Ken Wilber's Integral Metatheory. It also includes a specification of the IMP by showing how the Integral Map is actually able to weave together the major paradigms of methodologies or modes of inquiry in use. I then apply the Integral Map to show how the seemingly incompatible theoretical perspectives that I bring to this study can be reliably put to work together, toward a more balanced and comprehensive approach of the commons dilemma. In the Chapter three, I apply the Integral model to map the partialness in the main perspectives on collective action and the so-called evolution of cooperation. Clare Graves's ECLET conception is presented in the Chapter four. In the

Chapter five I situate the ECLET conception in the framework of the institutional analysis and develop the idea of an Integral Approach to the commons dilemmas, as introduced in the Chapter two. The detailed description of empirical strategy and the refinement of the working hypotheses presented here are developed in the sixth chapter. Chapter seven brings the results and the discussion of our findings. I conclude with the Chapter eight, by discussing the limitations and the affirmative implications ensuing from enough results in support of the thesis of the study.

As an earlier notice respecting the limitations of the study, I must mention that the results are limited to the accuracy of the test instrument, as developed by Hurlbut (1979), in addition to the cultural issues implicated in this attempt to extend its application to Brazilian subjects. The value of such attempt, however, connects with Graves's own proposition respecting the cross-cultural validity of his theoretical constructs. Furthermore, as Hurlbut indicates, the accuracy of the test instrument is proportional both to the test taker's awareness and understanding of his own personality and to his truthful disclosure of that personality. Indeed, the epistemological issue in this chapter helps to put in the proper perspective the fact that attitudes, beliefs, and thought patterns are not accessible for direct measurement and therefore can be measured only indirectly by self-report. Finally, the study is conditioned by the sufficiency of the proposed suitability of Graves's constructs to explain behavior in experimental conditions that are distinct of those from which the constructs originally emerged.

2. THE INTEGRAL VIEW OF KEN WILBER AND THE INTEGRAL OPERATING SYSTEM

2.1. Introduction

Ken Wilber's Integral Metatheory develops from a current of unity and continuity running throughout over twenty-five years of publication. Captained by his magnum opus, *Sex, Ecology, Spirituality*: *The Spirit of Evolution*, first published in 1995 (WILBER, 2000a) and conceived as the first volume of an in progress series entitled *Kosmos Trilogy*[14], the philosophical works of Ken Wilber (e.g. WILBER, 1980, 1981, 1996, 2000b, 2001) are recognized as the primary reference on integral theory.

However, despite the undoubted importance of Wilber's contribution, integral theory is not the work of one man. Prominent author whose work provide support for the integral perspective include Jürgen Habermas (e.g. 1979), Arthur Koestler (e.g. 1989), Jean Gebser ([1949] 1985), Jean Piaget ([1929], 1973), Abraham Maslow (e.g. 1971), Lawrence Kohlberg (e.g. 1976), Clare Graves ([1971] 2005), and Robert Kegan (e.g. 1982, 1994), to name a few.

Also, since the publication of SES (*Sex, Ecology, Spirituality*) and the foundation of the Integral Institute (www.integralinstitute.org), in 1998, Wilber's

[14] Wilber's use of the term "Kosmos", rather than "cosmos", will be clarified in the Section 2.2.

Integral Approach has developed into a lively field of inquiry with numerous contributors from diverse disciplines around the world. Much of this recent work has focused on theoretical and practical applications of Integral Approach to specific disciplines, resulting in new fields like Integral Ecology, Integral Politics, Integral Business, and Integral Life Practice.

In addition, as Riyed (2005) observes in introducing his application of Wilber's framework to advance an Integral Approach to sustainable development, the effort of integration is proceeding under different names in many fields, from public policy (e.g. MICHAELIS, 2003) to system theory (e.g. KAY et al., 1999) to political ecology (e.g. NORGAARD, 1995) to ecological economics (e.g. O'CONNOR, 1994). He points that while integrators and synthesists in these fields may not identify their work with Wilber's contribution, their objectives are often compatible, so that there is a vast body of literature on which to drawn when writing about the integral perspective.

As suggested in the previous chapter, the special interest of exploring collective action dilemmas from the standpoint of Wilber's version of integral theory is that it seeks to integrate the objective and the subjective perspectives using a *developmental* orientation. Starting from the proposition that there are four main epistemological approaches to any problem, namely: *behavioral* (objective), *systemic* (interobjective), *psychological* (subjective), and *cultural* (intersubjective); Wilber's Integral Theory reconstructively describe the process of development/evolution of an undivided, simultaneously ontological and epistemological reality, as a series of *tetra-enacted* emergent modes of being and knowing that have *corresponding* expressions in all those four basic domains of existence/experience. In this way, the metatheoretical scaffolding provided by Integral Theory allows us to address issues of subjectivity that have received inadequate attention in mainstream approaches to social dilemmas, while also providing the theoretical grounding for the developmental aspect of cultural institutions, worldviews, and the forms of socioeconomic organization.

However, as we saw in the previous chapter, cognitivist-developmental approaches to moral theory are at odds with the epistemological underpinning of

the modern empirical sciences (i.e. the correspondence theory of truth and the foundationalist model of justification), since value systems and moral stances cannot be ultimately grounded in any direct experience of the physical world. On the other hand, in the wake of the linguistic turn, the developmentalist interpretation is entirely banned by an extreme postmodernism that actively fight to deconstruct anything that looks like a metanarrative, particularly concerning culture.

In this context, Wilber's philosophy of **universal integralism** emerges partly as a reaction to the restricted concepts of knowledge, rationality, and truth that are oriented to the modern empirical sciences, and partly as a reaction to a paralyzing postmodern intellectual climate according to which pluralistic relativism is paradoxically presented as the only acceptable worldview. The situation Wilber describes as underlying the very emergence of Integral Theory is of great interest in this reassessment of the commons dilemma, as he recognizes in the downsides of both the modern and the postmodern worldviews the expression of a *degenerated* opposition between *pathological agency* (on the side of modernity) and *pathological communion* (on the side of postmodernity) – one that, according to him, hinders the development of more balanced and effective approaches to cope with our present ecological crisis. Furthermore, as we shall see, it is not by chance that Wilber's account stands in remarkable coherence with the cyclical turn between the agency and communion capacities of the self, as built in Graves's scheme of the biopsychosocial development of adult humans.

Therefore, before introducing the tolls of the Integral Operating System (IOS), as often called in analogy with the language of informatics, I offer a brief outline (§ 2.2) of the aspects of this modern-postmodern dilemma in respect of which Integral Theory is presented as a possible "way out". Next (§ 2.3), I summarize the cornerstone of Wilber's Integral view; the so-called Twenty Tenets of evolution, upon which the Integral Operating System (IOS) in turn applies. The compounding elements of the IOS are presented in the section 2.4, together with an illustration of how it supports the epistemological integration in the way of the Integral Methodological Pluralism. In the final section (§ 2.6), I

discuss the implications of the Integral Theory for investigating the commons dilemma.

2.2. The "collapse of the Kosmos" and the "paradox of damage": the epistemological puzzle under the present ecological crisis and the emergence of integral theory

Created by a cross-cultural comparison of most of the known forms of human inquiry, including the enduring insights of premodern, modern, and postmodern sources, Integral Theory endows us with a rather logical, profoundly insightful, and outstanding comprehensive framework designed to coherently approach virtually any human endeavor (including educational problems, business and organizational leadership, environmental and ecological matters, political and military issues, and personal transformation and spirituality) from the plural perspectives of science, morals, aesthetics, Eastern and Western philosophy, and the world's great wisdom traditions.

In order to accomplish this, Ken Wilber evidently does not take on the level of details. Rather, he lays out a number of *orienting generalizations* conceived as "a way to suggest that the world is really one, undivided, whole, and related to itself in every way" (WILBER, 2000a, p. xii). As he explains it:

> My point is that if we take this types of largely-agreed-upon orienting generalizations from various branches of knowledge (from physics to biology to psychology to theology), and if we string these orienting generalizations together, we will arrive at some astonishing and often profound conclusions, conclusions that, as extraordinary as they might be, nonetheless embody nothing more than our already-agreed-upon knowledge. The beads of knowledge are already accepted: it is only necessary to provide the thread to string them into a necklace (WILBER, 2000a, p. 5).

As suggested, the thread Wilber provides entails a *structural* bound linking ontology and epistemology along with the developmental orientation. However, in order to bring about the said orienting generalizations and string them in a coherent fashion, Integral Theory does not simply embrace the contributions of the greatest number of perspectives without first identifying and jettisoning the sources of potential contradictions between them. The guiding

principles of such critique endeavor are just the three heuristic principles of the Integral Methodological Pluralism (IMP): namely, *nonexclusion*, *enactment*, and *enfoldment* (see Section 1.4). In this way, Wilber organizes the development of human knowledge by jettisoning that that, according to his view, impedes for instance the postmodern worldview of enfolding the enduring contributions of modernity, and that that impedes the modern worldview of embracing the enduring contribution of the premodern approach to reality.

Accordingly, as we shall see, in order to restore the meaning of *ascesis* (transformation of consciousness), the qualitative distinctions and value rankings, together with the spiritual understanding of the life and the universe characteristics of the premodern worldview, Integral Theory proceeds in a thorough *postmetaphysical* fashion, so as to harmonize those attributes with the perspectives of both the modern empirical sciences and the postmodern constructivism and contextualism. As well, in order to overcome the modern-postmodern quandary, the philosophy of universal integralism must proceed in a post-postmodern fashion, in order to conciliate postmodern contextualism with the modern objectivism.

To proceed in this fashion, Wilber begins with a specific definition or description of modernity that allows us to see the underpinnings of *both* its positive and negative expressions and results. Hence, he mentions that various scholars, from Max Weber to Jürgen Habermas, have suggested that what specifically defined modernity, as "rich panoply of *still evolving* beliefs which arose with the successes of Western science and political organization in the Europe and North America during the past three centuries" (NORGAARD, 1995, p. 6-7, my italics), was something called "the differentiation of the cultural value spheres", which specifically means the differentiation of the art, morals, and science. As he explains, where previously these spheres tended to be fused, modernity differentiated them and "let each proceed at its own pace, with its own dignity, using its own tools, following its own discoveries, unencumbered by intrusions from other spheres" (WILBER, 2000b, p. 60).

To illustrate the implication of this differentiation, Wilber mentions that, in the Middle Ages, Galileo could *not* freely look through his telescope and report results because art and morals and science were all fused under the Church, and thus the morals of the Church defined what science could – or could not – do. Accordingly, the differentiation of the value spheres means that Galileo *does* could look through his telescope without fear of being charged with heresy and treason. In this way, science was free to pursue his own truths unencumbered by brutal domination from other spheres; and likewise with arts and morals. "Artist could, without fear of punishment, paint nonreligious themes, or even sacrilegious themes, if they wished. And moral theory was likewise free to pursue and inquiry into good life, whether it agreed with Bible or not" (ibid).

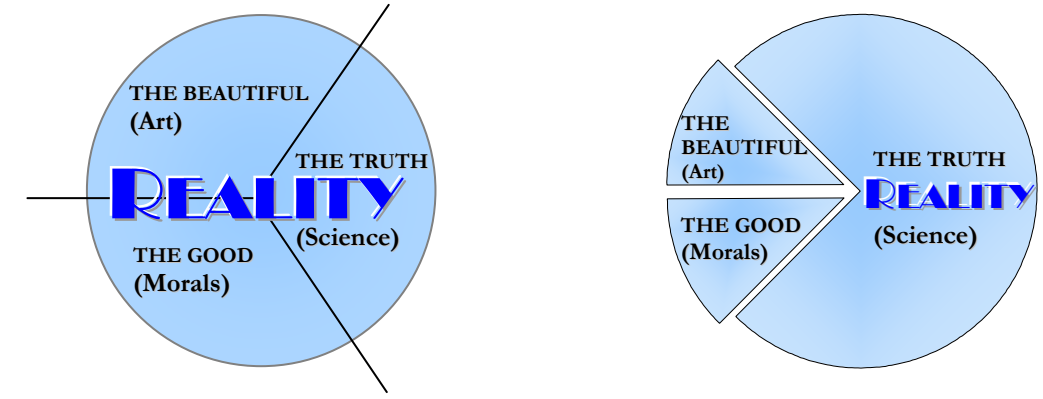
Wilber indicates that this differentiation was in part responsible for the rise of liberal democracy, the end of slavery, the growth of feminism, and the staggering advances in the medical sciences. "For all those reasons and more", he says, "these *differentiations* of modernity have also been referred to as the *dignity* of modernity" (ibid, emphasis in the original). However, according to him, the *disaster of modernity* was that these values spheres did not just peacefully separate, but rather often flew apart completely.

> The wonderful differentiations of modernity went too far into actual dissociation, fragmentation, and alienation. The dignity became disaster. The growth became cancer. As the value sphere began to dissociate, this allowed a powerful and aggressive science to begin to invade and dominate the other spheres, crowding art and morals out of any serious consideration in approaching "reality". Science became scientism – scientific materialism and scientific imperialisms – which soon became the dominant "official" worldview of modernity (WILBER, 2000b, p. 61).

The idea is represented in Figure 2.1. The dignity of modernity is depicted in the panel, (a) where art, morals, and science seize their own share in approaching reality. The panel (b) shows the fragmentation of the value sphere, promoted by the imperialistic encroachment on the part of science respecting the capacity to produce valid knowledge.

Wilber explains that the implication of this shift was a *subtle* reduction of human subjective experience, and actually of all interiority, into

fundamentally empirical and naturalistic terms[15]. As a result, seeing as in a merely empirical-sensory world there is merely empirical-sensory motivation, the scientific monopoly over the access to reality brought about a "*homogenization of motives* for all beings" – a homogenization that lacked, moreover, all *degrees of depth*, since it would have to assign to "the *lowest common denominator* for all beings: namely, sensory pleasure/pain, hedonic happiness". Hence, the disaster of modernity stands for a leveling of moral distinctions that "makes it impossible to see that there are altogether different degrees of moral virtue" (ibid, p. 435).



(a) Differentiation of the value sphere (b) Dissociation and scientism

Source: Author's configuration.

Figure 2.1 – The good and the bad news of modernity.

Wilber notices that this is why the instrumental and interlocking-order theorists (from Locke to Bentham) always had recourse to a hedonistic theory of motivation. Further, he adds that closely related to this ethics of pleasure/pain, in terms of motivation, is the simple survival or self-preservation, so that systems theorists often make self-preservation of the system – autopoiesis – the mono-

[15] As indicated, by *subtle* reductionism Wilber means the modern denial that subjective experiences needed to be explained in any terms *other* than those found in the empirical assessment. On the other hand, as a renewed Epicureanism, *gross* reductionism stands for the further claim that everything could actually be reduced to relations between elementary particles.

motivation that all beings possess. Of course, Wilber acknowledges that this is partially true in an abstract fashion, but the point of interest is that the mono-motive of functional fit is incapable of sustaining qualitative distinctions in the subjective and intersubjective domains (as those of arts and morals) – distinctions that, nonetheless, configure the core of the enduring contributions of the so-called perennial philosophy and virtually all spiritual traditions world over.

This limitation of the modern scientism is exemplarily illustrated in Crabtree's (2006) defense of what he denotes as the "universal egoism" theory. He argues that biologists, sociologists, philosophers and above all, psychologists, have held to the said theory according to which "all apparent altruism is *really* selfishness in disguise" (CRABTREE, 2006, my italics). As he explains:

> People behave altruistically [i.e. morally] for a number of selfish [and ultimately self-preserving] reasons. We are programmed genetically to behave in a way conductive to the sociability of the species. This unconscious specie-instinct is the closest thing we have to true altruism. In nearly every other conscious sense altruism is an illusion. We behave well because social good behavior fires off pleasant neurochemicals in our brains (the pleasure reward), because consciously or unconsciously we want others to see us as good person (the social reward) or to feel good about ourselves (for pride and self-esteem). Altruism is image and illusion (CRABTREE, 2006).

The striking feature of Crabtree's assertion is the way how it exposes so unmistakably the scientific monopolization of the capacity to tell us what is supposed to be *really* real. Accordingly, although in principle there is no contradiction in the fact that a *morally motivated* act may or may not, depending on the situation, either elicit or be accompanied by objective rewards, the point is that the reality of what is exterior (objective) is been candidly placed over the reality of the interior (subjective) motivations. This self-estranged interpretation actually exposes what has been called the *central problem of modernity*: namely, human subjectivity and its relation to the perceived world.

As Wilber (2000a, p. 441) explains, by reducing the subjective sphere to an empirical dimension, systems theorists do not actually explain the fit of the *subject* into the empirically holistic net (see Figure 2.1, panel b). Rather, that move explained away the subject as being merely another *object*. "It erased the very terms it was supposed to explain". As a result, the subject of the moral

reasoning is left *disengaged* and *dangling* with its own "self-defined" and "illusory" moral agency "with no way, in participatory terms, that the *subject* could actually fit into the interlocking world of *objects* it had so seamlessly described" (ibid, emphasis in the original).

One should notice that this epistemological attitude corresponds to the superseding authority attributed to the correspondence theory of truth and the foundationalist (monological) model of justification in approaching reality, as discussed in the previous chapter. Accordingly, Crabtree's account on "universal egoism" actually aligns with the evolutionary game-theoretical approach to human sociability and norm-governed behavior. Yet, it exposes more overtly that the acknowledgement of other-regarding preferences in the game-theoretic analysis of social dilemmas (i.e., analysis that relies only on the instrumental conception of rationality) does not really means that the cooperative alternatives have been admitted in the "game" as entailing the affirmation of whatsoever *moral* aspect concerned in the situation – after all, even if the subject of the cooperative *choice* might think of herself as behaving morally, the "universal egoism" theory is there to remind us that such judgment is no more than image and illusion; no more than a misapprehended byproduct of amoral evolutionary processes. This is why we see today a burst of interest in exploring the coevolutionary (sociobiological) foundations of trust, reciprocity, and reputation, for instance, *as expressions of an univocal instrumental rationality*, in the explanation of the cooperative behavior (e.g. BOWLES; GINTIS, 1998; AXELROD, 1986; BOYD; RICHERSON, 1985 and 1990; RICHERSON et al., 2003; HENRICH et al., 2003; OSTROM, 2003; KURSBAN, 2003; McCABE, 2003).

In this way, as a further consequence that collides frontally with the developmental perspective on morals, the exclusionary objectivistic account evidently does completely away with the cognitive bounds linking the process of *transformation of consciousness* with the discrimination of *different degrees of moral virtue*. In other words, more than denying the "rational foundation of normative validity", as pointed out by Habermas (see Section 1.3), moral

noncognitivism simply throws away the whole significance of *ascesis*, *not* as authoritarian self-denial, but as the arduous process of interior transformation that underpins subjects' ability to exercise *autonomy as self-determination*, and then to engage in *active self-control* on behalf of the collective good, precisely in situations where the collective interest challenge that of the individual.

As Wilber (2000a, p. 438) explains it, "since the weakest-noodle motivation [of pleasure/pain] is *already* present in most adults [as it is present, in fact, in most sentient beings as a true fundamental] – since it is *already* present in awareness, then one *does not have to transform* in order to live *this* "ethical" life: one simply extends one's present state to as many others as possible" (emphasis in the original). Accordingly, as "there is no longer deeper or higher happiness, [but] only wider happiness", the utilitarian ethics consequently makes it "impossible to see that deeper human happiness might reside in a profound transformation of consciousness, and not just in an expanding translation of one's present state" (ibid, p. 439-440).

Such account is basically what Wilber brings up as "the collapse of the Kosmos", where the term "Kosmos", first introduced by the Pythagoreans, is taken back in order to make a distinction between the truly holistic perspective on reality, i.e., on the Kosmos, and the "flatland holism", as he puts it, that underlies the contemporary notion of "cosmos" as something limited to the physical universe. As he explains, the original meaning of Kosmos was "the patterned nature or process of all domains of existence, from matter to math to theos (i.e. the divine domain), and not merely the physical universe, which is usually what both 'cosmos' and 'universe' mean today" (WILBER, 2000a, p. 45). Accordingly, in order to bring forth genuine holism, a truly integral view of the Kosmos, we are required first to distinguish, and then to reliably integrate the enduring contributions or knowledge *validated in the subjective sphere*, using its own tools, following its own discoveries, etc., with the knowledge constructed on the side of the objective sciences. Conversely, by discrediting the methods of inquiry into, the postulations on, and eventually the very reality of the subjective domain (e.g. "altruism is image and illusion"), we get only to flatland holism, to

a collapsed view of Kosmos wherein all qualitative distinctions, degrees of moral virtues, depth of consciousness and meanings are altogether reduced to a mere shadow of mutually interlocking and interpenetrating relations defined exclusively in procedural and action terms[16].

Now, what is of most interest respecting our present subject matter is that, as the modern worldview reduces the rational-ego to a cog in the interlocking order, and bounds the transformation of consciousness from above, it locks the self, i.e., itself, the *subject of awareness*, into a harmful battle between the *agency* and *communion* capacities, *at* the egoic-rational level (WILBER, 2000a, p. 442). As Wilber explains, this is *not* a battle of higher *transformation* toward further reaches of human potentials, since this possibility was just ruled out. Rather, it is a battle of mere *translations*, according to which the existential puzzle (e.g. "How should I/we live?", "What constitutes the good life?") is addressed, not entirely consistently, as a matter of either to follow the autonomous *agency* of the rational-ego in order to generate its own self-assured morals and aspirations, or else connect ego with the wider ground of its shared *communions* in the natural world.

Wilber refers to this battle as the battle between the Ego and Eco, and to the harmful situation it engenders as "the paradox of damage", since it undercuts both the actual emergence of higher levels of awareness, and the development of genuinely holistic approaches to most of the pressing social dilemmas. Of course, if we talk in terms of individuals, a degenerated opposition of the kind Wilber describes as the paradox of damage is quite exceptional – it would configure something like a "syndrome of identity dysfunction"[17]. Normally, such extreme hostility between the agency and communion capacities of the self turns up in the collective sphere, where different groups of people cluster around those camps,

[16] This is of course just the tread of the game-theoretic approach to human sociability.

[17] According to Akhtar (1984), the classic syndrome of identity diffusion entails six clinical features: 1) contradictory character traits, 2) temporal discontinuity in the self, 3) lack of authenticity, 4) feelings of emptiness, 5) gender dysphoria, and 6) inordinate ethnic and moral relativism. This syndrome implies severe character pathology and is distinct from adolescent identity crisis. The conceptualization has a significant bearing on the differential diagnosis and psychotherapy of personality disorders.

depending on a series of factor, above all of developmental nature[18]. Accordingly, in historical terms, standing at first as the two poles of the fundamental Enlightenment paradigm, the mutual discord between the Ego and the Eco remains, in many ways, Wilber says, as "the archbattle of modernity and postmodernity". In this fashion, as typical example of Wilber's orienting generalizations, the Western scientism and cultural imperialism of modernity's disaster (Figure 2.1, panel b) personifies the *pathological agency* of the Ego camp. On the other end of this polarization, the *pathological communion* appears as an extreme postmodernism that not only denounces the drawbacks of the modern times – such as the commodification of life, the brutalities of capitalism, the replacement of quality by quantity, the loss of value and meaning, the fragmentation of the lifeworld, existential dread, polluting industrialization, rampant and vulgar materialism; in short, all that Max Weber summarized as the "disenchantment of the world" – but also actively fight to deconstruct its very dignity, eventually endangering the sphere of values of regressing to a situation of fusion and indissociation.

Again, while the typical diversity and ambiguity of the postmodern thought defies definition, it is nonetheless possible to distinguish what defines postmodernity – at one and the same time as a historical period, an intellectual movement, an ideology, and a body of theory – so as to recognize its enduring contributions and separate them from its jettisonable shortcomings. Accordingly, to briefly situate, postmodernists in general argue that the assumed universal appeal of Western values and truth is inconsistent with the evidence of cultural diversity, the social construction of reality, and the crucial role of context and perspectives in knowledge generation. As Riyed (2005, p. 49) summarizes it, "Postmodernists therefore emphasize the relativity of truth, the reality of cultural diversity and the centrality of interpretation". He notices that while the degree to which postmodern theorists embrace the relativity of truth varies, they in general

[18] At this point, I must remind the reader that the cyclical turn between the agency and communion capacities of the self is but of the essence of Graves's model of adult personality systems development. As we shall see in a moment, according to Integral Theory, the permanent tension between these two normally equilibrating capacities configures the driving force underlying the evolution of the whole Kosmos, from matter to mind to soul to spirit.

deplore the modern emphasis on hierarchy and Western sense of superiority as self-congratulatory, oppressive, and marginalizing. Accordingly, the grand theories or metanarratives stressing the successes of the modern Western institutions and mindset in giving rise to unprecedented scientific knowledge, technological progress, and material wealth, configure the primary target of the postmodern attacks.

According to Wilber (2000a, p. x), the enduring contributions of the postmodern critique comprise important truths such as constructivism (the world is not only a perception but also an interpretation), contextualism (all truths are context-dependent, and contexts are boundless), and integral-aperspectivism (no context is finally privileged). Constituting the very warranty of Integral Theory, these central contributions of the postmodern thought are fully incorporated in the two first principles of the Integral Methodological Pluralism (IMP): namely, *nonexclusion* and *enactment* (see Section 1.4). On the other hand, by fighting quite actively to deconstruct the very notions of development and hierarchy, extreme pluralistic relativism has carried out even further the leveling of the qualitative distinctions, particularly regarding morals, so that the postmodern view entirely rejects the fundamental principle of *enfoldment*, and does not recognizes universal values whatsoever.

As Wilber (2000a, p. xi) explains, having "heroically developed" beyond rigid universal formalism and oppressive schemes of modernity, postmodernists "became suspicious of any universal at all". In addition, having commonly cut the plural perspectives free from *any* grounding in objective reality or universality, the postmodern thought became paralyzed by radical relativism and could go no further. As various theorists have indicated (e.g. LYOTARD, 1984; KAHN, 1999; WILBER, 2000a, 2002 and 2006a), the indiscriminate embrace (i.e. communion) of the postmodern perspective engenders a number of theoretical contractions. In particular, Wilber reports on postmodernists' paradoxical elevation of plural relativism as the only acceptable worldview as an awful state of affairs that has once completely invaded academia in general and cultural studies in particular. "Even the alternative colleges and institutes were

[at that time] specking postmodernese with an authoritarian thunder" (WILBER, 2000a, p. ix; see also WILBER, 2002). Wilber illustrates the theoretical inconsistence in the extreme relativist attitude by pointing that at the same time as postmodernists claim that there are no universal truths and that all truth is culturally situated, they affirm *this* very truth as an universal, i.e., as true for all cultures. Likewise, at the same time as they claim there are no transcendental truths, they render their own pronouncements transcendental to specific contexts. Also, the typical postmodern claim that *all* hierarchies or value ranking are oppressive and marginalizing is itself a value judgment that is been ranked above whatever view that entail value ranking, and that underpins, moreover, deconstructionists' vigorous struggle to marginalize them (ibid, p. x).

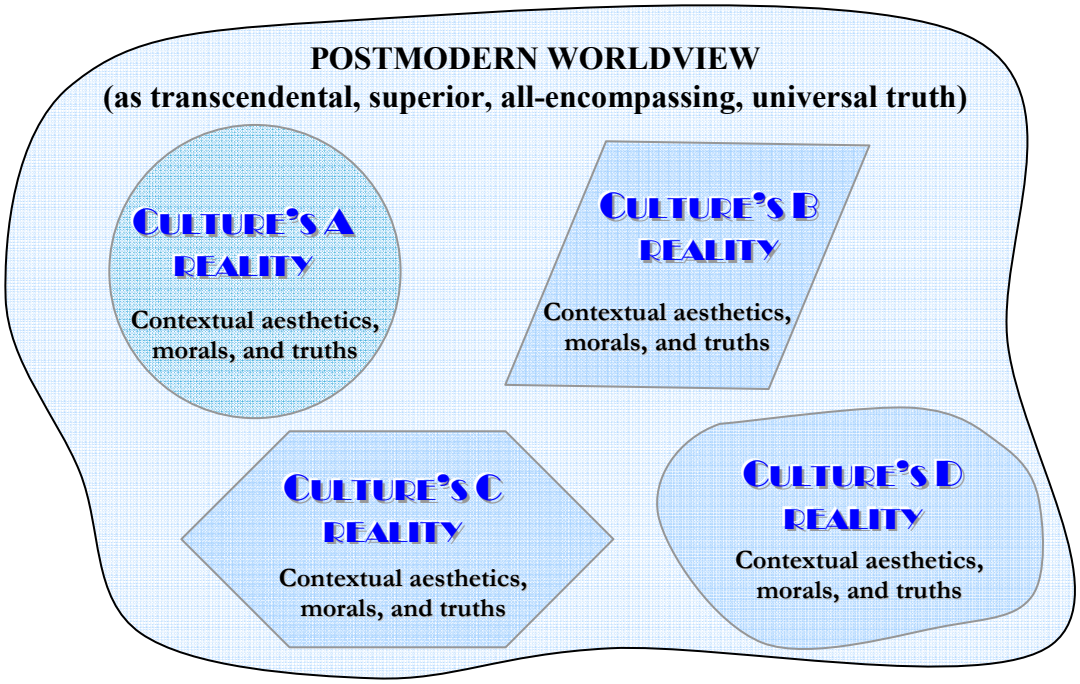
I attempt to illustrate this paradoxical situation as in Figure 2.2. On the one hand, the plural and supposedly incommensurable *intersubjective* realities appear as mere contextual aesthetics, morals, and truths, which are placed fully under the jurisdiction of the sweeping notion of "culture". On the other hand, the contradictory stand of such construction is represented by the transcendental position of the postmodern worldviews, represented by the outer bound, since its universalistic assertions actually place over those of whatever alternative worldviews.

Now, the suggestion conveyed by Wilber's "paradox of damage" is that the upper bound on interior growth, which sustained by *both* the modern and the postmodern worldviews, prevents them of being ever synthesized in a higher union. As he explains, without further transformation (self-transcendence), the Ego and the Eco turn around each other as completely and mutually incompatible perspectives, so that the *more* there is of one, the *less* there must be of the other. Hence, the more the Ego is emphasized as the true ground of a rational morality and an ethical will *not found in nature*, then the more the Eco is devaluated, the more it comes to be perceived as a source of *heteronomy* ("other-direction") opposed to the rational autonomy of the ego-subject. After all, as the "universal egoism" theory reveals so insensitively, nothing in the system of nature appear to seek universal compassion, or mutual understanding, or ethical restraint.

Therefore, to the extent that rational-ego is able to avoid collapsing in tandem with the Kosmos, its task becomes to free itself "from the nasty, short, and brutish net of nature", so as "to establish its agency over the communions of the great and grinding system", as well as over the cultural communions and herd mentalities. In short, the task to the Ego is seen as to *disengage from the Eco*, so that the more the agency, and the less the communion, the freer and the more whole the self would be. Ideally, the purest freedom would be the Pure Reason (of Kant) or the Pure Subject (of Fichte), transcending nature and objects altogether to find the ultimate ethical clarity. However, when confronted with the paradox of how Reason could in effect oppose itself to its own self-presenting bent as just another strand in the great natural web, Reason became merely instrumental. This is why the Ego camp has offered instrumental or calculative reason as a way to orient the individual ego to the great empirical interlocking order[19].

On the other side, following Wilber's account, the Eco camp accuses the Ego of being not just establishing its agency but actually severing and repressing its connection with nature – both external nature and, more importantly, internal nature (sensual, desirous, sexual, vital). Accordingly, for the Eco camp, the self-isolated agency, however important, had sealed off the rich networks of unions and communions with nature and other cultures, and thus the more reenchanted the world, the better. In this vein, for the Eco camp, moral freedom would be found in the *wholeness*, through the reintegration with nature and social community. Thus, the more the communion, and the less the agency, the freer and the more whole the self would be. The purest wholeness would be the Pure Eco, or the Pure Immanence (of Spinoza), free of separate egos altogether. Accordingly, as Wilber puts it, in opposition to Ego camp's instrumental reason, the Eco camp offered but *instrumental feelings*, "feelings that allowed us to fit in the Web of Life as a part and not to try to dominate on the whole".

[19] In this regard, Taylor (1989) observes that a new and important *spiritual* meaning was given to the very instrumental stance of Reason toward the world. As stated by no less than John Locke, quoted by Taylor (ibid, p. 243), "Instrumental rationality, properly conducted, is of the essence of our service to God. This is how we participate in God's purposes. Not through instinct, like the animals, but trough conscious calculation, we take our place in the whole".



Source: Author's configuration

Figure 2.2 – The postmodern worldview as a contradictory transcendental universalistic stance that denies transcendental universalism in favor of exclusionary pluralistic relativism.

As indicated, quite apart from the much and indeed complementary truths conveyed by both sides, the point is that, according to Integral Theory, without genuine self-transcendence toward higher levels of awareness the two partial views cannot be superseded.

In this context, seeing as, in the way of these key orienting generalizations, both Wilber and Graves see in the development of the individuals an *ontological recapitulation* of the phylogenic path of humankind (e.g. from premodern to modern to postmodern perspectives), the suggestion here is clearly that such blocked opposition between agency and communion configure a likely and rather important obstacle when people attempt to achieve agreement on how they should regulate their interactions in collective action

settings[20]. As well, it not by chance that the modern-postmodern quandary is clearly recognizable in the current academic debate on how internal values, belief systems, social institutions, and cultural worldviews shape the logic of the social interactions, how they affect the economic performance and the standards of living, and how all these notions and displays are supposed to influence human well-being (cf. e.g. HARRISON; HUNTINGTON, 2000; MANTZAVINOS et al., 2004; NORTH, 2005).

This is why, on the one hand, not just the developmental orientation but chiefly the admission of *higher* levels of awareness (beyond the rational-ego) occupies a central position in the thesis of this study. Likewise, seeing as the *meanings* of the moral contents embedded in the developmental sequence from preconventional to conventional to postconventional stages of moral reasoning, just like those related to further reaches of human potentials are altogether inaccessible from the collapsed perspective that underlies the "universal egoism" theory, for instance, the defense of the thesis of this study cannot go without the great epistemological integration and genuine paradigm shift resulting from Ken Wilber's Integral view. In the next section, I outline the structural foundation of Integral Theory, prior to present the tools for processing the Kosmos, as reunited in the Integral Operating System (IOS).

2.3. The pattern that connects and the validity claim of mysticism: toward the reenchantment of the world

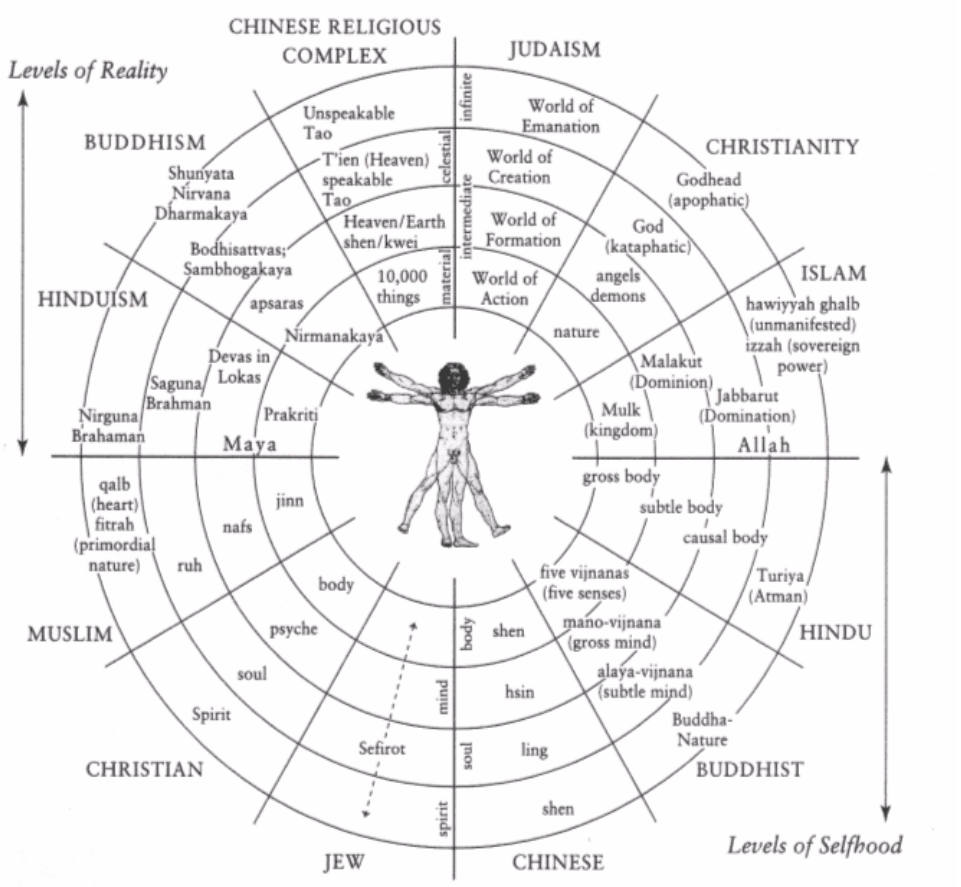
In his search for orienting generalizations that could underpin universal integralism, and therefore allow the important truths of premodern, modern, and postmodern worldviews, Wilber started by studying and comparing hundreds of examples of scientific, spiritual, philosophical, and traditional forms of human knowledge. He concluded that, with the exception of extreme postmodernism, all

[20] This same sort of onto-phylogenetic parallel is sustained by Habermas, as he attempts to show that the stages in Lawrence Kohlberg's model of the development of sociocognitive and moral reasoning are *isomorphic* to the stages that occur in his own phylogenetic account of communicative action (cf. HABERMAS, 1990). See also Simon (1962).

of this knowledge conveyed some hierarchical understanding of reality, or could be readily translated in hierarchical terms (cf. WILBER, 2000a, p. 23-29). As we saw, even postmodernists cannot escape the notion of hierarchy they seek to deconstruct, since they obviously value the postmodern perspective more highly than other perspectives, such as modernism.

As indicated, Integral Theory then proceeds by discriminating what should be kept and what should be jettisoned along the historic development of the worldviews (i.e. development over time, as they actually happened), in order to arrange the enduring contribution of the premodern, modern, and postmodern approaches to reality *as a great hierarchical process of increasing inclusiveness*. This is why the notion of *nested* hierarchy is central to the Integral view of Ken Wilber, which in this regard basically patches up the Whiteheadean theme known as "development is envelopment".

In this way, with reference to the premodern (ancient) worldviews, the notion of hierarchy is manifestly a central one, permeating the whole organization of social life, often in authoritarian terms. On the level of big explanatory models of reality, Wilber points that the *common core* of the world's great spiritual traditions (as built in the so-called "perennial philosophy") entails a view of reality as composed of various *levels of existence*, generally expressed as a "Great Chain of Being". Wilber indicates that the traditional Great Chain of Being is usually given as something like: matter to body to mind to soul to spirit. Drawing on Huston Smith's (1976) compilation, Wilber suggests that interpretative schemes similar to the Great Chain can be found in most of the wisdom traditions of the premodern world and suitably cohered as in Figure 2.3.



Source: Wilber (2001, p. 68).

Figure 2.3 – The great chain in various wisdom traditions, compiled by Huston Smith.

Regarding the nature of the relationship between the levels of reality (upper half of the figure) and the levels of selfhood (lower half), the traditional interpretation of the Great Chain is clearly dualistic. Also, concerning the relationship between the upper and the lower levels of both the reality (above) and the selfhood (bellow), the traditional interpretation is without doubt metaphysical. Indeed, in most of the premodern societies such metaphysical account of the Great Chain typically used to, as it still continue to underpin sources and markers of authority. As we shall see, the Integral Operating System (IOS) configures the conceptual scaffolding that allows Wilber to honor fully the *spiritual* and *mystical* contents of the Great Chain of Being within a thoroughly

postmetaphysical explanation – one that honors as well the contemporary sources of individual liberties (as part of the enduring contributions and dignities of both modernity and postmodernity).

As for the contemporary exclusionary objectivistic (i.e., collapsed, or flatland) approach to reality, following a temporary phase in complexity studies, when the influence of chaos theory challenged the notion of hierarchy by fostering a view of complex systems as consisting of components which were all equally important, more recently it has become clear that a view which underplays hierarchy also tends to underplay the fact that complex systems have *structure*, that it, that they are not homogenous things (CILLIERS, 2005). Accordingly, contemporary systems theorists widely recognize that hierarchy is not an accidental feature of complex systems, but an essential one. Naturally, as already implied since Simon's (1962) "near decomposability" hypothesis[21], and Koestler's (1973) concept of "holon" (a structure or organization that being a *whole* in one context is simultaneously a *part* in another), complex systems are *not* viewed as strictly hierarchical, so that component entities may be part of more than one higher-level entity, and entities at multiple levels of organization may interact (cf. e.g. ARTHUR et al., 1997, p. 6).

As cross-cutting or nested structures, the hierarchical order of complex systems does not entail the rigid, oppressive, and authoritarian character normally associated to the popular understanding of hierarchy, which is the type of hierarchy that postmodernists criticizes, normally with good reasons. Accordingly, as Arthur et al. (op. cit.) explain, in cross-cutting hierarchies reciprocal causation operates between different levels of organization, thus accounting for disperse interactions with no global controller, continual adaptation, perpetual novelty, and out-of-equilibrium dynamics.

In other words, as Wilber (2000a, p. 432) puts it, the collapsed and disqualified world of interlocking order is *still a hierarchy*, still a *holarchy* – i.e.,

[21] Simon's near decomposability hypothesis holds that cross-cutting connections or interactions among subsystems are weak, but not negligible. He then claims that at least some complex systems can be approached as nearly decomposable systems, so that approximations in hierarchical terms would be close enough to the truth to enable a proper understanding (cf. SIMON, 1962).

a hierarchy that has holons as its component entities. Its vital limitation is that such "self-organizing open hierarchical order" (SOHO), to use Koestler's (1973) terminology, has been estranged from the very *subject* of the self-organization, and then plays out strictly and only in the *horizontal* plane. Hence, as we saw with reference to the development of human cognition and reasoning, any vertical move (up or down) has been entirely disqualified. On the one hand, any *prescientific* (and noninstrumental) style of learning and approaching reality is swiftly downgraded as sources of superstition, ideologies, and other "flawed belief systems" (c.f., e.g. MANTZAVINOS et al., 2004, p. 76). On the other hand, as Wilber (2000a, p. 432) notices, any attempt to found knowledge in perceptions, insights, and discernments associated to *higher* levels of awareness is summarily denounced and often stigmatized as the *sin of pride*. After all, the exclusionary objectivistic account does not indeed require anybody to rise above anything in order to have access to all truth that, as presumed, is fit to know. As Wilber (ibid, p. 433) puts it, one is encouraged just to look more carefully at the empirical world that is more or less available. "No arduous *transformations*, but simply more precise *translation* (mapping of the empirically accessible)". However,

> The whole point in that once the weight of the Univocal One[22] is lifted from the shoulder of awareness, the Big Three[23] jump instantly back into focus, and interior depths once *forbidden of serious discussion* – depths forbidden with the shuddering collapse of the Kosmos – now unfurl before mind's inward eye: the surface are not all, she shadows hide something else. The appearances don't just reveal, they conceal: something other is going on. There are nutritional supplements that this thin soup of shiny surfaces does not contain, secrets not apparent to the senses, items that are more… interesting (WILBER, 2000a, p. 430, emphasis in the original).

To see this clearly, the reader may just weigh her or his own moral and possibly aesthetic concerns vis-à-vis the contemporary social chaos against the deflated perspective that comes out from the "universal egoism" theory, for instance. Once again, the affirmation of the reality disclosed before the *inward*

[22] Wilber uses the expression "Univocal One" to denote the dominium of scientism and scientific materialism, i.e., the *disaster* of modernity, as shown in Figure 2.1, panel (b).

[23] "The Big Three" is Wilber's short for the modern differentiation of the sphere of values, as in Figure 2.1, panel (a).

eye does *not* mean that the correlations between morally motivated behaviors and the pleasant neurochemicals in our brains and social rewards are inexistent. After all, as Wilber (ibid, p. 433) puts it, "every event in the real Kosmos does indeed have its [objective] correlate, its empirical and sensorimotor footprint (its exterior form), and some sort of *evidence* for the empirical and positivistic course could *always* be supplied" (emphasis in the original). The whole point is then that the attempt to restrict valid assessment about reality, and eventually reality itself, to the description of its exterior correlates (no matter how much complexly intermingled and dynamically animated) configures in fact an awful epistemological mistake – one whose denunciation embodies the very *dignity of postmodernity*. As Wilber notices, even postmodern poststructuralists are quite certain that truth is not merely a clean game of positivistic and empirical mapping. As he puts it, "The surfaces are shot. The flat and faded system is anemic to the core. This thin soup is causing malnutrition: epistemological, ontological, motivational" (ibid, p. 430).

Hence, to return to the matter at hand, with reference to the notion of hierarchy, "what distinguishes Wilber's epistemology and ontology from that of system theory is its application not only to exterior observable aspects of systems, but also to the subjective interior structures through which humans interpret reality" (RIYED, 2005, p. 53). Furthermore, apart from the extraordinary, profound, and sometimes also seemingly arcane conclusions Wilber is able to take out from this integration of the interior and the exterior realities along with the SOHO structure, such extension of the notion of hierarchy to the subjective sphere it is not, in itself, particularly radical. Indeed, as soon as we are able to break from the collapsed ontology, so as to recognize the knowledge brought forth by the methodologies designed to approach the subjective domain, one discovers the same nested hierarchical foundation in the very heart of the paradigm of psychological structuralism. In particular, the principle of hierarchical inclusiveness rests just in the core of Graves's model of biopsychosocial development of adult humans. The same is true with regard to Habermas's account on the genesis of discourse ethics and communication action

(cf. HABERMAS, 1990). Finally, the same principle appears as well in Cummins's et al. (2004, p. 294) proposition that "there is a *vertical* epistemic ordering of [cognitive] systems"[24]. To be sure, according to them, the relevant vertical architecture to describe cognitive systems "seems to be a partial constituency hierarchy, that is, a hierarchy ordered by partial constituency, or *part-whole* (mereological) relations", wherein "architectures at one level are components of architectures at the levels above" (ibid, my italics). This is of course precisely Koestler's concept of holarchy. Furthermore, Cummins et al central proposition that "to properly explain a system's cognitive capacities, you must refer to the stratum which defines the appropriate norm [of right reason]" (CUMMINS et al., p. 287) is in full agreement with the implications ensuing from Graves's summary statement (section 1.2), and therefore also with the fundamental principle that motivates this empirical research on how different sociocognitive strata, configuring different value systems and worldviews, may matter in explaining behavior in collective action settings[25].

As discussed, the postmodern denial of such interior structuralism results from a failure to really escape from the collapsed *ontology* brought about along the development of the modern sciences, for insofar as it is tacitly held, anything that cannot be grounded in the empirical world, such as value judgments or moral issues, turns out to be mere relative and incommensurate. Hence, to reiterate, postmodern deconstructionist objections notwithstanding, what there is of really unique in Wilber's Integral Approach is neither its recognition of the part-whole relations built in the architecture of complex systems, nor its extension of such

[24] The authors' central thesis is that differences in architectures, goals and resources of different cognitive systems matter at the *normative* level, as "they constrain what principles of normative epistemology [i.e. rules of right reason] can be used to describe and prescribe their behavior" (CUMMINS et al., 2004, p. 287).

[25] Cummins and coworkers actually limit the discussion of differences in architecture, resources and goals supporting their proposition that the science and individual cognition configure *different* cognitive strata, so that "science is not a good model of individual cognition, and individual cognition is not a good model of science". Yet, the authors explicitly suggest that a series of *substrata* may be similarly characterized *within* individuals or brains, as well as along the historical development of science (cf. CUMMINS et al., 2004, p. 320). My specific suggestion (which is just that of Graves himself) is that Graves's scheme provides a model of such substrata (as levels or stages of psychological development) when defined from the perspective of value systems – that is, when defined from the perspective that matters to approach individuals' dissimilar responses to controlled incentive structures in social dilemma situations.

structure to the subjective domain, but rather the sharp insightful appeal with which the Integral Operating System strings these already-agreed-upon beads of knowledge together, so as to engender a series of path-breaking and indeed profound conclusions that emerge from a truly integral view of the undivided (simultaneously ontological and epistemological) Kosmos. As we shall see, such conclusions bring forth nothing less than a reliable answer to the world-knot represented by the so-called "mind-body problem" – an answer that is neither reductionist nor dualist, but genuinely *nondual*. Furthermore, such proposed solution brings forth a validly claim on behalf of mysticism. Before exploring these issues, however, I will provide a brief outline of Integral Theory's Twenty Tenets of Evolution, so as to establish the structural foundation of Wilber's Integral view of the Kosmos.

Accordingly, in Wilber's view, the twenty tenets of evolution describe the "pattern that connects" all domains of existence into a pluridimensional, multilevel (transfinite), interconnected Kosmos, or *Great Holarchy of Being and Knowing*. While he has never systematically listed the said twenty tenets, I will follow Reynolds's (2004, p. 226) numbering scheme in order to present them concisely. In this fashion, Integral Theory puts forward the following tenets (or fundamental conclusion) describing an essentially *creative* evolution and structuration of the Kosmos (which is quite more than the system theoretic notion of spontaneous self-organization):

1. Reality as a whole is not composed of things or processes but of holons (wholes that are part of other wholes);
2. Holons display four fundamental capacities:
   a. self-preservation (or agency);
   b. self-adaptation (or communion);
   c. self-transcendence or self-transformation (or Eros);
   d. self-dissolution (or Thanatos);
3. Holons emerge;
4. Holons emerge holarchically;
5. Each emergent holon transcends *and* includes *all* its predecessors;

6. The lower holon (in the holarchy) sets the *possibility* of the higher, while the higher sets the *probabilities* of the lower;
7. The number of levels that a holarchy comprises determines whether it is "shallow" or "deep;" and the number of holons on any given level we shall call its "span;"
8. Each successive level of evolution produces *greater depth and less span*;
9. Addition 1: The greater the depth of a holon, the greater its degree of *consciousness*;
10. If one destroys any holon, one will destroy all of the holon above (in the holarchy) but none of the holons bellow it;
11. Holarchies coevolve (or "tetra-evolve", actually);
12. The *micro* is in relational exchange with the *macro* at all levels of its depth;
13. Evolution has directionality;
14. Evolution has increasing complexity;
15. Evolution has increasing differentiation/integration;
16. Evolution has increasing organization/structuration;
17. Evolution has increasing relative autonomy;
18. Evolution has increasing the telos or intentionality;
19. Addition 2: Every holon issues an IOU (Incomplete or Uncertain) to the Kosmos;
20. Addition 3: All IOUs are redeemed in Emptiness.

Now, in order to gather the fruits the twenty tenets may bestow to our understanding, Wilber (2000a, p. 43) argues that the principle of holarchy "undercuts the traditional argument between atomism (all things are fundamentally isolated and individual wholes that interact only by chance) and wholism (all things are merely strands or parts of the higher web or whole)". Indeed, what the first tenet radically means is that "both of those views are absolutely incorrect", for "there are no wholes, and there are no parts. There are only whole/parts" (i.e., holons). Moreover, Wilber argues that this approach "also undercuts the argument between the materialist and idealist camps". In this regard, the first tenet means that "reality isn't composed of quarks, or

bootstrapping hadrons, or subatomic exchange; but neither is it composed of ideas, symbols, or thoughts. It is composed of holons" (ibid).

In this fashion, holons define "event horizons" or phenomenological worldspaces wherein ontology and epistemology are inseparable. By taking on the general proposition that the within of things (holons) is *consciousness* and the without of things is *form*, or that the within of things is *depth* and the without is *surface*, Integral Theory then holds that "*all surfaces are surfaces of depth, which means, all forms are forms of consciousness*". Hence, all the way through the levels of the holarchy, "*the spectrum of evolution is a spectrum of consciousness*" (WILBER, 2000a, p. 117)[26]. However, in moving to a still higher level of abstraction, Wilber suggests that the Kosmos is ultimately composed of *perspectives* – not feelings, not awareness, not perceptions, not consciousness, for all of those are *always already* perspectives (WILBER, 2006, excerpt G). In this way, Wilber conciliates the notion of hierarchy with the postmodern constructivism and contextualism, for holons do not simply reflect a pregiven world. Rather, "according to their capacity", Wilber explains, "holons select, organize, give form to, the multitude of stimuli cascading around them"[27]. In short, holons coevolve (eleventh tenet).

As well, by emphasizing "the indefiniteness of holarchy, its openness, its dizzifyingly nested nature" (nineteenth tenet), the integral view agrees also with postmodern aperspectivism (no privileged context). As Wilber puts it, "one cannot point to any thing, to any holon, and say it is just *that* and nothing else, because every holon is simultaneously a superholon and a subholon: it is composed of holons and composes holons (across both space and time); [so that] noting is ever simply present" (WILBER, 2000a, p. 45).

The IOU principle, then, "represent the tension and instability (incomplete or inconsistent) *inherent* in *all* manifest holons by virtue of the fact

[26] Wilber mentions that Spinoza, Leibniz, Whitehead, Schopenhauer, Aurobindo, Schelling, and Radhakrishnan are just a few of the major theorist who have explicitly recognized that the interiority of individual holons is in essence the same as consciousness.

[27] With reference to the development of human cognition (narrowly defined), Wilber points that such is precisely the basis of the Piagetean scheme.

that that they are all whole/parts" (ibid, p. 531, emphasis in the original)[28]. Wilber indicates that another way to say *exactly* the same thing it to say that all holons are agency-in-communion (wholes that are parts of other wholes). Hence, "As *agency* (as wholeness), a holon must seek consistency in order to endure. But as communion (as partness), it must seek union with other holons in order to be more complete). In this way, the IOU principle implies that, at any given or *same* level in the holarchy, the more agency, the less communion and vice versa, and neither side can finally win without destroying the holon. As Wilber (ibid, p. 49) puts it, "Both of this capacities or tendencies are absolutely crucial and equally important; an excess of either will kill the holon immediately (i.e., destroy its identifying patterns); even an immoderate imbalance will lead to structural deformity".

Clearly, here is the structural foundation of Wilber's account on the paradox of damage. In this way, it concurrently shows that, on the one hand, the Ego and the Eco camps are trapped into a *horizontal* battle, that is, that both camps share a same "event horizon" (which is just that of the collapsed Kosmos), as it shows, on the other hand, that such battle cannot the won by any side. Accordingly, as it extends to the form of pathological opposition it may cause the regression of the human holon toward lower levels of development (if not the whole destruction of the human holon). On the other hand, the harmonization of these two capacities requires an upward move toward a higher horizon of events, coenacted at higher phenomenological worldspace or level of awareness.

Furthermore, by fully recognizing the all agency is always agency-in-communion, Wilber points to the interdependence of microevolution and macroevolution (twelfth tenet), therefore making it clear that a distinction needs to be made between the individual (singular) and the social (collective or plural) aspects of holons. In discussing this distinction, Wilber puts forward a path-breaking challenge to the virtually undisputed understanding according to which

[28] Wilber remind us that, in mathematics, the IOU principle shows up as Tarski's Theorem and Gödel's Incompleteness Theorem, both of which are taken to mean that in any sufficiently developed mathematical system (mathematical holon), the holon can be either complete or consistent, but not both (WILBER, 2000a, p. 529).

"man is part of nature". This contemporary dogma is of course intimately connected to the collapsed view of the Kosmos, so that humans are reduced to being just another strand in the great "web of life". Wilber explains that a central error of this approach is a confusion and conflation of the individual and the social aspects of holons, and that this is mostly because the flatland approach leads scholars to construct hierarchies which almost always confuse the micro (individual) and macro (social) dimensions of holons.

To provide an example, Wilber (2000a, p. 88-90) draws on Karl Popper's biospheric holarchy, presented in the Figure 2.4. The figure indicates that Popper counts individual and populations as being at the same logical (and existential) type, and thus arrangeable one above the other, as different "levels" on the same scale. He regards, for instance, the *population* of unicellular organism as a level (level 8) higher than the level of *individual* cells and unicellular organisms (indicated as level 7). The same holds for levels 10 and 11 (Figure 2.4). As Wilber explains, this is quite incorrect, for in any holarchical sequence, either individual or social, if we destroy any level we destroy *all* the levels *above* it and *none* of the levels *below* it (tenth tenet). Thus, a higher-level holon is in part composed of its lower-level holons, and if we destroy any lower-level, we will also destroy any levels above because we have taken away some of their component parts. But these parts themselves existed before the whole emerged and therefore can, in general terms, exist without it. Accordingly, Wilber points that level 8 and level 7, in Figure 2.4, are not different levels, as Popper imagines, but the individual and the social aspects of the *same* level. If one were really lower, then it could exist without the higher. However, as Wilber (ibid) puts it, "you cannot have cells without a society (population) of other cells, and you cannot have a population of cells with no cells at all". Actually, the nature of the relationship between the individual and the social is one of mutual interdependence (the coevolution of micro and macro), where the individual holon is a *participant* or a *member*, not a component part, of the social holon. In other words, Integral Theory reminds us that nothing exists alone without an

environment of similar (eleventh tenet), which is just another way of saying the all agency is agency-in-communion.

12) Level of ecosystems (total biosphere)
11) Level of populations of metazoa and plants
10) Level of metazoa and multicellular plants
9) Level of tissues and organs
8) Level of populations of unicellular organism
7) Level of cells and of unicellular organism
6) Level of organelles (and perhaps viruses)
5) Liquids and solids (crystals)
4) Molecules
3) Atoms
2) Elementary particles
1) Sub-elementary particles
0) Unknown: sub-sub-elementary particles?

Source: Wilber (2000a, p. 89).

Figure 2.4 – Popper's Holarchy (Popper and Eccles, 1983, p. 17).

Wilber indicates how Popper's holarchy should be corrected by asking us to examine what would or would not be destroyed if we theoretically disappeared with the biosphere of total ecosystem. "Looking at Popper's diagram", he says, "we find that if we destroyed the biosphere (if a thermonuclear holocaust destroyed all life forms on earth), we would destroy all the levels drown to and including his level 6, *but no lower*. That means that the biological ecosystem *starts* at his level 6, and not, as he lists it, as level 12" (ibid, p. 98-90). Wilber points that such crucial error appears in the ontology of many ecotheorists as well. As example, he quotes the ecofeminist Julia Russell, when she says:

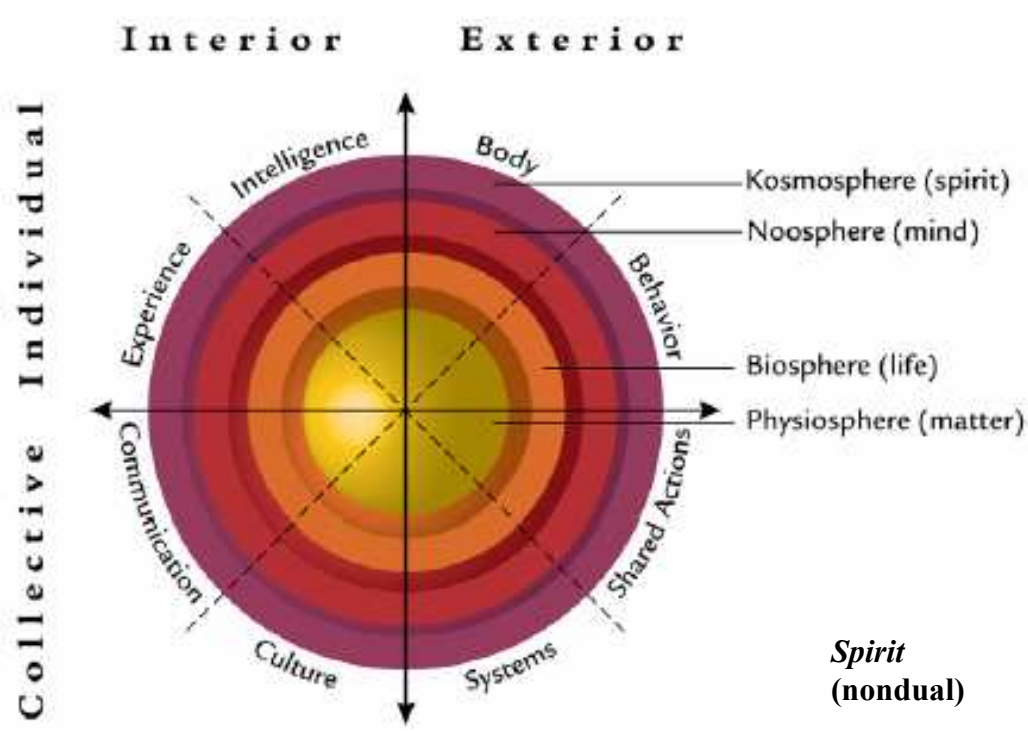
> At the same time, a body, a nation, exists in a larger context in which it functions as a part. The nation exists in the context of all nations that make up the whole political body of humanity. And all of humanity exists in a biosphere that is the body of the whole Earth. We exist as parts of a seamless whole in which everything is connected to everything else (DIAMOND; ORENSTEIN, 1990, p. 227).

Yet, as Wilber (2000a, p. 89) explains, if we truly existed as *parts* of a larger biosphere (Popper's level 12), so that the biosphere were really a higher level of organization than nation-states, as Russell ascertains, then (a) you couldn't have a biosphere *until* you had nation-states (sixth tenet), and (b) if you destroyed all nation-states you would destroy the biosphere (tenth tenet). Both of these propositions are obviously not true. In fact, it is just the opposite, meaning that the *human* sphere, i.e., the sphere of what is *distinctively* human, which we shall denote as *noosphere*[29], is *not* in the biosphere, but rather the *bios is a part of the nouns* – man is *not* part of nature but just the other way around. Of course, the human organism (e.g., organs, tissues, cells) does is part of the biosphere, as much as its own constituent parts (molecules, atoms, quarks…) pertain to the greater whole called the physiosphere (the realm of matter). However, it is not our physical or the biological constitution that make us distinctively humans. Actually, in this level there are no significant *qualitative* distinctions between humans and apes, dolphins, or even crocodiles (there are only quantitative distinctions, mostly restricted to the brain).

Accordingly, it is precisely because the biosphere is a component part of the noosphere that the destruction of the biosphere guarantees the destruction of all of humanity, *but not vice versa* (tenth tenet). Similarly, the biosphere is not a component or a part of physiosphere. The bios is not a part of the cosmos, but just the opposite: the cosmos is a part of, a component of, the bios (WILBER, 2000a, p. 99). Wilber then recommends that what is necessary is to construct a series of true holarchies of compound individuals (individual holons) and then indicate, *at the same level of organization*, the type of compound environment (social holon) in which the individual holon is a *participant* and on whose existence the individual depends (as a *member*, not as a component). According to him, this needs to be done in all of the three great realms of evolution: namely, physiosphere, biosphere, and noosphere.

[29] This is a term popularized by the Russian geochemist Vladmir Vernadsky ([1926] 1998) and French philosopher Teilhard di Cardin (1965). Literally, it means the sphere of human thought or consciousness, and is composed of interacting human minds and ideas.

As a first approximation in this direction – one that put together the contributions of many theorist working in seemingly unrelated fields and using disparate methodologies – Wilber affords us the "All quadrant, All Levels" (AQAL) map of the Kosmos, as illustrated in Figure 2.5. Some details of the AQAL map (pronounced *ah-qwul*) are indicated in Figure 2.6. In the chapter 6, after we have seen Graves's theory in more detain, we will zoon-in the AQAL map to examine the onto-phylogenetic parallel describing the development/evolution of the human holon, i.e., the development and structuration within the noosphere.



Source: Integral Sustainability Seminar (INTEGRAL INSTITUTE, 2006).

Figure 2.5 – The AQAL map of the Kosmos as the Great Nest of Being (each higher level transcends and includes all previous levels in the holarchy).

As a central piece of the Integral Operating System (IOS), the AQAL map begins to build up the Integral view of the undivided Kosmos first by factoring it in its four most basic perspectives: namely, the interior and the exterior of the singular (individual) and plural (collective). In this vein, Figures 2.5 and 2.6 represent a long-term perspective of the evolutionary process, where, according to the contemporary (objectivist) account, the physical realm of matter (the physiosphere) would have emerged with the "big bang", giving rise to the atoms, molecules, galaxies, stars, planets, and so on, along an evolutionary period lasting for billions of years (see the lower levels in both Right-Hand quadrants of the Figure 2.6). Eventually, conditions emerged in which molecules were able to combine to form cells, giving raise to the realm of life, or biosphere, comprising the collective organism and forms of life. Still according to the contemporary narrative, evolution continued in the biosphere, giving rise to numerous forms of life and eventually to the human realm, comprising human mind/brain functioning, tolls and technology, social organization, culture, etc., which together configure the noosphere.

Hence, the holarchy presented in these figures is in agreement with the sequence of the evolutionary process along time, as presently understood. As to its structural arrangement, Wilber notices that, in addition to the conflation between the individual and social aspects of holons, the difficulty in recognizing the proper order in the Great Holarchy has also to do with a confusion between size (or span) and embrace (or depth). He explains that the problem with span is that it distorts and confuses the actual nature of the holarchical Kosmos, because it is, in reality, based upon greater embrace and greater depth, not span (eighth tenet). Accordingly, "just as an atom is in a molecule but a molecule is not in an atom – even though the span of atoms is much, much larger than molecules – so the cosmos is in the bios but the bios in not in the cosmos, even though the span of the cosmos is astronomically larger" (ibid). As Reynolds observes (2004, p. 233), what makes the bio stand at a higher position in the holarchy is ultimately the fact that biology had indeed added "something extra" beyond the plain physical matter. The only thing "greater" about the physiosphere is its span, not

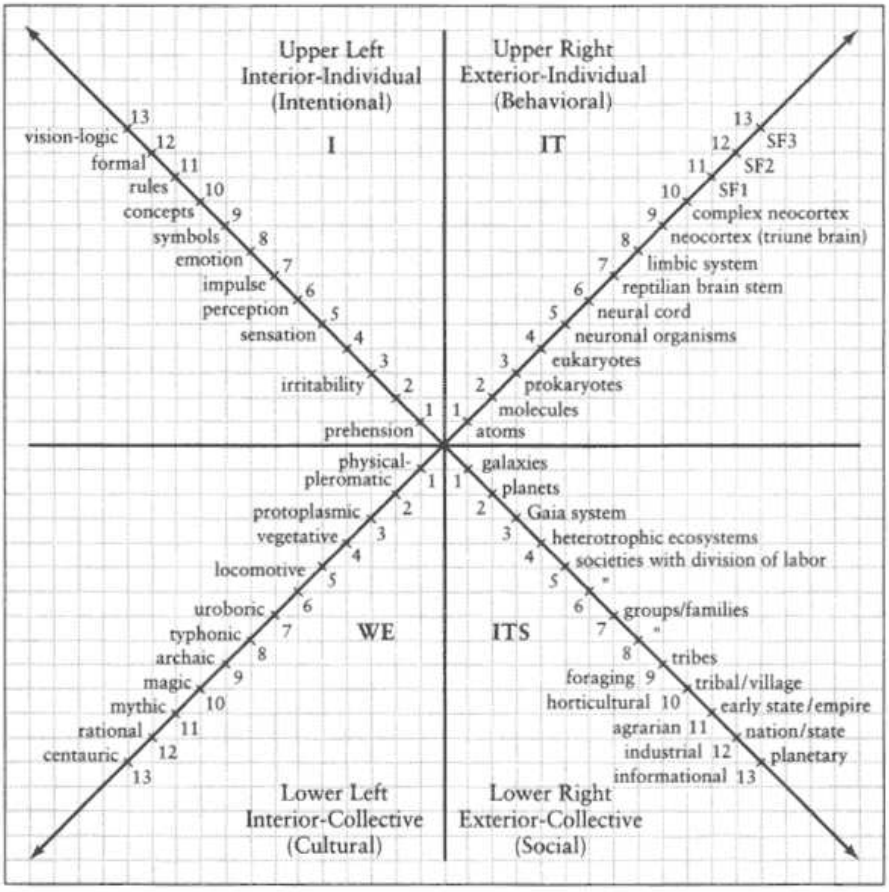
its depth or wholeness. Likewise, "Just as matter had pushed forth life, the self-transcendence drive *within* biology pushed forth something *beyond* biology…", it pushed forth the realm of what is *distinctively* human, the realm of symbols and tools, the realm of human mind and culture. In short, the biosphere has pushed forth the noosphere (WILBER, 2000a, p. 106).

Furthermore, as indicated, thus arranged the AQAL map provides a reliable key to address the most profound existential puzzle: namely, the so-called "mind-body problem". Accordingly, in unsnarling this world-knot Wilber puts forward two basic phases, amounting to the recognition of (1) the fourfold nature of the evolutionary process (eleventh tenet), and (2) its transfinite character or vertical openness (nineteenth tenet). Concerning the first step, Wilber points that it is not enough to say that organism and environment coevolve, as well as it is not enough to say that culture and consciousness coevolve. Rather, what it is necessary is "an approach to consciousness that differentiate-and-integrates all four quadrants" (WILBER, 2000b, p. 184). Hence, Wilber explains that, as recognized:

> The objective organism (the Upper-Right quadrant), with its DNA, its neuronal pathways, its brain systems, and its behavioral patterns, mutually interacts with the objective environment, ecosystems, and social realities (the Lower Right), and all of those indeed coevolve. Likewise, individual consciousness (Upper Left), with its internationality, structures, and states, arises within, and mutually interacts with the intersubjective culture (Lower Left) in which it finds itself, and which it in turn helps to create, so that these too, coevolve. But just as important, subjective intentionality and objective behavior mutually interact (e.g., through will and response), and cultural worldviews mutually interact which social structures, as does individual consciousness and behavior. In other words, all four quadrants – organism, environment, consciousness, and culture – cause and are caused by the others: "they tetra evolve".

From this, the second step requires the acknowledgment that (a) there are interior *stages* of consciousness development, and (b) that there are *higher levels* of consciousness development (all of which is indicated in Figure 2.6). As Wilber (ibid, note 15, p. 281) explains, "By acknowledging that the interior develops (as do their exteriors), we can see that mind (interior mental consciousness) and body (*interior* feelings) are related as transcend and include". In other words, "The felt body is in the mind", or "the mind transcends and

includes the [felt] body", as indicated in Figure 2.6 (Upper-Left quadrant), where symbols, concepts, rules, etc., are the interior aspects of higher holons that transcend and include vital feelings and sensorimotor awareness such as emotion, impulse, perception, sensation. As Wilber notes, this is "precisely why the mind can causally *operate on* the body". As well, this is also why interior growth is required so that moral concern (as in concepts and rules expressing deontic constraints) can *operate on* the choice behavior and then take precedence over impulsive and emotional drives. Figure 2.6 shows the exterior correlates of this interior holarchy as the evolution of the human *triune brain*, as composed by the reptilian brain stem (ruling rudimentary sensorimotor intelligence and instinctive drives or impulse), the limbic system (experienced as feelings and emotions), and the neocortex (which connect with the emergence of noosphere, at level 9, in all four quadrants).



Source: Wilber (2000a).

Figure 2.6 – Some details of the AQAL map of the Kosmos.

Accordingly, part of the reason for the persistence of the "mind-body problem" is that the collapse of the Kosmos simply rules out the reality of the interior holarchy, as already extensively discussed. Hence, as in the flatland world the material brain is indeed a part of the whole body (i.e., a part of the whole biological organism), thus consciousness must somehow be a function of the brain. The puzzle is than that, as many authorities have shown (e.g. PINKER, 2003) there is absolutely nothing in that brain that even vaguely corresponds to qualia or experiences or realities of mind and consciousness. In vein of that, as Wilber (2000b, p. 179) explains, "We must then either reduce consciousness to brain (and thus deny consciousness in its own terms), or accept the dualism as real, whereupon we can't even explain how I can move my arm (or how one reality affect the other)". Hence, just like his account on the battle between the Ego and the Eco, Wilber is suggesting that both of those solutions occur within the flatland paradigm.

Now, the second step in unsnarling the existential puzzle must move on toward the acknowledgement of the *transpersonal stages of development*, because the other part of the reason why both physicalists (or materialist) and dualists have generally agreed that the mind-body problem is irresolvable, "is not that they aren't smart enough to figure it out, but that it is only solved in postrational stages of development, stages which are generally suspect, ignored, or actively denied by most rational researches" (WILBER, 2000b, p. 181). Wilber illustrates the issue by pointing that just like a person at the prerational stage of development will not agree with, or even understand the rationalist claim that there is a *proof* for the Pythagorean Theorem – a proof that is true enough to virtually anybody who develops to the rational level and studies geometry, the "genuinely nondual solution for the mind-body problem in not something that can be fully grasped at the rational level (ibid)[30].

[30] In fact, as Wilber explains, "simply stating, in a rational fashion, that subject and object are nondual leads to all sorts of intractable problems and paradoxes". For example, "To say that subject and object are two aspects of an underlying reality begs the question as to what this underlying reality is, since it cannot be stated in terms that are not merely combinations of 'subjective' and 'objective.' Either this third entity, the underlying reality, has subjective and objective properties, or it does not. If it does, it is not really underlying; if it does not, it is really not unifying" (ibid, p. 276, note 13).

Accordingly, Wilber argues that those who develop to the nondual stages of consciousness unfolding are virtually unanimous that consciousness and matter, interior and exterior, self and world, subject and object are *both* distinct realities *and* aspects of the same thing: a true unity-in-diversity. Yet, that unity-in-diversity cannot be stated in rational terms *in a way that makes sense to anybody who has not also had a transrational experience*. To illustrate Wilber's point, I quote bellow a passage from Master Li Hongzhi, the founder of the spiritual practice named Falun Dafa (meaning, "The Great Cultivation Way of the Law Wheel"). In addressing this issue, Hongzhi says:

> In intellectual circles there's always been the question of whether matter is primary or mind is primary. They've been talking about this and debating it for a long time. I'm going to tell you, in reality, matter and mind are one and the same. The scientists in the field of human-body science are now thinking that the thoughts the human brain produces are matter. So they exist materially. But aren't they also in people's minds? Aren't they one and the same? It's just like what I've said about the universe—it has its material side, and at the same time, it has its nature side. The universe's nature, to be True, Good, and Endure, isn't something ordinary people can sense, and that's because ordinary people are all on this one level's plane. When you go beyond the level of ordinary people then you can experience this nature (HONGZHI, 2003, p. 14).

As Wilber puts it, this means that the "proof" of the nondual solution can only be found in the further development of the consciousness "of those who seek to know the solution". Of course, Wilber recognizes that this sort of solution ("you must further develop your own consciousness if you want to know its full dimensions") is not satisfactory to the rationalist (whether dualist or physicalist). Nonetheless, "it is the only acceptable form of solution according to a genuinely integral paradigm" (WILBER, 2000b, p. 181).

But even though, as I presume, most readers have not tapped into these transrational stages of development (just as I have not myself), we can recognize its logical necessity in the very structural basis of the Kosmos, as set forth by the twenty tenets, seeing as it requires it to be both *transfinite* (nineteenth tenet) and *nonfoundational* (twentieth tenet). Accordingly, Integral Theory holds that evolution continues in the noosphere, toward the realms of *soul* and *spirit*, to use the Christian terminology (see Figure 2.3). Wilber denotes the realm of spirit as

kosmosphere, while the realm of soul is represented as a transitional domain, connecting the noosphere to the kosmosphere (Figure 2.5).

The transfinite and nonfoundational nature of the Kosmos is then represented by the double nature of "spirit", which is at one and the same time fully transcendent (as, for example, the *Pure Subject* of Fichte), and fully immanent (as the *Pure Immanence* of Spinoza). The realm of fully transcendent spirit comprises the kosmosphere, whereas the fully immanent Spirit (as the paper sheet) represents the *common* (nondual) *ground of the entire sequence*. In this regard, Wilber mentions that the patriarchal religions tend to emphasize the transcendent "otherworldly" aspect of spirit (the Celestial Father), while the matriarchal, neopagan religions tend to emphasize the fully immanent or "thisworldy" aspect of Spirit (as the Virgin Mother, the Mother Earth).

Now, Wilber's postmetaphysical account of the kosmosphere requires first the recognition that "An individual holon [e.g. an individual human being] can eventually *embrace the entire Kosmos*" (WILBER, 2000a, p. 96), and thus experience its infinite depth and unbroken fruition. The first step in recognizing the reasonability of such possibility is actually implied in Figures 2.5 and 2.6, as the noosphere embraces the whole biosphere and physiosphere. Accordingly, as Wilber notes, "For the evolution of individual holons, *greater embrace* means that more of the universe is being taken into to holon", as it "is actually internal to that holon" (ibid). Once the proper holarchical arrangement of the Kosmos has been acknowledged, the second step requires definitely overthrowing the Univocal One, so as to admit that the production of valid knowledge is not confined to the methods of empirical sciences, but also amenable to the methods of interior inquiry. In this regard, Wilber (2001, p. 77) puts the flowing postmetaphysical validity claim for spirituality and mysticism:

> My thesis is simply this: deep spirituality involves direct investigation of experiential evidence disclosed in the higher stages of consciousness development. (I have called these stages psychy, subtle, causal, and nondual—which are simply summarized as "soul" and "spirit" in the figures). These deep-spiritual investigations follow the three strands of all good science (not narrow science, good science). They rely in specific social practices or injunctions (such as contemplation); they rest their claims on data and experiential evidence; and they constantly refine and check these data in a community of the

> adequate—which is why they are correctly referred to as contemplative science (which is certainly how they understand themselves).

Accordingly, quite apart from its logic-mathematic necessity (see note 15), Wilber's statement puts forward the epistemological-experiential basis of the twentieth tenet of evolution: namely, all IOUs (i.e., all holons) are redeemed in Emptiness. Hence, the twentieth tenet actually stands for a direct taste of nondual Spirit, seeing as

> Emptiness is neither a Whole nor a Part nor a Whole/Part. Emptiness is the reality of which all wholes and all parts are simply manifestations. In Emptiness I do not become Whole, nor do I realize that I am merely a Part of some Great Big Whole. Rather, in Emptiness I become the opening or clearing in which all wholes and all parts arise eternally.
> … Not in Emptiness but as Emptiness, I am released from the fate of a never-ending addition of parts, and I stand free as the Source and Suchness of the glorious display. I taste the sky and swallow whole the Kosmos, and nothing is added to me; I disappear in a million forms and nothing is subtracted; I rise as sun to greet my own day, and nothing moves at all (WILBER, 2000a, p. 532).

But Wilber still underlines that part of the novel claim of the Integral Approach is that "the interior data and experiences of soul and spirit (in the Upper-Left quadrant) *have correlates in the sensorimotor evidence in the Upper-Right quadrant*" (see Figure 2.6), so that they can be investigated (as they have been so) by hooking a meditator to an EEG machine, for instance (e.g. WALLACE, 1970; TEBECIS, 1975; LEHMANN; STIGSBY, 1977; HEBERT R et al., 1981; see also WILBER, 1997). Of course, in order to adequatelly interprete the subjective experience, and figure out the inner *meaning* linked to the EEG imagery produced in transcendental meditation (such as patterns of delta waves, which normally occur only in deep *dreamless* sleep), the experimenter must have first-person experience of that interior state, so as to be acquainted himself of that type of experience. Otherwise, she or he shall remain as clueless on this regard as she or he actually is in normal deep sleep[31].

[31] The substantive matter here is that such subjective experiences typically entails "a sense of expanded consciousness, and increase in love and compassion, a felling of encountering the sacred and numinous in both herself and the world at large". In short, the direct experience of that so-called delta state, bring about interior disclosures for which the word "spiritual" seem most fitting (WILBER, 2001, p. 65-66).

Moving further in the way of this astounding epistemological integration, Wilber also indicates that, according to the spiritual traditions, the entire process of evolution or "un-folding" from matter to body to mind to spirit could never occur without a prior process of involution, "in-folding", or "fall". Therefore, before evolution of nondual Spirit can occur (see Figure 2.5), involution of Kosmic spirit (i.e., spirit in the kosmosphere) must occur as a successive step down of the higher into the lower levels.[32] Hence, the higher levels appear to emerge "out of" the lower levels during evolution because, and only because, they were first deposited there by involution. Hegel (1974) conveys the idea:

> Everything that from eternity has happened in heaven and earth, the life of God and all the deeds of time simply are the struggles for Spirit to know itself, to find itself, to be for itself, and finally unite itself to itself; it is alienated and divided, but only so as to be able thus to find itself and to return to itself …

In other words, contrary to the flatland (system theoretic) account on "emergent properties" and "perpetual novelty", which is somewhat reminiscent of the idea of "spontaneous generation", the suggestion here is that one cannot get the higher out of the lower unless the higher were *already* there, in potential, sleeping or quiescent, so to speak, as a seed, waiting to emerge. Actually, in computational simulation of complex system dynamics, the interesting (meaningful) patterns that seem to emerge spontaneously from the system autopoiesis have in fact been planted there, inadvertently or thoroughly, in the form of encoded information by the very spirit of the analyst, as she or he experiment with alternative parameters (seeds) and models. These same seeds then grow and blossom in the analyst's consciousness in the form of meanings she or he recognizes or bestows to those emergent patterns.

In this vein, Integral Theory presents spirit as a "perfect expression of the self-transcending thrust of evolution", so that evolution is seen as a process of *creative* emergence, where any holon (including quarks, atoms, and molecules) is fundamentally a compound individual, a *sentient being* in a compound

[32] This seems to comprise the idea or reincarnation (of the causal spirit, not of individual personalities), seeing as the whole Kosmos is supposed to by in permanent cyclical movement. At this point, it is important to notice that the actual geometry of the Kosmos seems to be that of tire chamber, or a torun (see in this regard YOUNG, 1976a, 1976b, 1985 and 2005c).

environment (see Figure 2.6)[33]. Thus, an "all quadrant, all level" approach intimately integrates science and spirituality across many different fonts. As Wilber (2001, p. 78) puts it, it integrates spirituality or *deep* religion with *broad* science by showing that deep spirituality is in part a broad science of the farther reaches pf human potential. It also integrates spirituality with *narrow* or flatland science, because even deep-spiritual data and experience (such as mystical experiences) nonetheless have real correlates in the material brain, which can be carefully investigated with narrow science (as in the case of the meditator hooked to an EEG). And it even makes room for *narrow* (mythic and dogmatic) religion, as conveying lower level construal of the same spiritual realities and experiences disclosed in deep spirituality.[34] "In all of these cases", Wilber concludes, "an all-quadrant, all-level approach offers at least the possibility of a seamless intermeshing of what were previously thought to be nonoverlapping magisteria" (ibid).

2.4. The IOS and the IMP: upgrading our minds for integrally processing the Kosmos

In the Integral community's parlance, the Integral Operating System or IOS is often used as another phrase for the Integral Map. This expression is applied in analogy with the language of informatics, where an operating system is the infrastructure that allows various software programs to operate in an information network. In the present context, the "software programs" represent the multiple methodological perspectives according to which one may approach (or process) a given problem situation. Accordingly, the IOS provides the scaffolding that allows us to proceed the inquiry according to the principles of

[33] Of course, Wilber is not suggesting that quarks, for example, present reflexive intentionality as human beings (the anthropomorphic fallacy that Lovejoy called "retrotension"). He is simply stating that all holons have their own forms of interior perception or prehension (see the eighteenth tenet).

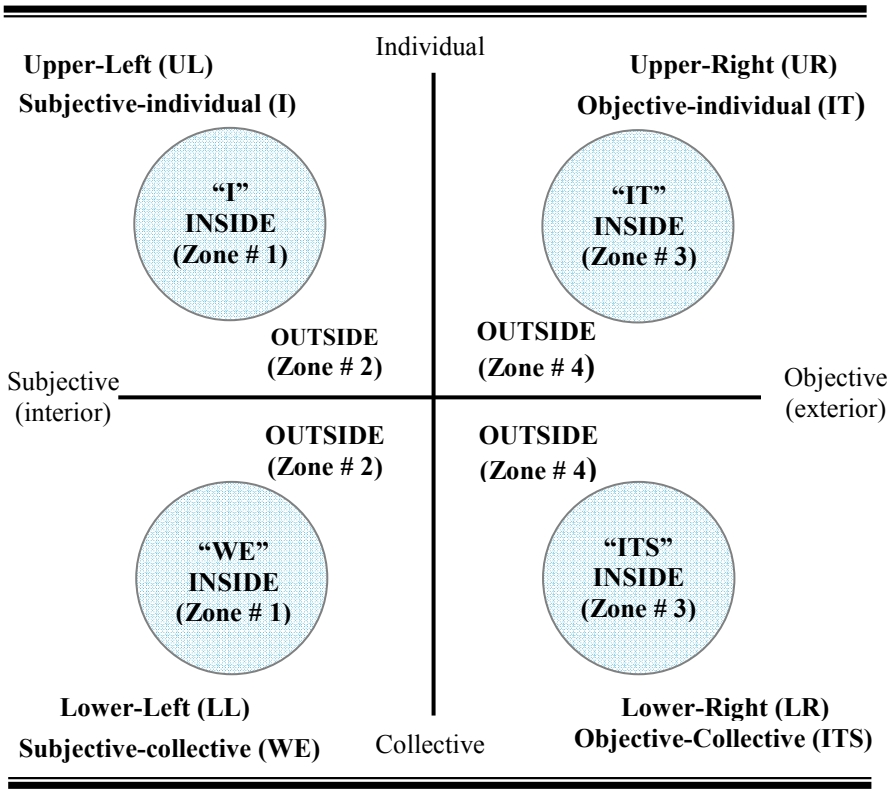
[34] What is implied in this assertion is that "the real core of [narrow] religion is deep religion or deep spirituality…" (WILBER, 2001, p. 80). Wilber further notes that this acknowledgement "tends to relax and lessen narrow-religion zeal, and thus, to the extent that you are alive to your own higher potentials, you will find narrow religion less and less appealing" (ibid).

the Integral Methodological Pluralism (see section 1.4), so as to approach any problem situation in more compressive and effective ways.

In this way, by relying in notions as elementary as those of *interior*, *exterior*, *singular* and *plural* (which jointly create the four quadrants), the IOS creates a *common language* that dramatically facilitates and accelerates the development of cross-disciplinary and transdisciplinary knowledge. Furthermore, as we saw, this integration endeavor is *not* restricted to the domain of science, as it is actually aimed at integrating the Big Three, i.e., the domains of science, art, and morals (see Figure 2.1). Because the IOS can be used to help index any activity – from dance to business to psychology to law to moral theory to ecology to spirituality – it allows each of those domains to talk to the others. We have indeed already seen how the IOS supported the outset of an outstanding view of the Kosmos as a colossal holon or Nest of Being and Knowing that *creatively* evolve from matter to body to mind to soul to spirit (Figure 2.4), simply by combining the basic elements of "quadrants" and "levels" into the AQAL map.

Evidently, as indicated previously, this was turned possible only on the grounds of rather large but also trustworthy orienting generalizations. In order to provide a more refined view of the Kosmos, i.e., one that points to aspects of holons that cannot be adequately addressed only within the scope of quadrants and levels, Wilber parsimoniously draws on just three additional and broad defined notions: namely, *lines*, *states*, and *types*. But before exploring these new simple factors of the IOS, it is important to notice that what renders it so much insightful and encompassing is that all of its five elements are *not* merely theoretical concepts. Rather, they are all *aspects of our own awareness*, so that they appear as *potential perspectives* available to all human beings. In this way, the Integral Approach does not ask the scholarly analyst, the ordinary observer, or the initiated witness of Emptiness, to adopt anything that they do not *already* have available to them; it does not offer some "outside" philosophy that people are asked to believe, but a *pointer* to perspectives and potentials that they already possess but perhaps are not fully utilizing or expressing.

Indeed, as we saw, amounting to the most basic perspectives on reality, the four quadrants emerge uncomplicatedly from the combination of the two primary polarities involving the "interior" and the "exterior", and the "individual" and the "collective" aspects of the *experience* of all holons. Now, if we recognize that each quadrant admits an "inside" and an "outside" view, the four quadrants give room for *eight indigenous perspectives*: namely, the inside and the outside views of the interior-individual realities/experiences, the inside and the outside of the interior-collective, the inside and the outside of the exterior-individual, and the inside and the outside views of the exterior-collective realities/experiences, as indicated in the Figure 2.7.



Source: Adapted from Wilber (2006, except B).

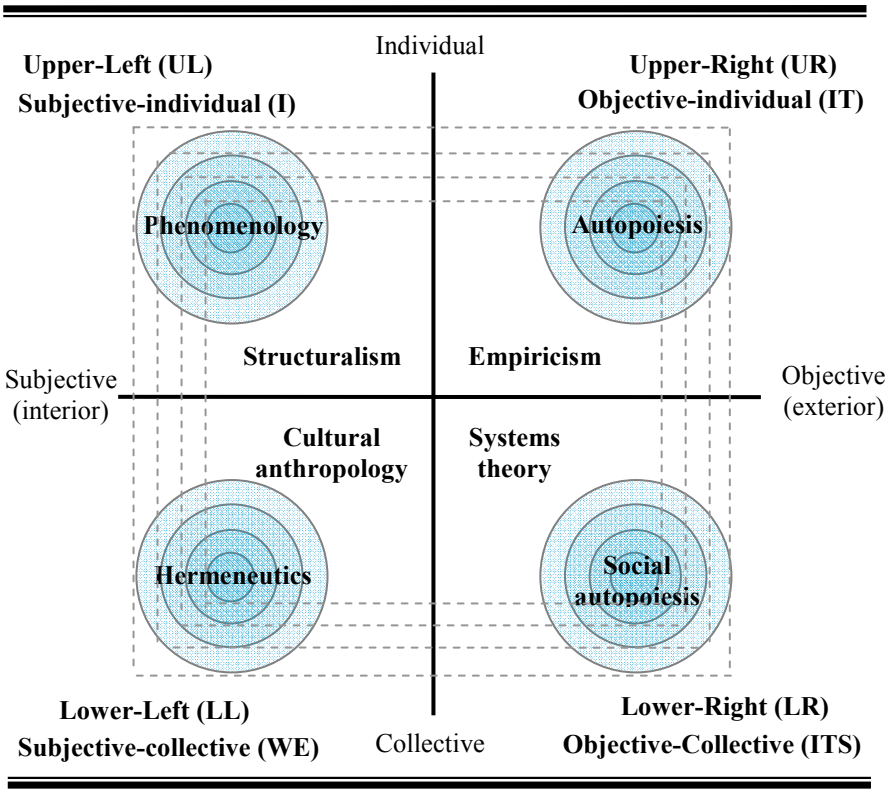
Figure 2.7 – Eight indigenous perspectives.

Wilber explains that these are the eight primal perspectives that all holons have available to them. This is why, far from being some sort of abstract systematization or new holistic theory without any territory, the IOS maps quite concrete phenomenological spaces from which most of the major forms of human inquiry have been launched. As examples, Figure 2.8 presents the eight major paradigms of methodologies that correspond to those native perspectives: namely, phenomenology and structuralism (in the Upper-Left), hermeneutics and cultural anthropology (Lower-Left), autopoiesis and empiricism (Upper-Right), and social autopoiesis and systems theory (Lower-Right).

The Figure 2.7 further indicates that phenomenology and hermeneutics (in Figure 2.8) are paradigms of the same kind, in the sense that both of them hold up methodologies designed to address the "inside of interiors", i.e., the first-person (felt) realities of both individuals and groups (or cultures). Wilber then clusters these two event horizons in the Zone # 1 (Figure 2.7). Accordingly, Zone # 2 encompasses the *same* interior realities when seen from the "outside", i.e., descriptively, so that psychological structuralism and cultural anthropology are examples of methodologies designed to describe systematically the "outside of the interiors". Zone # 4 is the realm of the conventional science (e.g. empiricism and systems theories), which descriptively approach (also according to their own evolving rules or right reason) the "outside of exteriors", i.e., the superficial appearances of the so-called objective phenomena taking place "out there", as they look like when perceived through our five sensory organs and their technological extensions. Finally, in Zone # 3, autopoiesis is the paradigm of the methodologies designed to elucidate the laws of formation or the regnant nexuses that are internal to the most diverse range of exterior (observable) phenomena, from the morphogenesis of an organ to the growth of a cancer to the formation of geological landscapes to the emergence of behavioral and spatial patterns in a wide range of social phenomena, such as industrial organization, trading structures, asset pricing, racial segregation, voting decisions, and so on.

Furthermore, by integrating ontology and epistemology along the structural foundation provided by the twenty tenets, the holarchic organization

provides the pattern that connects the levels of development identifiable in each quadrant by means of the adequate methodology. In the Figure 2.8, I attempt to represent this by connecting the holons in each quadrant using the doted lines (see also Figure 2.6).



Source: Adapted from Wilber (2006, except B).

Figure 2.8 – Eight major paradigms of methodologies.

Hence, this is how the IOS parsimoniously arrange the pieces of the epistemological puzzle not in a grab bag of unrelated paradigms, but in a meta-paradigm that elegantly weaves together its many threads into a true unity-in-diversity that slights neither the unity nor the diversity. It seems quite clear, or thus I regard, that Integral Theory's IMP move us quite beyond from Gintis's (2003) proposal of unifying the *behavioral* sciences under the canon of the instrumental conception of rationality.

Now, with regard to the new elements of the IOS, while *levels* denotes *enduring states of equilibrium*, which Wilber sees as morphogenetic fields, "Kosmic habits of evolution", or developmental spaces through which the holons flow on the road of strange attractors, the element of states denotes *temporary* or *passing states* that do not present developmental quality in themselves[35].

With reference to the human interiority, there are all sorts of states of consciousness that provide profound motivation, meaning, and drives – and that can be experienced as really fleeting (e.g. fury and many sorts of peak experiences), more or less transitory states (as all sorts of passions, including grief and joy), or even quite lasting states (such as depression, melancholy, and detachment). In any particular situation, states of consciousness may not be a very important factor, or they may be *the* determinant factor driving behavior in that situation. Also, most states of consciousness are triggered by contingent factors beyond the control of the individual, while others can be induced by intense experiences like making love or listening to exquisite music, by drug consumption, or bought forth through practice, like meditative states induced by yoga, meditation, contemplation, and so on.

Needless to say that states are "all quadrants", so that there are all sort of intersubjective states of consciousnesses (e.g., during religious rituals, military training, political movements, and so on), as well as states or transitory phases that can be observed in the dynamic of objective systems (e.g. embryogenesis, both chaotic and periodic dynamics of various natures). Most important in this regard is that interior states do not hove in the air, dangling and disembodied. On the contrary, as Wilber (2005, p. 23) puts it, "For every state of consciousness, there is a felt energetic component, and embodied feeling, a concrete vehicle which provides the actual support for any state of awareness".

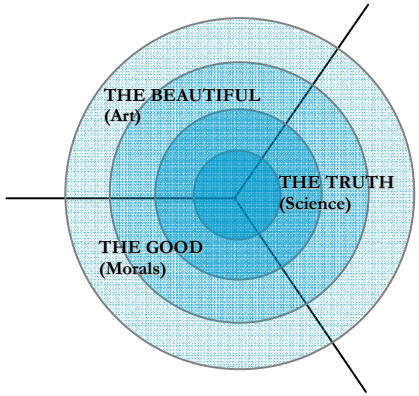
In particular, Wilber explores the quite important link between the three natural states of consciousness: waking, dreaming, and deep sleep, and their corresponding *bodies* or energetic mediums: namely, gross body, subtle body,

[35] It should be noticed, however, that states of consciousness present connections with the holarchic structuration of organism, particularly the triune brain, along the evolutionary processes.

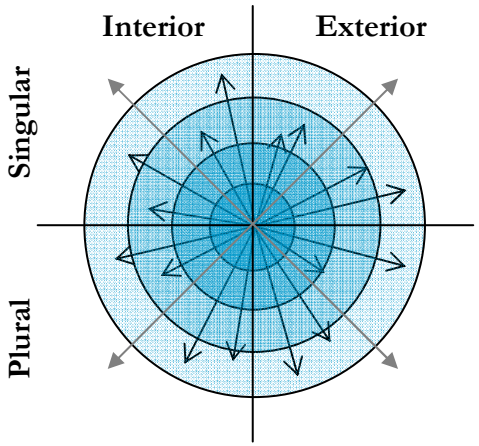
and causal body (see Figure 2.9, panel e). Again, there is nothing metaphysical here, seeing as (a) the *gross* body is simply the ordinary physical body we are aware of *when* waking; (b) the *subtle* body consists of the energy fields associated to the fluid and flowing images and emotional feelings we are aware of in the dream state (as much as we are unaware of both the gross body and the whole Right-Hand, i.e., exterior reality), and (c) the *causal* body stands for the finest energetic field configuration running and sustaining our lives when consciousness of both gross and subtle bodies vanishes in deep (dreamless) sleep. Wilber (2005, p. 24) points that the great wisdom traditions maintain that in this (deep sleep) state – which might seem like merely blank or nothingness – "we are actually plunged into a vast formless realm, a great Emptiness or Ground of being, an expanse of consciousness that seen almost infinite", and "out of which creative possibilities can arise". Of course, many people do not experience that state in any other way but as total obliviousness. Yet, as Wilber indicates, "the wisdom traditions are unanimous that this *formless state* and its *causal body* can be entered in full awareness, whereupon they too, yield their extraordinary potentials for growth and awareness" (ibid).

Again, to illustrate, I quote bellow Master Li Hongzhi's instructions on how Falun Dafa practitioner should do their meditation practice:

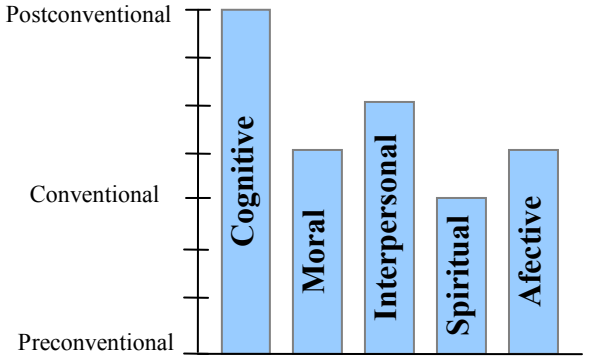
> We have a rule that your master consciousness has to be aware, and that's because this practice cultivates the real you. So you have to improve yourself while you're fully aware of it. Now, we have a meditation. And how do we do our meditation? As a rule, we say that no matter how deep your concentration gets, you have to know that you're practicing here, and you just absolutely can't get into that kind of state where you're totally oblivious. So what exactly will you experience? What will happen is that when you go sit there you'll feel wonderful, like you're sitting in an eggshell, and you'll feel really good, you'll know that you are cultivating, but you'll feel like your whole [gross] body can't move. These are things that have to happen in our practice. There's another cultivation state: you sit and sit, and then you discover that your legs are gone, you can't figure out where they went, and your body is gone, too, your arms are gone too, your hands are also gone, and only your head is left. As you keep meditating, you discover that your head is gone, too, and only your mind is left – there's just a little bit of awareness left knowing that you're cultivating here. If you can reach that point it will be more than good enough. And why is that? When a person does the exercises in that state his body is being evolved to its fullest extent. It's the best state (HONGZHI, 2003, p. 169).



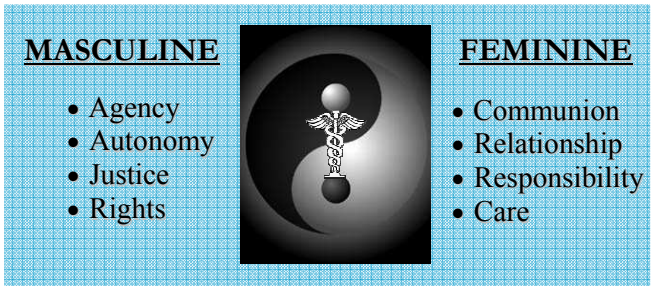
(a) THE BIG THREE



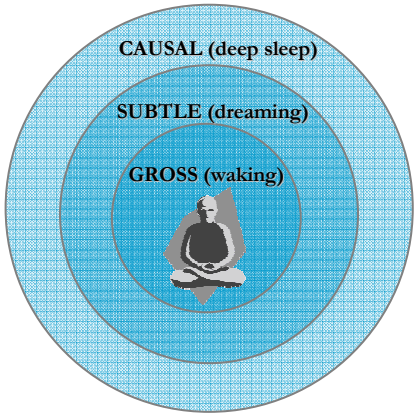
(b) AQAL map (with multiple lines)



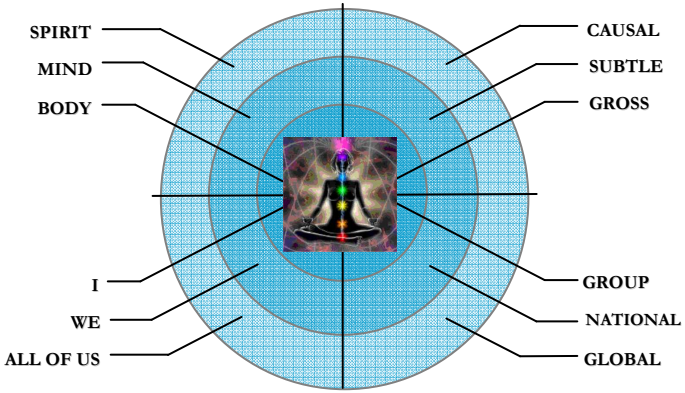
(c) PSYCHOGRAPH (LINES AND LEVELS)



(d) TYPES



(e) STATES AND BODIES



(f) AQAL (in humans)

SOURCE: Adapted from Wilber (2005).

Figure 2.9. The Integral Operating System (IOS).

We shall return to this subject in the next chapter, and again in chapter six, in order to explore a little further the connection between this ability to sustain awareness (direct experience) of what lies beneath our gross material bodies and the steadfastness of the moral dispositions, such as truthfulness, compassion, and forbearance (or self-control), which are qualities patently significant in fostering cooperative solutions in social dilemma situations.

As this issue concerns chiefly to the development of spirituality, which *may* have little to do with the development of instrumental reason, for instance, the present discussion brings up IOS's element of *lines* or *streams* of consciousness. Howard Gardner (1983 and 1993) made this concept fairly well-known using the idea of multiple intelligences, such as emotional and interpersonal intelligences, moral reasoning, logical thinking, and so on. Actually, I have already tacitly alluded to this aspect of the IOS in the first chapter, when commenting on Wilber's table of correlations involving over one hundred developmental models covering multiple lines of interior growth. I mentioned there that there is currently a great deal of consensus in the field of developmental psychology that these various lines unfold along progressive stages or levels of development, and that the upward move entails the decreasing of the egocentrism and the broadening of the moral embrace, i.e., of the range of beings that are believed as worth of moral concerns. The fine-tuning now entails the recognition that the development of different lines occurs in a *relatively independent* fashion. In this way, the Integral Approach accounts for the recognized fact that one may excel in cognitive development while do poorly in emotional or interpersonal intelligence, for example. The idea of multiple developmental lines unfolding in their own pace, in connection with different factors and variable circumstances, also allows for the almost infinite number of combinations that bring forth the distinctiveness of individuals' global personalities.

We will return to this issue in the chapter six. Respecting the completion of the IOS's toolkit, by putting lines and levels together Wilber brings up the so-called "psychograph", which displays the developmental profile of an individual

in some point in time. An example of a common psychographic configuration is presented in the panel (c) of the Figure 2.9. The regular feature embedded in this illustration is the more advanced development of the cognitive line in relation to the others. This is usually the case because cognitive development appears to be a necessary, but not a sufficient requisite to foster the development of other streams of consciousness (WILBER, 2000b).

Once again, bearing in mind the (tetra) enactment principle of development and evolution (eleventh tenet), the multiple intelligences or subjective drives (Upper-Left) contribute to bring about multiple lines of behavior in the Upper-Right quadrant, including conduct toward the natural environment, the opposite sex, or the other participants in collective action situations, response to rules and regulations, hygiene habits, birth control use, money management, and so on. In the collective-exterior domain (Lower-Right) the line element of social development shows up as any of the concrete, material form of social embeddedness, including the development of modes of tools and technologies, architectural styles, forces of production, and concrete institutions or forms political organization. Finally, in the Lower-Left the multiple lines of the cultural dimension stand for different elements compounding the overall culture such as worldviews, shared values, perceptions, and meanings, styles of intersubjective justification, concepts of justice, motivation, and authority, rituals, taboos, informal institutions, and so on. The idea is conveyed in the Figure 2.9, in the panel (b). The crucial point is that all lines go through similar qualitative transformations, which configure the said "Kosmic habits" of evolution, as fashioned by the principle of hierarchical inclusiveness (see first, fourth, and fifth tenets).

Finally, *types* refer to items that can be present at virtually any stage or state. Accordingly, there are numerous "horizontal typologies" that can used to explain behavior in the face of diverse situations. Once again, I commented in the first chapter about the growing interest in examining the implications of personality types – like Friedman and Rosenman's (1974) "Type-A/Type-B behaviors" (BOONE et al., 1999); Rotter's (1966) internal versus external "locus

of control;" high versus low "self-monitoring" (SNYDER, 1974; SNYDER; GANGESTAD, 1986), among others – on behavior of individuals in collective action situations and Prisoners' Dilemma games. I also suggested there that a more interesting analysis would ensue from the combination of types and levels (as in Graves's scheme), since a person can be a particular type at virtually any of the levels, and levels, contrary to types, present an internal dynamic that should be first understood, and then taken into account when considering which type of incentives one should create so as to change behavior in any desired direction.

To illustrate the aspect of types, a contrast between the male (yang) and female (yin) orientations is presented in the panel (d) of the Figure 2.9. The figure points to the idea that "the typical male orientation tends to be more agentic, autonomous, abstract, and independent, based on rights and justice; whereas the female orientation tend to be more permeable, relational, and feelingful, based on care and responsibility" (WILBER, 2000b, p. 46). It is interesting to notice that such polarity touches basis with the basic conflict inherent to collective action problems, and one might expect that females would tend to be more cooperative in response to the moral dilemma implicated in the conflict between agency and communion. Nonetheless, drawing mostly in Carol Gilligan's (1982) enormous influential book *In a Different Voice*, Wilber indicates that both men and women tend to develop through three or four major levels or stages of moral development – namely, preconventional, conventional, postconventional, and integrated – that are quite similar to Kohlberg's stages of moral reasoning, which Gilligan began by criticizing. Further, Wilber (2005) indicates that Gilligan refers to these stages as *hierarchical* because each stage has a *higher* capacity for care and compassion, but that she highlighted that women progress to these stages using a different type of logic – they develop "in a different voice". Also, according to Wilber's report, Gilligan specifically calls these hierarchical stages in women *selfish* (which is egocentric), *care* (which is ethnocentric), *universal care* (which is worldcentric), and *integrated*. Once more,

Wilber emphasizes that the Gilligan's stages brings about increasing capacity for care and compassion[36].

Finally, it is important to notice that the element of type also entails the distinction between the *healthy* and *unhealthy* versions of types, levels, and states. This is precisely the aspect of types under Wilber's account on the paradox of damage, discussed previously. Accordingly, the Ego camp's pathological agency can be seen an unhealthy expression of the masculine or yang principle.

> "If the healthy masculine principle tends toward autonomy, strength, independence, and freedom, when that principle becomes unhealthy or pathological, all of those positive virtues either over- or under-fire. There is not just autonomy, but alienation; not just strength, but domination; not just independence, but morbid fear of relationship and commitment; not just drive toward freedom, but drive to destroy. The unhealthy masculine principle does not transcend in freedom, but dominates in fear" (WILBER, 2005, p. 21).

Correspondingly, "If the healthy feminine principle [as healthy communion] tends toward flowing, relationship, care, and compassion, the unhealthy feminine flounders in each of those. Instead of being in relationship, she becomes lost in relationship. Instead of a healthy self in communion with others, she loses herself altogether and is dominated by the relationship she is in. Not a connection, but fusion; not a flow state, but panic state; not communion, but meltdown. The unhealthy feminine principle [just like the Eco camp] does not find fullness in connection, but chaos in fusion" (ibid).

Once again, it seems worth mentioning at this point that the wisdom traditions have ever associated the cultivation to higher levels of consciousness to the achievement of a balance between the yin and yang principles. In particular, Tang Tantrism includes the so-called male-female dual cultivation, whose specific purpose "is to collect yin to supplement the yang, and collect the yang to supplement the yin, so that yin and yang supplement each other and cultivate each other to achieve the goal of getting the yin and yang in balance"

[36] Wilber comments Gilligan's work is so widely misread and distorted because the postmodern stance denies hierarchies in general. In this fashion, he points that many feminists confused the idea that females tend to not to think hierarchically with the idea that females to not develop hierarchically. He clarifies the while the former is true, the latter is false, according to Gilligan herself (WILBER, 2000b, p. 47).

(HONGZHI, 2003, p. 181). Yet, as Master Hongzhi observes, "whether it's Buddhist or it's Daoists, specially the Daoists' yin-yang theory, they all talk about how the human body naturally has both yin and yang", so that one is still able to cultivate toward higher levels by bringing these primordial principles into balance within ourselves.

2.5. The implications of the integral approach for the investigation of the commons dilemma

The *general* implication of the Integral Approach for the investigation of the commons dilemma is just the same it has for the investigation of any problem: namely, to elicit more balanced, comprehensive, and effective solutions, by acknowledging and incorporating the full range of resources for any situation – in all quadrants, all levels, all lines, all states, and all types – without excluding or denying any of them. This is so because, as elements of our own awareness, and therefore constitutive of any perspective on the problem situation, all of the IOS's elements are certainly affecting the situation and the problems being generated.

By contrast, the methods of the other major approaches appear to have chief biases built into them, because they typically advance their truth as the only fundamentally correct approach while condemning the others as inferior or even dangerous (WILBER, 2005). We saw that this sort of predisposition is quite visible in all of the three great periods of the formation of the contemporary Western thought, as postmodernism paradoxically attempt to absolutize the perspective of plural relativism in order to confront the scientific imperialism brought forth with modernity. Such imperialism, in turn, seems to embed a kind of vengeance to the historic brutalities instigated by the religious absolutism that characterized the premodern worldview and mode of life (but that remain quite alive at the present time, in different contexts).

Hence, when discussing the collapse of the Kosmos, we saw that the rational-scientific method, in its exclusive form, focuses problem analysis (and

solution) on systems and processes, and for the most part excludes issues associated with individual meaning, aesthetics, and group culture. In other words, the empirical sciences absolutize their own favorite Right-Hand quadrants. Even systems theory, which claims to be "comprehensive" and "all-inclusive", in fact greatly privileges the Lower-Right quadrant – and eventually explicitly denies irreducible reality to all the subjective experiences in the domains of aesthetics and morals (WILBER, 2005), as exemplarily attested in Crabtree's (2006) sentence that "altruism is image and illusion", or in Dawkins's (2006) "God Delusion" (see also DAWKINS, 1976).

Likewise, postmodern pluralism tends to absolutize the Lower-Left quadrant, by granting reality to the social or cultural dimensions, while denying or at least greatly discounting the objective realities, as described by the paradigms of methodologies in the Right-Hand quadrants. Eventually, extreme postmodernists look upon science as a mere interpretation, much like poetry. "But clearly", as Wilber (2005) puts it, "a diamond will cut a piece of glass no matter what culture it appears in", meaning that there are important objective truths that need to be honored if any enduring solutions to the world's problems are to be discovered.

The Integral Approach then assists us to move in this direction by accepting the *partial* truths of the greatest number of perspectives, while denying that any of them alone have the only truth. In this vein, pervaded with genuine vitality rather than being instructed by mere theoretical concepts, the Integral Approach acknowledges and thoroughly combines all of their important contributions to facilitate fresh, comprehensive, and truly exciting approaches to resolving some of the world's recalcitrant problems.

With reference to collective action situations, Figure 2.10 illustrates how *this* particular move toward an Integral Approach of the commons dilemma looks like. The *specific* contribution of the Integral Approach in this context is that it allows us to address the vital *moral* aspect of the commons dilemma forthrightly, i.e., in its own terms, without either denying its irreducible reality or turning it into a mere contingent feature of cultural traditions – and this is achieved without

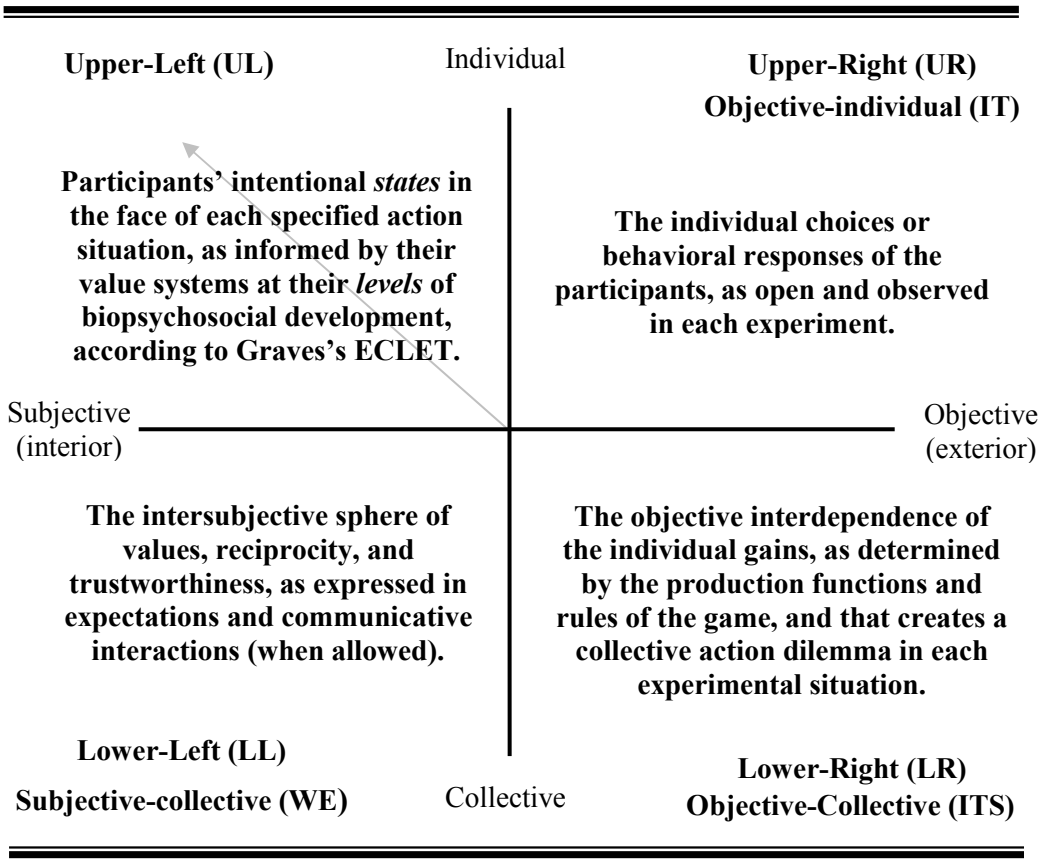
denying as well the significance of these contingent determinants of behavior in both the objective-systemic and the cultural domains.

This is a rather important achievement because, as we saw, the collapsed ontology entraps the torrent of perspectives into a huge dye vat that levels all qualitative distinctions, and then makes it impossible to uphold the profound intuition – in the core of virtually all wisdom traditions, as well as deep inside all of us – that there are after all quite different degrees of moral virtue. Think, for instance, in which basis we can distinguish between a *type* of spirituality that uphold compassion to all sentient being from a *type* of spirituality in which self-destruction is justified, and believed to be recompensed by Heaven, provided it is used as weapon to kill infidels. Note as well in which basis this sort of self-sacrifying conviction is uphold. Evidently, the problem here is not that there are no different degrees of moral virtue, but that there is no consensus about how these degrees could be hierarchized. Once more, following the differentiation of Big Three and the disaster of modernity, the reason for that difficulty resides fundamentally in collapsed view of the Kosmos, which underpins a flatland ontology wherefrom the postmodern view is not really able to escape. Indeed, perhaps for the first time in the *official* history of humankind, we are in a position to rectify this terribly awful situation.[37]

This is why, in the present contribution, I am primarily concerned with the implications of the *developmental* aspect of moral reasoning, i.e., with the element of *levels*, in the value systems developmental *line*, and that underlies the intentional *states* (all in the Upper-Left quadrant) and then the individual choices and behaviors (Upper-Right) of the participants embroiled in collective action (Lower-Right) dilemmas (see Figure 2.10). Note that the experimental designs warranty the interdependence that creates the objective conflict between the individual and the collective tangible benefits in the three experiments, but the *moral* conflict between the *agency* and *communion* is an essentially subjective phenomenon that must be approached using the Left-Hand methodologies (such

[37] According to great wisdom traditions, such as the Buddha Scholl and the Taoist system, there were many (at least 81) civilization periods prior the current one (cf. HONGZHI, 2003, p. 8-11).

as Graves's ECLET), not simply disregarded on the grounds on the absolutism of the Right-Hand methodologies (see Figure 2.8).



Source: Author's configuration.

Figure 2.10 – Toward an integral approach of the commons dilemma.

The intersubjective sphere is present in the form of the individuals' expectations regarding the behavior of the other participants. These expectations involve the degree in which the values and worldviews embraced by each participant are believed to be shared. When communication is allowed (in the second condition of the appropriation dilemma experiment), the intersubjective sphere constitutes the phenomenological space where the extent in which meanings and significations are shared supports mutual understanding and expectations of trustworthiness and reciprocity. This is the underlying reasoning under the working hypothesis that normative commitment in the appropriation

dilemma may be facilitated not only by the presence in the group composition of individuals centered at some of the sacrifice-self type of systems, but first and foremost by the presence of individuals that have achieved the *higher stages* of development in Graves's scheme. The point here is that the holarchical structure of development means that the individuals centered at lower levels do not form expectations based on styles of moral reasoning that are open only at the higher levels of development, while the individuals centered at higher levels do realize the styles of moral reasoning that guides the argumentation ensuing from lower levels of development.

It should be notices in this regard that, insofar as rational choice is concerned, the same inference is supported by the theoretical scaffolding provided by Habermas's philosophical justification of discourse ethics and theory of communicative action. As we shall see, this is so because in Graves's scheme the level six brings forth an orientation toward principles of justice and procedures for justifying norms that underpin Habermas's discourse ethics and the communicative *type* of rational action. Furthermore, while Graves's level six is a personality system of the sacrifice-self *type*, the enfoldment principle of evolution assures that these orientations are subsumed in the level seven, which is an express-self type of system. Yet, as Wilber (2000b, p. 83) points out, Habermas lamentably leaves out and totally ignores any of the stages of consciousness beyond vision-logic (i.e., transrational stages open at level eight, in Graves's scheme). "Moreover", Wilber continues, "in placing his reliance on linguistically generated structures of understanding, Habermas places an unfortunate wedge between human and nonhuman nature, so that this approach to nature is essentially instrumental". In view of that, Wilber claims that Habermas view is, in short, "inadequate to both prerational and the transrational domains" (ibid). On the other hand, Graves's ECLET and Wilber's spectrum of consciousness are both essentially open-ended conceptions that explicitly include transrational stages of awareness, as indicated.

Finally, as we saw, Integral Theory's Twenty Tenets of evolution confer an enormous significance not only to the agency and communion tension that is

in the core of the commons dilemma, but also to the aspect of increasing *telos*, i.e., purposiveness or intentionality (eighteenth tenet). In this vein, Integral Theory roots the justification for the nested hierarchical pattern of moral development (as well as of any other line of development), and for the qualitative changes marked by decreasing egocentrism and increasing moral embrace right into structural foundation of the whole Kosmos. In turn, following this deep rooted pattern, increasing telos eventually bring forth the capacity exercise autonomy (agency) taking into account the interest of others (communion).

This fundamentally means that the ability to overcome the social dilemmas resides first and foremost *inside* the individuals, and that the detailed analysis of the contingences in different action situations is instilled with much more significance after we have *explicitly* recognized that, i.e., that the interior disposition to behave in this way is neither a given (no matter how or if at all) nor can be brought about by an education designed to "reveal to all" that everything is connected to everything else, in the style of the systems theories. Rather, it requires above all "learning ways to foster many arduous waves of interior growth" (WILBER, 2000b, p. 138).

Clearly, I am at this point just restating the central thesis of this study. However, I would like to notice that by digging a little deeper than customary when running into our own narrow working area, I take the commons dilemma not simply as an subject matter to discuss the efficiency and the sustainability of resource use, however important that may be, but as an exceptionally convenient theme to reflect upon the old question: namely, "what is original motivation behind the complex phenomena of this existing universe?"

3. UNCOVERING PARTIALNESS

3.1. Introduction

As suggested in the previous chapter, the Integral Theory provides the broader framework I employ in this research to expose the partialness involved in the regular view which identifies rationality with instrumental action, and sees both norm-governed behavior and unbinding willingness to cooperate as ensuing from some sort of nonrational source. Though the assumption of instrumental rationality may impart just a methodological strategy, which clear limits of applicability, theorists commonly ask for generalization, especially currently, as a series of theoretical developments and new research techniques allow extending the principle of the "survival of the fittest" to explain anything that endures in time, including cultural traits, social institutions, dress styles, philosophical ideas, music styles, and so on (e.g. AXELROD; HAMILTON, 1984; AXELROD, 1986; HENRICH, 2000; HERNRICH et al., 2003; McELREATH et al., 2003; GINTIS, 2003).

As summarized by Heath (forthcoming), the problem with this evolutionary approach to culture and human behavior is that while biological factors may contribute to the cultural fitness of a behavioral pattern, and thus exercise an influence over which patterns get selected, they cannot act as a direct

force of selection, so as to determine the *content* of particular intentional state, because the structure of the cultural platform is so fitness-enhancing that it permits the transmission of all sorts of contents that are biologically maladaptive. Heath explains that while in the purely formal sense one may say that no matter a given behavioral trait is interpreted as genetically programmed disposition or as culturally transmitted "meme" they are still subject to the same replicator dynamics, what makes the "survival of the fittest" principle useful in biology is that there we have a relatively independent concept of fitness and reproductive success. This allows us to make confident predictions about the effects that particular mutations will have over the representation of a particular allele in the population. Conversely, there is no clear and independent criterion to define what determines the "fitness" of a cultural trait or "meme". As a result, the only way to determine the "fitness" of a meme, such as a value judgment or idea of morality embedded into a behavioral pattern, is to see how well such pattern is represented at time t_1 and at time t_2, then infer that the one with increased representation is more fit and the one with decreased representation is less so. As Heath point out, there is nothing to stop one from doing so, at a strictly formal level. But the problem is that it turns the old "survival of the fittest" slogan into a tautology, as the fit becomes nothing other than that which survives. Finally, this approach renders as merely conventional the motivational structures associated with value systems and morality, and truncates the account of the social interactions, especially communicative acts, to entail only its instrumental or strategic facet. Although this does not mean that explanations grounded upon functional fit undermine the significance of deontic constraints and responsibility, it does foster that kind of moral skepticism that we see in Hardin's (1968) attack to "conscience", as a possible case to be made in order to deal with the commons dilemma.

I will return to this issue below in this chapter, when reviewing the evolutionary approach to cooperation. The necessity of introducing an additional logic (i.e. communicative rationality) for grasping the full nature of social interactions will be discussed in the chapter four, along with the Habermas's

theory of communicative action (HABERMAS, 1984 and 1987). Before that, I put forward a tentative illustration of what an Integral approach to collective action should look like, and then use the suggested outline to weigh up the incompleteness involved in the most important approaches to our subject matter thus far.

3.2. An integral map to come across partialness

As indicated, the Figure 3.1 maps into the four quadrants the principal elements connected with an Integral Approach to collective action. The entries are of course far from comprehensive. The aim is just to arrange the main intervening variables in the four domains of the AQAL matrix, while suggest the evolutionary/developmental aspect constitutive of most of them, as indicated by the diagonal line. When the entries refer to individual agents, the diagonal is read in the ontogenetic/developmental sense. When they refer to characteristics of our species (still individual) or societies (collective), the line represents the phylogenetic/evolutionary time. I do not mean that every individual entry in the Figure 3.1 ought to be read in this way. The specific Integral framework in effect in this research will be presented in the chapter four (see Figure 5.1), after we have seen the characteristics of Graves's ECLET conception.

Accordingly, in the Upper-Left quadrant I clustered the principal *lines* or streams of psychological development, plus the suggestion that horizontal *types* (e.g. traits of personally) may similarly influence individuals' attitudes toward collective action problems in ways that are relatively independent of psychological development. As indicated, I shall fill the gaps left in this quadrant using Clare Graves's ECLET conception, since the said openings relates mainly with developmental aspect of human subjective life. Indeed, there are an increasing number of contributions exploring the effect of personality traits or particular aspects of individual inclinations on collective action dilemmas. Boone et al (1999), for example, also argue that economic agents differ as to their inclination toward cooperation, and investigate the effect of personality on

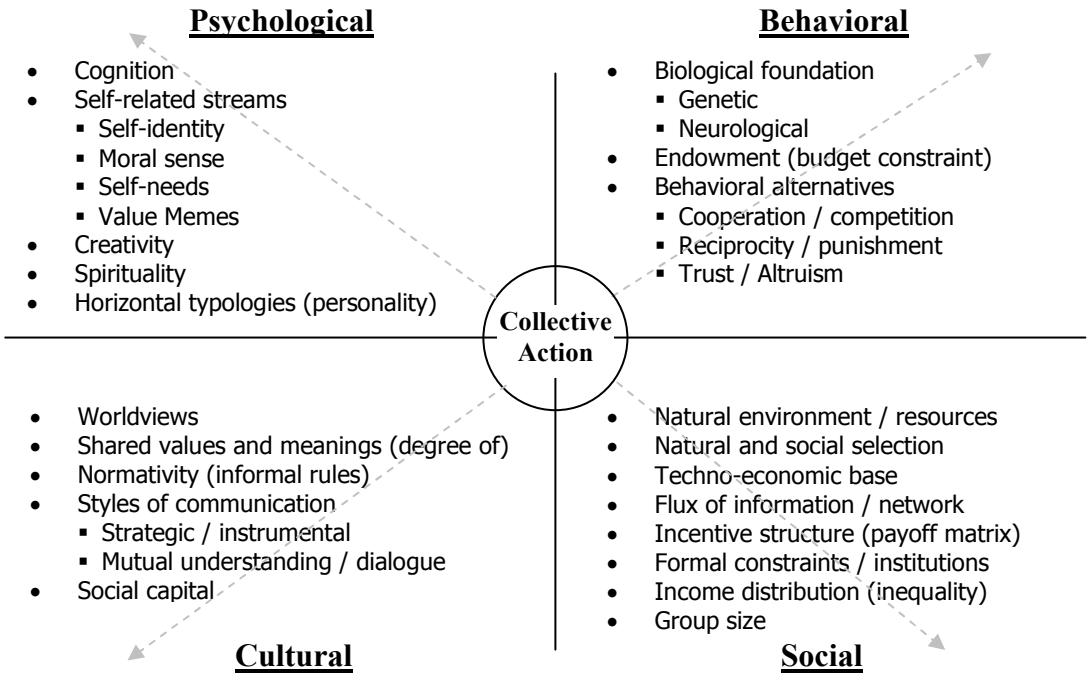
competitive versus cooperative behavior in repeated Prisoner's Dilemma (PD) games. The author's results clearly demonstrate that the type of personality of the player matters and are robust[38]. Also in the setting of repeated PD games, Curry (2005) examined the effect of "patience", i.e., the weight attached by subjects to immediate versus delayed benefits, whereas Price (2005) showed that guilt-prone people contribute more in anonymous public good games. While both Curry and Prince work within the framework of evolutionary psychology, personality is likewise treated as a matter of *type*, and no attempt is made toward the integration of such types into a full developed theory of the formation and transformations of the adult personality, as I do here.

In the Lower-Right quadrant, I mention the evolution of worldviews, shared values, meanings and normativity. It should be noticed that cultural evolution is a sensitive matter, partially for good reasons, partially out of a radical relativist prejudice. The issues on this hot topic will be discussed in the chapter four. There we shall suggest that the possibility of an ontogenetic/psychological-phylogenetic/cultural parallel ensues directly from the conception of the Human Holon and configures a central trait of the Integral Theory. As well, such onto/phylo parallel is present in Graves's ECLET conception as a linkage between the psychosocial development of the adult human (Upper-Left) and the historic evolution of the social living conditions (Lower-Right) and cultural institutions (Lower-Left). Moreover, the same kind of parallel is developed at length by Jürgen Habermas, in his *The Theory of Communicative Action* (HABERMAS, 1984 and 1987)[39]. As suggested, Habermas's contribution is essential for a more balanced approach to collective action which intends to address communication. By incorporating *speech act theory* into his model of social action, Habermas's offers a typology of social interactions that sets the analysis free from the exclusionary instrumental

[38] The personality types in case relates with the locus of control, degree of self-monitoring, a kind of "hurry sickness" personality and a sensation seeking type (BOONE et al., 1999).

[39] Habermas discusses the historic reconstruction of worldviews in the chapter 5 of his The Theory of Communicative Action (op. cited), in connection with Emile Durkhein and Herbert Mead contributions. The onto/phylo parallel is addressed at length in connection with Lawrence Kohlberg's model of the development of sociocognitive and moral reasoning (HABERMAS, 1990). See also Habermas's *Communication and the Evolution of Society* (HABERMAS, 1979).

conception of rationality that dominates the game theoretical and the behaviorist approaches to communication in collective action settings. Habermas's basic typology of social interactions is indicated in the Lower-Left quadrant, since it involves different subjective dispositions toward communication (strategic interaction and instrumental communication vis-à-vis dialogue and mutual understanding). I suggest that both Habermas's typology and Graves's ECLET conception enrich enormously our understanding on thrust, reciprocity and the formation of social capital (Figure 3.1).



Source: Author's configuration.

Figure 3.1 – Toward an integral approach of collective action.

The basic behavioral alternatives associated with plural subjective dispositions are indicated in the Upper-Right. There, I remind that such behavioral possibilities are connected with both individual/species biological foundations and individual economic endowments. As suggested above, despite

the implication on moral skepticism, the evolutionary approach has been moving toward the explanation of both the biological and social foundations of altruism and cooperative behavior, considering various dimensions and dynamic interactions taking place in complex socio ecological systems. The most important aspects considered in this regard are generically and tentatively indicated in the Lower-Right quadrant (Figure 3.1). Before reviewing these developments, and in order to highlight the pros and cons associated with the relaxation of the assumptions of the standard model of economic actor, I begin by presenting the conventional economic approach to the commons dilemma.

3.3. The neoclassical economic theory of collective action

In agreement with Gödel's prove of the inherent incompleteness of *all* logical systems (GÖDEL, 1931), one should expect to find a number of unproven axioms and important abstractions supporting the logico-mathematical formalism of neoclassical economics. As Jones and Wilson (1987) explains, Gödel proved that within any given branch of mathematics, there would always be some propositions that couldn't be proven either true or false using the rules and axioms of that mathematical branch itself. Accordingly, one might be able to prove every conceivable statement about numbers, for instance, within a system by going *outside* the system in order to come up with new rules and axioms. But by doing so one will only create a larger system with its own unprovable statements (see also HOFSTADTER, 1980).

As an inexorable trait of all logical systems of any complexity, such sort of incompleteness should not be taken as a defect. It is just the price to be paid for internal consistence, as is the very heart of the concept of holon. However, if one loses sight from this fundamental tradeoff, one may misguidedly take the very internal consistence as an sign of the sufficiency of the theoretical construct, which is just what it is not, and intend to extend its rigorously deduced conclusions to situations outside the scope of proper application.

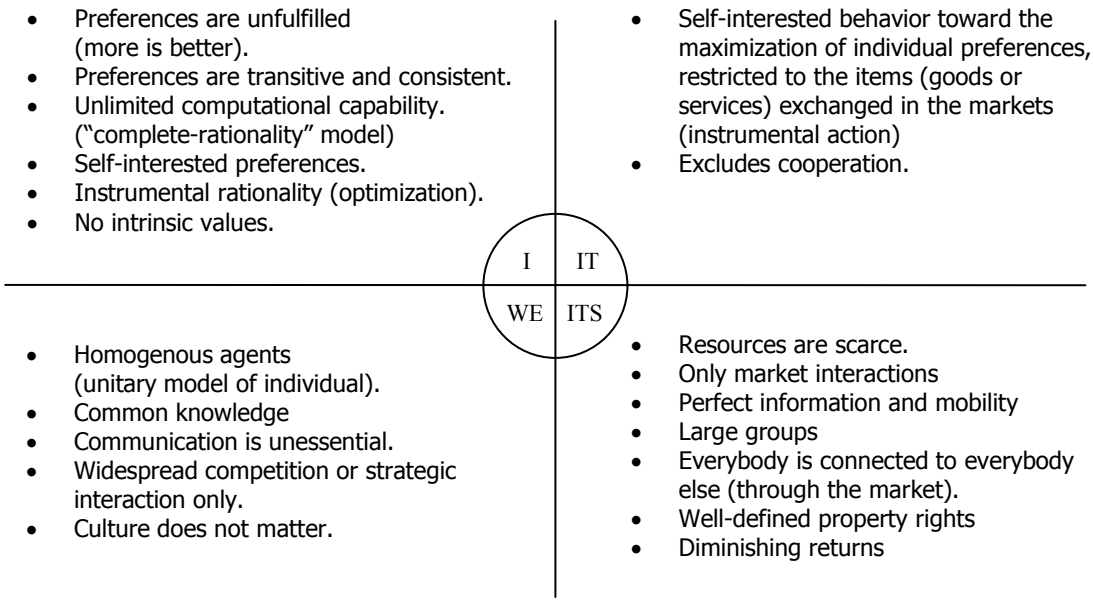
In order to address this aspect of the conventional analysis of collective action, I commence by mapping, in the Lower-Right quadrant of the Figure 3.2, the foundational assumptions regarding the circumstances of action that outline the *original* scope of economic theory, namely: resource scarcity and market interactions only. In addition, the conventional approach further abstracts the so-called transaction-costs; so that perfect information and perfect mobility establishes a system where everybody is connect with everybody else through the markets signals, which flow costless everywhere. Though usually viewed as simplifying assumptions, I note that these conditions have decisive implications for the plausibility, i.e., the internal consistence connecting the circumstances of action with the model of individual that shall fill up both quadrants above. Similarly, by implying costless transactions, the assumptions on information and mobility render unessential the account of social institutions other than property rights. In other words, the entire incentive structure affecting behavior is assumed to be imparted by the market signals. Diminishing returns is the most important technological assumption which assists equilibrium analysis. Again, while increasing return or positive feedbacks has had a long presence in economic analysis, its consideration within this general scheme is not an easy one (ARROW, 2000). As Arthur (2000) reports, while ideas that invoke some form of increasing return have become highly fashionable, as recent as the mid-1980s many economists still regarded increasing returns with skepticism. As suggested, this is a sign of misapprehension of the tradeoff involving completeness and consistence, since the very existence of positive feedbacks, as an occasion outside the original construct, is denied in order to protect the consistence of the theoretical system from the disturbing and unwelcome implications due to increasing returns. It should be noticed, as Arthur (2000) points out, that nonconvexities and positive feedback mechanisms are now central to modern theorizing in international trade theory, growth theory, the economics of technology, industrial organization, macroeconomics, regional economics, economic development, and political economy.

But resuming with the original framework (Figure 3.2), at the start individual endowments are taken as given and final income distribution ensues from the equilibrium conditions, as guided by an "invisible hand", that is, an abstract entity outside the systems as well. Group size defines three basic market structures: perfect competition, monopoly, and oligopoly. We shall discuss the implication of group size ahead; seeing as the traditional characterization of the collective action problems presumes that individuals are not able to coordinate (i.e. that group size is large enough to prevent coordination).

Accordingly, in the both upper quadrants (Figure 3.2), the conventional model of individual excludes cooperation. As Gintis (2000) puts it, the *Homo economicus* is self-interested, caring only about personal (or more broadly, familial) bundle of commodities, work, and leisure acquired. Also, the *H. economicus* is outcome oriented, caring about social interactions only insofar as they affect his final consumption and wealth. In these circumstances, rationality is exclusively instrumental and there is no room for the attribution of rewarding values to the actions itself. In this very restrictive context, preference satisfaction is equated to well-being. Most important, *H. economicus* comes to a choice situation with exogenously given and determinate preferences. Otherwise put, the model of economic actor presupposes, without argument, that agents come equipped with no matter what desires and beliefs prior to any practical interventions in the world (HEATH, 2001). Also, individuals' needs and wants are presumed to be unsatisfied. This is not to say that individuals are imagined as being insatiable, as indicated by the usual assumption of diminishing marginal utility. Notwithstanding, the analysis proceeds only in so far as needs and wants are unfulfilled, so that "more is better" is an essential assumptions concerning the subjective *state* of the individuals. Finally, the model assumes that economic agent has a rate of time preference that allows him to allocate consumption over time in a consistent manner, reflecting his welfare and his concern for the welfare of future generations. Such choice consistence requires unlimited computational capabilities, which is made reasonable in connection with costless transactions

and perfect information with regard to the circumstances of action (Lower-Right).

On the collective domain (Lower-Left), in so far as the actors are situated in the context of interdependent choice, the unitary cognitive foundation come together with the assumption that "more is better" to homogenize all expectations and establish "common knowledge". Consistently, such homogenous self-interested dispositions create a general "culture of competition" that renders communication unessential. The implication is that the richness of intersubjective relationship, mutual understanding and all its complexities (Figure 3.1) collapses into strategic interactions only – and this may well be equated to the assumption that *culture* does not matter (Figure 3.2).



Source: Author's configuration.

Figure 3.2 – The main assumptions of the neoclassical economics.

Given the framework outlined above, the standard economic approach to collective action dilemmas is usually addressed under the rubric of *externalities*. As indicated, the crucial feature defining an externality is that either the

transaction itself or some spillover from the economic activity does not pass through the market. Accordingly, an externality is viewed as a *market failure*, and "it is this lack of markets for externalities that causes problems" (VARIAN, 1993, p. 546). The problem externalities cause is that of inefficient provision or inefficient appropriation of the resource, and the failure of the market mechanism is attributed to the poor definition of the *property rights*. If one solves the problem on property rights one presumably solves the problem of externality. As Varian (1993, p. 562) puts it, "if everything that people care about is owned by someone who can control its use and, in particular, can exclude others from overusing it, then there are by definition no externalities [nor the inefficiencies associated with it]. The market solution leads to a Pareto efficient outcome".

But note how Varian's inference is reliant upon the particular conception of rationality outlined above. After all, it is not uncommon to come across the argument of private property to justify owner's supposed *right* to use her property whatever she pleases, which should not be taken as a vow of efficiency. As we saw, difficulties arise when theorists start to indiscriminately recommend political measures, such as the better specification of the property rights to address concrete social dilemmas, seemingly forgotten that the certainty or the determinacy ensuing from the theoretical predictions based on the neoclassical model ensues right from the incompleteness of its highly abstract framework. Of course, this does not mean that such theoretical predictions are useless or bear no relationship with the end results of actual social puzzles. On the contrary, a deeper problem arises, out from the *symmetry thesis*, when repeated empirical observations cohering with the theoretical predictions are taken as evidences of the appropriateness of the axiomatic structure in the "hard core" of the theoretical construction. That is, as if accurate predictions were the same as accurate explanations.

Notwithstanding, the premium for higher abstractions is more straightforward theoretical predictions, and clear theoretical predictions are of uttermost importance, not because they are evidences of accurate explanations, but because they provide the reference points against the "empirical anomalies"

can be confronted with in order to either change the theory or better delineate its limits of applicability, or both. Since the formal model of economic behavior indeed provides the theoretical predictions against with the experimental observations produced in this study will be confronted with, I will illustrate its formal reasoning, in connection with the appropriation dilemma, before moving toward the theoretical developments that change core assumptions regarding the subjective base and behavior of the human actor.

3.4. The formal model of the appropriation dilemma: an illustration

To illustrate the economic approach to collective action let us pose this problem in the original context of a common grazing land. Thus, consider an agricultural village in which the villagers graze their cows on a common field (the CPR). Following Varian (1993), we then want to compare two allocation mechanisms: (1) the private ownership solution and (2) the solution where the field is owned in common by the villagers and access to it is free and unrestricted.

In any case, how much product, say milk, a cow produces will depend on how many other cows are grazed on the land. Let $f(c)$ be the value of the milk produced if there are c cows grazed on the land. The value of the milk per cow is thus the average product, $f(c)/c$. Suppose further that a cow costs a dollars. For simplicity, we will neglect other production costs and revenues. Thus, in order to maximize the total amount of wealth, we set up the following problem:

$$\max_{c} f(c) - ac.$$

The first order condition establishes that the maximum profit will occur when the marginal revenue equals the marginal cost of production. Given our simplifying assumptions, in this case the marginal cost and the average cost of a cow are both equal to its constant price, a. So, we have that the first order condition is given by:

$$\mathrm{MP}(c^*) = a.$$

For if the marginal product of a cow were greater than a, it would pay to put another cow on the land; and if it were less than a, it would pay to take one off.

Provided the adequacy of the assumptions of the conventional "rational actor" model (Figure 3.2), this is the solution that would result if the grazing ground were owned by someone who could restrict access to it. For in this case the owner of the grazing ground would purchase just the right amount of cows to maximize his profits. Yet, if the grazing ground is a common field, the rent it may yield is not capable of being appropriated by anyone. In this setting, each village is more or less free to graze how many cows he pleases, and as far as he acts as assumed, when he decides whether or not to purchase a cow he looks at the extra value he will get [$f(c)/c$] and compares this to the cost of the cow, a. Therefore, it will be profitable to add a new cow until the average product equals the average cost a.

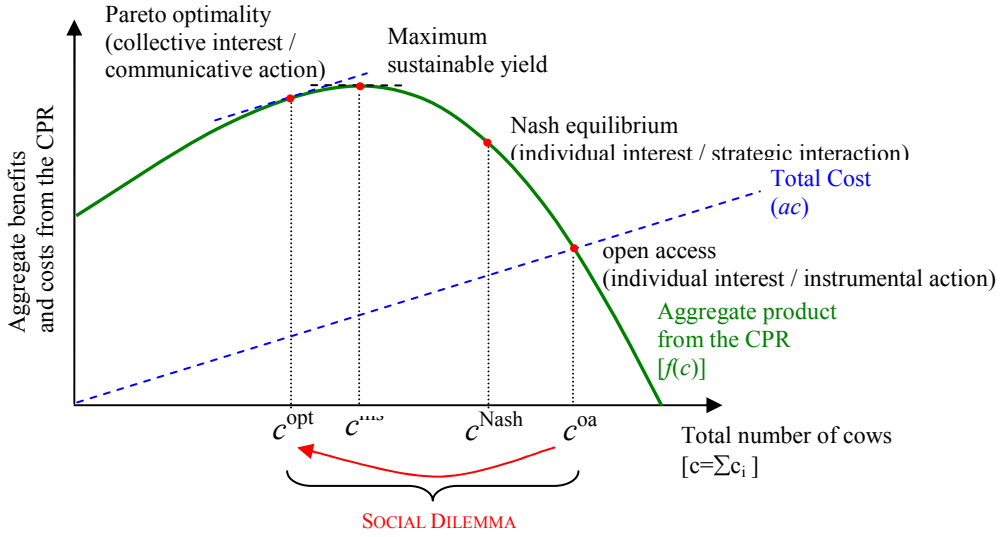
$$\frac{f(\hat{c})}{\hat{c}} = a\,.$$

As Varian (1993) notes, this is fine for him, but what has been left out of this calculation is the fact that his extra cow will impose a negative externality on the other villagers by reducing the output from all the other cows. Assuming that each individual villager has a number of cows that are negligible relative to the total number grazed on the common field, the model puts forward that each and every one village, since they are presumed to be all equally "perfect rational actors", will ignore the *social cost* of his own purchase, so that too many cows will be grazed on the common ground.

This argument is illustrated in Figure 3.3. Given the ecological characteristics of most renewable resources, we can assume a concave function, $f(c)$, that describes the flow of economic benefits that the villagers can perceive

from putting different number of cows in the common field. The concavity of the $f(c)$ function is assumed to reach a maximum given the maximum sustainable yield that the field can produce in terms of biomass, and beyond which productivity decreases to the point of exhaustion or collapse. We can also assume that the social benefic ensuing from the field is positive even when no cow is grazed on it [$f(c=0) > 0$], given that the common-land provides other types of ecological services even if no economic production happens (e.g. water regulation, pollinization, aesthetic services, nutrients recycling, erosion control, biodiversity conservation).

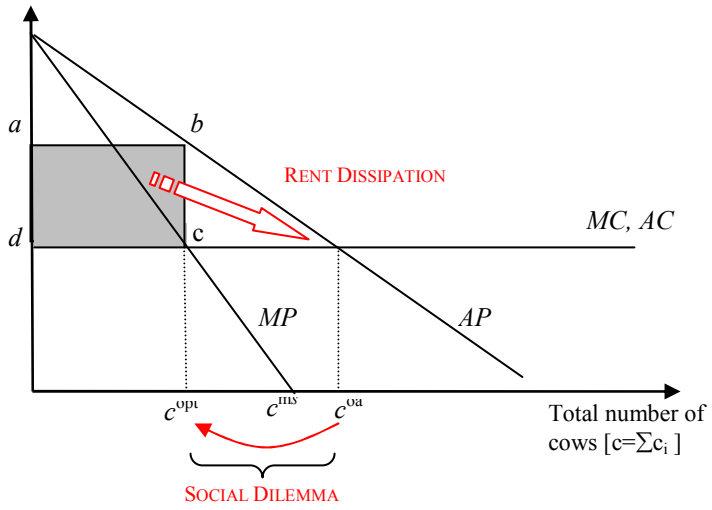
As assumed, for simplicity, the total cost of production is given by ac, which increases by the constant rate, a, with the number of cows, c, in the common field. The efficient output [MP(c^*) = a] is found at c^{opt}, where the function $f(c)$ has the same inclination as the line ac. We saw that this would be the solution if the land had a private owner and that it corresponded to the collective interest. However, as far as the entire abstract framework is tenable, the market failure is presumed to induce the villagers to continue to add cows to the common field. As already discussed, this process continue until $f(c)/c = a$ at c^{oa} (open access equilibrium), were no rents can be draw from the CPR.



Source: Adapted and modified from Cardenas (2000).

Figure 3.3 – The Common-Pool Resource (CPR) dilemma.

The rent dissipation can be better visualized in the Figure 3.4. There, the declining average productivity curve (*AP*) indicates that the output per cow decreases as more and more cows are grazed on the common land. The relationship between the average productivity (*AP*) and the marginal productivity (*MP*) of the grazing field is the same as that between average revenue and marginal revenue in imperfect competition theory. Thus the *MP* bisects any horizontal between the ordinate and *AP*.



Source: Adapted from Gordon (1993).

Figure 3.4 – The CPR dilemma.

Since, for simplicity, the fixed costs of production were neglected, and the cost of a cow was assumed to be unaffected by the number of cows purchased, the marginal cost (*MC*) and average cost (*AC*) are identical and constant, as shown in the Figure 3.4. If we assume that these costs include an opportunity income for the villagers, the shaded area *abcd* represents the maximum net economic yield from the grazing land, when the optimum number of cows, c^{opt}, are grazed. Gordon ([1954] 1993) observes that this area can be regarded as the rent yielded by the grazing field, which reflects the productivity

of that ground, not an artificial marked limitation. Yet, owing to the externality crated by the poor specification of the property rights, this economic rent is not capable of being appropriated by any villager in isolation. The predicted result is a pattern of competition among the villagers which culminates in the dissipation of the economic rent [$f(\hat{c}) - a\hat{c} = 0$] at c^{oa}.

As indicated in both figures, this intensity of grazing is over land's maximum sustainable yield, c^{ms}. In this example, the occurrence of a true ecological disaster will depend on the relation between $f(c)$ and ac. If the grazing costs are low enough, the number of cows in the ground may exceed the ecologically sustainable threshold, causing the collapse of the ecosystem – hence the tragedy of the commons.

Thus far, however, the analysis remained within the framework of strict rational choice and does not take into account the effect of strategic interactions that may ensue from the interdependence of the individuals' choices. As suggested, the internal consistence of doing so is sustained by the incompleteness of the framework, which presumes that the number of individuals in the commons is large enough to render the effect of the strategic interactions inconsequential vis-à-vis the theoretical equilibrium of open access, given that as the entry of additional cows in the common ground is free, the strategic conjectures about the behavior of others ultimately converge to the $f(c)/c = a$ condition discussed above. Notwithstanding, in the Figure 3.3 I kept Cárdenas (2000) indication, due to Cornes and Sandler (1983, 1986 and 1996) analysis, that strategic reasoning alone may prevent individuals from increasing the exploration of the commons to such high levels and shift the equilibrium outcome to the left of c^{oa}, such as at c^{Nash}, for instance. It should be noticed, however, that such step, configuring the essence of noncooperative game theory and rational expectation models, in fact weakens the internal consistence of the conventional framework, for it will require that the assumptions regarding the cognitive and computational capabilities of the human actor, which seemed just reasonably within the confines of market interactions only, be extended beyond the acceptable limits. Moreover, if one changes the action situation in order to

allow actual social interaction between the villagers, such as through communication opportunities, the reasonability of assuming only strategic interaction is totally lost, for the villagers will engage into another mode of rational action, namely, communicative action (HABERMAS, 1984 and 1997), to be discussed in the chapter four. In the Figure 3.3, however, I am indicating that a better result, eventually near to the optimum, may be attainable via communicative action, if the villagers have the opportunity to get to mutual understanding on how they should handle the situation in order to prevent "the tragedy of the commons" and assure the their collective interest. As indicated, the likelihood of successful communication is posited here, *ceteris paribus*, to be a function of the levels of psychosocial development of the individuals. The results of our empirical tests are presented and discussed tin the chapter six. We must now go ahead with the theoretical developments that support the consideration of distributed intentional states and personal inclinations, which will be then assembled according to the finds of the stage developmental psychology and that provide the basis for mutual understanding.

3.5. Rethinking the model of individual in economics: relaxing assumptions and approaching complexity

As done in the Figure 3.2, respecting the framework of the standard economic theory, I now map in the Figure 3.5 the most significant transformations observed in the economic approach to the commons dilemma, as it tags along improved representations of the human agent and relaxes restrictive assumptions regarding the action arena.

As Ostrom (2005) explains, once one decides to explore alternative theories and models of human behavior, the number of choices that the researcher has to make about: (1) the way that participants acquire, process, represent, retain, and use *information*, (2) the *valuation* that participants assign to actions and outcomes, (3) the processes or *choice criteria* that participants use for selecting particular actions or strategic chains of actions in light of their

resources, and thus their combination, is immense. She observes that alternative assumptions are not likely to be as clean and elegantly simple as those of rational choice models. As well, such a move toward inclusiveness must be paid with increasing ambiguity and uncertainty of the behavioral predictions ensuing from complex patterns of interactions. As Ostrom (2003, p. 63) puts it, "the kind of theory that emerge from such an enterprise does not lead to global bivariate (or even multivariate) predictions that have been the ideal to which many scholars have aspired". On the contrary, theoretical predictions ensuing from models of heterogeneous agents, for instance, are necessity complex, iterative, and conditional (MARWELL; OLIVER, 1993)[40].

Of course, if complexity gets to the point where no discernible pattern of behavior takes place, our attempt to model action situation possibly loses its main purpose. Ostrom then observes that one strategy adopted by many scholars (herself included) is not to change all of the assumptions at the same time when trying to model a particular type of situation. Accordingly, the entries in Figure 3.5 do not comprise the framework of any full developed and internally consistent theory of collective action, as it does in Figure 3.2. Rather, they represent a set of relaxations spread all over a number of different and some times incompatible approaches.

The problem is that these assumptions regarding information, valuation, and choice criteria are not quite separable as one would like. As in Simons's theory of "bounded rationality", the recognition that actual human beings have in fact limited capabilities to acquire, process, represent, retain, and use *information*, is just in the root of his substitution of the maximizing principle for the "satisfying" *criterion of choice*. Therefore, according to Simon's scheme (cf. SIMON, 1955 and 1982), in order to chose, "bounded rational" actors set an *aspiration level* which, if achieved, they will be *happy enough* with, and if they don't, try to change either their aspiration level or their decision (see Figure 3.5). However, the idea that the actors may change either the decision itself or the

[40] See also Selten (1990, 1991), Güth (1995), Kreps (1990), Rabin (1993), Palfrey and Rosenthal (1988) and Abbink et al. (1996).

aspiration level creates manifest problems for the purpose of theoretical prediction of behavior[41].

The link with valuation is subtler, since it requires, first, a change in the common assumption according to which what render the prediction of the human behavior difficult is the presence of uncertainty. At this point, Heiner (1983) offers a path-braking approach according to which it is just the presence of uncertainty that limits behavioral flexibility and is thus in the origin of predictable behavior. Heiner's approach begins relaxing the conventional assumption on information processing by acknowledging Simons's "bounded rationality" suggestion. Accordingly, he posits a competence-difficulty gap for decision-makers which is a function of perceptual powers of the agent and of the complexity of the decision problem to be solved. He then develops a simple reliability condition which determines whether agents with a given competence-difficulty gap will benefit from the flexibility to select potential actions (DRIVER, 1992). In this fashion, Heiner put forward that observed regularities of behavioral can be fruitfully understood as "behavioral rules" or heuristics that arise *because of uncertainty in distinguishing preferred from less-preferred behavior*. As he explains, such uncertainty – which connects with the assumption about valuation – requires behavior to be governed by mechanisms that restrict the flexibility to choose potential actions, or which produce a selective alertness to information that might prompt particular actions to be chosen. These mechanisms simplify behavior to less-complex patterns, which are easier for an observer to recognize and predict. Conversely, in the special case of no uncertainty, the behavior of perfect informed, fully optimizing agents responding with complete flexibility to every perturbation in their environmental would *not* produce easily recognizable patterns, but rather would be extremely difficult to predict (see also FOSTER; YOUNG, 2001). Thus, according to Heiner, it is in the limits of maximizing that we will find the origin of predictable behavior.

[41] As well as for the empirical disproving of the "bounded rationality" hypothesis, as far as the character of the individuals' preferences is private.

The implication Heiner advances is that, contrary to the dominant evolutionary justification for assuming agents are able to optimize[42], evolutionary selection processes will in general *not* produce approximations to optimizing behavior. Rather, they will tend to produce behavioral regularities or rule-governed behaviors that will evolve in the extent that uncertainty prevents agents from successfully maximize (HEINER, 1983, p. 561). Following the formal development of his approach, Heiner then presents a number of applications raging from an explanation of the instinctive behavior to punctuated dynamics for scientific inquiry, from the formation of social judgments to the reluctance to insure against rare disasters, from a clear prediction of the law of demand to the emergence of social institutions.

Of particular importance for our present concern is Heiner's insight on how uncertainty may generate *hierarchical structures of increasing flexible rules*. Among a number of applications discussed in his paper, he mentions both the significance of hierarchical structures in the design of cognitive and related learning processes, and Kohlberg's developmental model of moral reasoning. As evident, Kohlberg's model gets right into the core of the assumption about valuation. Further, as we shall see in the next chapter, such "hierarchical structures of increasing flexible rules" configure just the essential attribute of Clare Graves's theory on how value systems cyclically emerge turning the relative weight of self-regarding vis-à-vis other-regarding preferences. Finally, in the chapter four, we shall see that the same hierarchical conception is present in virtually all psychological developmental models.

Concerning learning processes, Heiner suggests that the hierarchical structure may be associated with resistance to knowledge not built up in recursive stages. This interpretation agrees with the so-called "cognitive institutionalism" (MANTZAVINOS et al., 2003), which provides a key ingredient of the conceptual framework I offer in this study. Accordingly, at the same time as the proponents of the cognitive institutionalism take on a

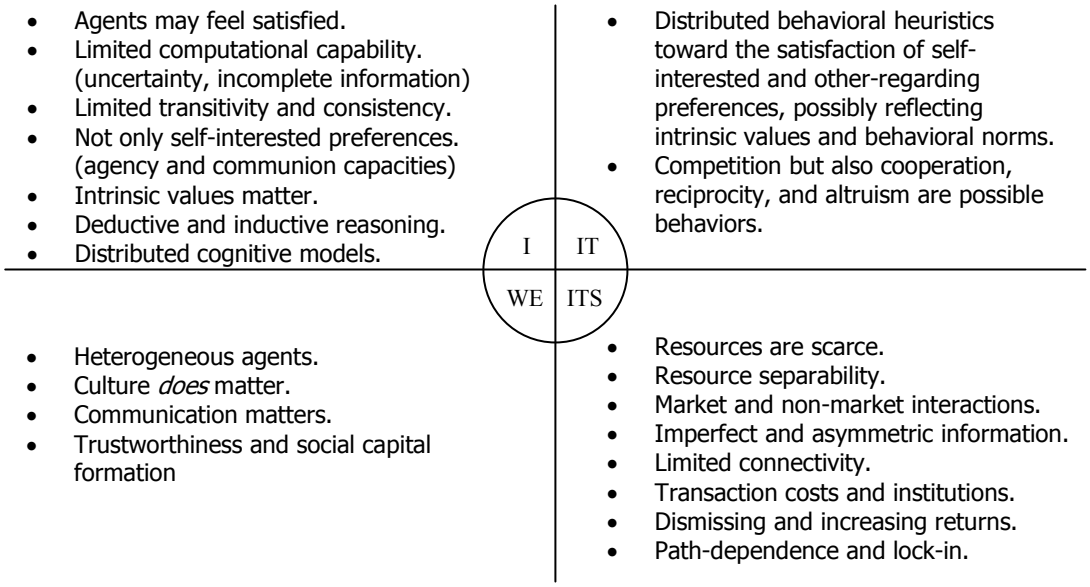
[42] See in particular Armen Alchian (1950) and also Sidney Winter (1964, 1971), Jack Hirshleifer (1977), Richard Nelson and Winter (1974) and Gintis (2003).

conception of individuals' mental models as flexible knowledge structures – which gradually evolves during the cognitive development to organize our perceptions and keep track of our memories – along the path of social learning they usually turn into relatively stable structures, resistant to abrupt changes, and shaped by the progressive involvement of parallel emotional adaptations.

The authors refer to such relatively stable mental structure as *belief systems*. Yet, the framework of "cognitive institutionalism" lacks a full developed psychological theory to arrange such belief systems or cognitive props into a hierarchical structure of increasing flexible rules. Fortunately, it is precisely this gap that the ECLET conception can profitably fills up. Also, while the cognitive institutionalism recognizes that no externally imposed conception of rational choice should work for interpreting the evolution of shared mental models and belief systems over time – or possibly just because of this – its elucidation of the appearance and change of informal institutions ends up adhering to the doctrine of spontaneous emergence of order. As the authors say, "The spontaneous emergence of informal institutions is a process of innovation and imitation that takes place in a social group that is learning collectively. Individuals respecting conventions, following moral rules, and adopting social norms cause (as an unintended outcome of their action) the emergence of social order" (MANTZAVINOS et al., 2003). As well, in the short-run, shared mental models are posited to ensue from repeated communication opportunities, as people in their sociocultural environment continually try to solve their problems (DENZAU; NORTH, 1994). In this vein, the cognitive institutionalism sees the informal institutions – as just the interior side of the shared mental models – as endogenous to a community, which opens the door for the understanding that culture matters (see Figure 3.5).

Yet, the adherence to a "pragmatic conception" of mental models naturally renders communication into an instrumental form of social interaction, and possibly the emergence of language into an unintentional result of instrumental action. As we shall see in the chapter four, it is here that Habermas's theory of communicative action can fill the gap by providing both a

more explicit typology of social interactions and a more convincing account of the emergence of social order in connection with the emergence of language – possibly the most fundamental of the human institutions. Moreover, as already indicated, Habermas's reconstructive account of the evolution of cultural institutional and worldviews is explicitly connected with Kohlberg's hierarchical model of the development of the sociocognitive capacities and moral reasoning, which is just a different stream of the same river of interior development described by Graves. In fact, the compromise with the instrumental form of rationally obscures the emancipatory power embedded in human true creativity, as it reflects not only human attempt to solve practical problems but also take in for questioning our human nature and the meaning of this existence. Once more, the possibility of changing the focus of concern and expenditure of energy from subsistence toward existential issues is itself a matter of hierarchical structuration of the human development.



Source: Author's configuration.

Figure 3.5 – Approaching complexity in the four-quadrants.

In summary, I suggest that the reason why Ostrom (2005) sees a "messiness" in the alternative theoretical structures explaining human behavior, pointed by her as a key factor explaining the hesitation of some theorists to adopt more realistic assumptions to model the human agent, is just that her efforts, as well as those of most scholars investigating collective action problems, is directed toward the attempt of predicting and explaining behavior (and then the adequate model of individual) only as instrumental rational reactions to the variable circumstances of action. While it is certain that alternative behavioral theories are marked by many differences and contradictions, I have already commented on the significant convergence ensuing from the consideration of the holarchical structure of human psychological development, and, in particular, how such a theoretical convergence draws into two qualities of most significance for the analysis of collective action dilemmas, namely: lessening of egocentrism and increasing moral embrace (WILBER, 2000b). Yet, as we shall see in the sequence, Ostrom's stance ensues right from her methodological discipline and the primary focus of the IAD framework on the institutional change, together with its subject-object perspective and methodological compromise with the instrumental conception of rationality. As indicated, this leaves out of the account the fact that the development of the human capacities to handle information, the kind and span of human valuations and concerns, as well as agents' standards or criteria of choice, all of them obey their own laws of transformations, so that they cannot be simply inferred from the structure of action environment, and then taken as more or less adequate *assumptions* to model the human behavior in such circumstances. I argue that such methodological discipline ultimately undercuts the original and quite significant emancipatory message ensuing from the IAD framework. This drawback, nonetheless, can be fixed through the combination of the IAD framework with some theory of adult psychosocial development, as put forward in this study.

3.6. A half step toward freedom: the emancipatory and the automating faces of the IAD framework

The noteworthy message emerging from the extensive research documented in the seminal *Governing the Commons*, by Elinor Ostrom ([1990] 1997), is that people are trapped by the prisoner's dilemma-like situations only if they treat themselves as prisoners by passively accepting the suboptimum strategy the dilemma locks them into. But if they try to work out a contract with the other players, or find the ones most likely to cooperate, or agree on rules for punishing cheaters, or artificially change the incentive ratios, they can create some self-organizing solution for their collective action problem that benefits them all.

This is a clear emancipatory message that challenged the accepted wisdom about institutions for collective action, as discussed in the previous chapter. Ostrom's founding role in the advancement of the IAD framework grew from her careful inductive examination of empirical studies on common-pool resource (CPR) management, and insistence on interdisciplinary analysis. As Ostrom et al. (2002) explain, the character of *framework* comes from the recognition that we presently have not a sufficiently developed theory of collective action to explain the empirical regularity that, under certain favorable circumstances and dispositions, groups of people do are able to solve their social dilemmas profitably and to sustain alternative social arrangements for extend periods of time. Up from such acknowledgement, the IAD framework provides a conceptual structure which helps analysts to organize information in order to better understand and discriminate the most important contingencies affecting the likelihood of emergence of successful self-organized solutions addressing collective action dilemmas. As such, the IAD framework is not restricted to the use of one particular theory or model of human behavior. On the contrary, the framework furnishes the foundation for investigating the predictive power of complementary or competing theories and models according to what appears to

be relevant to the context of each particular decision environment (OSTROM et al., 2002).

Nonetheless, having its roots in the classic political economy, neoclassical microeconomic theory, institutional economics, public choice theory, transaction-cost economics, and noncooperative game theory, the IAD framework remains firmly within the limits defined by the instrumental conception of rationality and the tradition of the methodological individualism. Though indeed open to experiment with alternative theories or models of the individual agent and their social interactions, the primary focus of the IAD framework rests not on the inquiry into the human actor by itself but on the situational variables (Lower-Right) that could *predict* the likelihood of such actors come out with successful and enduring self-organized solutions for their collective action dilemmas in the context of complex socio-ecological systems (SES). A passage from Ostrom (1997, p. 25-26) clears up the point:

> As an institutionalist studying empirical phenomena, I presume that individuals try to solve problems as effectively as they can. That assumption imposes a discipline on me. Instead of presuming that some individuals are incompetent, evil, or irrational, and others are omniscient, I presume that individuals have very similar limited capabilities to reason and figure out the structure of complex environments. It is my responsibility as a scientist to ascertain what problems individuals are trying to solve and what factors help or hinder them in these efforts. When the problem that I observe involve lack of predictability, information, and trust, as well as high levels of complexity and transactional difficulties, then my efforts to explain must take these problems overtly into account rather than assuming them away (OSTROM, 1999, p. 25-26).

As such, one of the most significant products of the IAD framework, ensuing from cumulative information on successful and unsuccessful management of CPRs encompassing different circumstances of action, is the conception of eight design principles that was found to support successful self-organizing initiatives. However, in this context, the IAD framework tends to assume a kind of *situational imperative* that, in a sense, represents a step back with its original emancipatory implication. As Ostrom (2003, p. 25) says *a propos* of the effectiveness of the convention model of rational actor in predicting behavior in highly structured and competitive environments, "the *situation* turns the individual into a determinate, calculating machine" (my

italics). She draws in Alchian's (1950), as well as in Gode and Sundar's (1977) contributions to argue that it is the structure of markets that leads participants to make efficient choices or act as "rational egoists" rather than the assumptions made in economic theory about the nature of individual rational choice. Ostrom also mentions a metaphor used by Latsis to summarize the point. According to her quote, Latsis (1972, p. 210-211) puts that: "To say that a seller under perfect [market] conditions deliberately chooses a course of action to maximize profits is analogous to saying that a member of the audience is maximizing if he runs out of the single exit available to him in a burning cinema". One should notice, first, the evolutionary paradigm underlining all the above reasoning. I have already introduced Heath's restrictive comments in this regard, and indicated that we shall return to the point more ahead in this chapter. Also, I must delay a comment on Latsis's comparison after we have corrected a misstep in how the IAD framework conceives the holarchical structure of the action situation.

Notwithstanding, in the way outlined above, the insufficiency of relying exclusively in the conventional model of rational actor is justified in the measure that highly structured and competitive environments, where perfect information may possibly be assumed, are not the only situations facing individuals that are of interest to institutional analysts (OSTROM, 2005). Accordingly, as Ostrom explains, some situations generate only partial information, whereas the distribution of the information may be asymmetric. Situations also differ in regard to the extent of repetition and whether the same individuals continue to interact with one another over time or whether interactions are largely with a continuous stream of strangers. As indicated, the focus of the IAD framework is on the issue as how all of these situational variables affect choice, and it is when the rational choice models fail to predict outcomes that other theoretical tools are seen as needed. At this point, and even if what follows is not necessary implied in the above reasoning, one should remind that good predictions are not the same as good explanations. Accordingly, the fact that the conventional model of individual adequately predicts the average final outcome of collective action in highly structured and competitive environments does *not* mean that the

individuals are in average behaving as the model assumes. I am insisting on this point because the main concern in this study is to expose, first, how individuals behave differently when set within the very same institutional settings and structure of objective incentives, and, second, to show that such differences can be explained by the internal laws of adult psychosocial development, as represented by Graves's theory.

As pointed out, the IAD framework does not hesitate in challenging the conventional assumptions regarding the standard of valuations that agents may assign to both outcomes and actions. As Ostrom puts it:

> The reason we can characterize participants as rational egoists in an open, competitive market is because of the institution *not* because all of the participants are narrowly selfish! The same individuals who energetically pursue profit maximizing strategies from 8 a.m. to 5 p.m. every workday, may also volunteer several evenings a month on neighborhood projects, contribute substantial funds to diverse charities, regularly vote, and be known to friends and co-workers as kind, considerate individuals who always do more than their share of any team project. These individuals are also likely to have biases in the way the world is interpreted and a variety of other human failings (OSTROM, 2005, italics in the original).

Ostrom adds that, in other to model human behavior and animate institutional analysis, one may be further confronted with the need to examine how individuals view risk, uncertainty and information asymmetries and how they react to the actions and perceived attributes of other participants. Still, whether or not, and in which extent such diversity is to be considered is seen, once again, as a function of the situation to be modeled. Therefore, assuming the separability of the set of assumptions describing the human actor, Ostrom suggests that in stable and repetitive situations where intrinsic values are important, one may retain the assumption regarding full information about the structure of the situation, while focusing primarily on the types of values of importance to individuals and the resulting strategies they select.

But as herself observes, "When other-regarding preferences and intrinsic values are assigned to outcomes and actions, the situation is one of incomplete, rather than, complete information because other players cannot know exactly how an individual is valuing these actions and outcomes" (OSTROM, 2005).

This fact clearly entails the distinctive quality of the collective action dilemmas, as these are situations where intrinsic values do matter, and where people will certainly weigh alternative actions and outcomes differently. Seeing as, moreover, most of the relevant situations involve complex socio-ecological systems – plenty of uncertainties and transactional difficulties – perfect information is not a sensible assumption to approach our subject matter. Finally, the attempt to predict behavior in such settings by assuming that the agents are rational optimizers requires endowing them with such a level of computational capability that we simply should not, as far as we want to bestow some explanatory quality to our model, instead of conceiving it only as distinct analytical or decision tools.

Therefore, none of the assumptions describing the conventional model of individual should be retained to understand and to address collective action dilemmas productively. Still, if we want to advance predictions in such complex settings (imperfect information, heterogeneous valuations, and limited computational capabilities), we should not take human behavior as mere instrumental reactions to the situational variables. Hence, as suggested here, the way out of the puzzle requires acknowledging, first, that human behavior is not regulated by a flat conception of instrumental rationality, but evolves according its own internal rules of development, and, second, that uncertainty and rule-governed behavior – not maximization with no uncertainty – is what is in the origin of predictable behavior, as put forward by Heiner (1983). If so, this should bring to an end the long tradition and ever renewed efforts to justify the "as if" maximization thesis on the basis of the "survival of the fittest" principle. As Heiner observes, the claim that any observed behavior is consistent with the maximization of *some* preference function is probably incapable of either theoretical or empirical disprove (see also BOLAND, 1981). Indeed, I would say that, by and large, such effort is not part of a theory about empirical phenomena. Rather, it is part of a theory on how to preserve the "hard core" of neoclassical microeconomic research program, and how to extend its influence as far as possible – eventually beyond the limits of its meaningful applicability. The

claim, as usual, is that such prediction concerning actors' adjustment to the circumstances of action is well supported by empirically established predictions. As Ostrom remind us:

> Armen Alchian (1950) demonstrated long ago that those who do not behave as rational egoists in an open, competitive, market will not be present in any significant number once demand and supply have led to an equilibrium. More recently, Gode and Sundar (1997) demonstrated that it is not even necessary to assume that all individuals are rational egoists – even at equilibrium – to establish efficient markets. Rather, they show that the efficiency of market exchanges derives from the set of rules constituting a market, rather than the sophisticated calculation of the participants. In their model, they show that "zero intelligence traders", reach efficient outcomes when seven essential rules constrain the actions of buyers and sellers and not otherwise.

But the problem is that neither the market rules – or other humanly devised institutions – are creations of "zero intelligence" agents, nor "zero intelligence" agents are creations of haven fallen institutions. Further, no "zero intelligence" judge or enforcer would unease him- or herself to ensure these rules are followed. Therefore, even if Gode and Sundar's approach is meant just to set the microeconomic argument free from the burden of unrealistic assumptions involving the human actor, it nonetheless entails the same inversion in the order of the genuine activity, which characterizes the *situational imperative* that emerge clearly not from Ostrom's view of the problem, but from her methodological discipline and style of approaching it. As suggested in the previous chapter, the issue is connected with the standpoint of moral skepticisms, underlining the view of the instrumental conception of rationality as the only form of rationality. As we shall see further ahead, it is also just an expression of the most fundamental environmental imperative that characterizes the mainstream evolutionary paradigm. Finally, it has also to do with the ordinary misconception of the holarchical structure of reality, which the IAD framework reproduces in its structuration of the action arena. Therefore, I must address this issue prior to our review of the general schema and main contributions delivered from the IAD framework.

3.7. A correction of perspective on the structure of the action arena

As we shall see in a moment, the central unity of analysis of the IAD framework is the *action arena*. Ostrom draws in the notion of *holon* to expose the nested structure of the action arenas, so that, as she puts it, "the IAD framework can be presented at scales ranging from exceedingly fine-grained to extremely broad-grained". As a result, the analyst trying to advance institutional analysis may possibly face from the start the puzzle of identifying the proper level or scale of analysis, since, as Ostrom explains, "the relevant theoretical concepts needed to understand phenomena at one level do not necessarily scale up or down", for the "parts used to construct a holon are frequently not descriptive of the holon they have created".

Nonetheless, as just mentioned, the IAD framework misarranges the holarchical structure of the action arena out of a failure to recognize the fourfold aspect of every holon. Accordingly, as did Popper, Ostrom takes the individual and the collective aspects as two distinct *levels* in the holarchy of the action arena, where the collective is set "on top" of the individual. In Ostrom (2005) words:

> Human decision making is the result of many layers of internal processing starting with the biophysical structure, but with layers upon layers of cognitive structure on top of the biophysical components. Further, many of the values pursued by individuals are intrinsic values that may not be represented by external material objects and their presence and strength are important parts of the individual to be examined. Building *on top* of the single individual are structures composed of multiple individuals – families, firms, industries, nations, and many other units – themselves composed of many parts and, in turn, parts of still larger structures. What is a whole system at one level is a part of a system at another [higher] level (my italics).

Note that the cognitive structures are, first, correctly set "on top" of the biophysical *components*. This view agrees with Figure 5.1, as far as the cognitive structures can stands for the broader concept of noosphere (the domain of human mind). Also, the addition regarding the intrinsic values as possibly having no adequate representation in the material gains or payoff matrix expresses the plain acknowledgment of the significance of the interior domain of human concerns as

entailing more than self-interested preferences over goods and services acquired, and where cognition takes its place together with about two dozen of other developmental lines, which we will address in the chapter four. Further, while still restricted to cognition (as a sign of the commitment with instrumental conception of rationality), Ostrom remark on "layers upon layers of cognitive structures" is still closer to the findings of the stage developmental psychology that I bring to joint the IAD framework in this study.

The conceptual error is that, as we saw, the structures composed of multiple individuals are not on top of the single individual, but are rather the collective aspect of the Human Holon, which is located in the axis of the AQAL matrix. The identification of what may be on top of the single individual requires first the acknowledgment that the individual is not a holon, rigorously speaking, but just the individual expression of the Human Holon. Moreover, it would require the inquiry into the exterior (material) expressions of the subtle and causal bodies. This would drive us to Wilber's discussion on the spectrum of subtle energies (WILBER, forthcoming, Excerpt G, Part II), which is beyond the scope this study. Notwithstanding, in the chapter four I will present Wilber's Integral vision on the Human Holon (see Figure 5.1) in connection with our discussion on the sociocultural evolution. Then we shall see that nations are not on top of industries, as in Ostrom's holarchy. Rather, industrial nations are on top of horticultural societies, or corporate states are on top of feudal emperies, as social structures that transcend-and-include their own past, not the individuals, who are *members* and not component parts of those social structures. Moreover, these social faces (Lower-Right) of the Human Holon co-evolve with both the individual capacities (both Upper quadrants) and the collective cultures and worldviews (Lower-Left).

This correction of perspective is important in order to highlight the insufficiency of approaching the collective action only from the perspective of the methodological individualism – since the individual has *no* prerogative over the collective, seeing as neither the individual can be conceived apart from the collective (we always need two individuals, the parents, to create a third

individual, the newborn), nor a collection of individuals can exist with no individual at all. In this regard, it should be noticed that Ostrom' holarchy, above, inherits the IAD framework's special concern with the institutional structure affecting the likelihood of the emergence of sustainable self-organized rules to handle the CPRs dilemmas. In this sense, however, the fact that the constitutional level is "on top" of the local rules-in-use, for instance, comprises just an institutional hierarchy (implying authority), not a holarchy. Hence, the fact that local self-organized rules that affect collective action situations are submitted to the constitutional rules "on top" does not mean that the constitutional level presupposes, for its sustainability as a higher level holon, the existence those local regulations "below". This is not to say that the nested structure or the levels of institutional analysis is mistaken; only that such levels do not comprise holarchies. Further, the criteria put forward by the IAD framework to classify rules is kept purposefully close to the categories that game theory uses to specify the game situation (i.e. rules apply to participants, positions, actions, outcomes, information, control, and costs/benefits), so as to cohere with the methodological individuals underlining the IAD approach to institutional change (cf. OSTROM, 2005, chap. 7).

As well known, the methodological individualism of the so-called new institutional schools of economics turned the "old" institutionalist position on its head – using the instrumental rationality that characterizes individuals' choices in the neoclassical economics to explain history, social relations and the emergence of institutions rather than the other way around, as the old institutionalists proposed, of using history and societal concerns to explain economic behavior, structures and patterns. Also, the sociological (collectivistic) approach draws attention to the essentially dialogical nature of the justification practice involved in the attempt to influence others regarding the appropriateness of the rule-governed behavior and other deontic constraints. This is of course central to approach the role of communication in solving collective action problems. Therefore, bearing in mind the implications of such correction of perspective on the structure of the action arena, let us see the general schema of the IAD

framework, before we can turn to examine its design principles and behaviorist approach to cooperation.

3.8. The general schema of the IAD framework

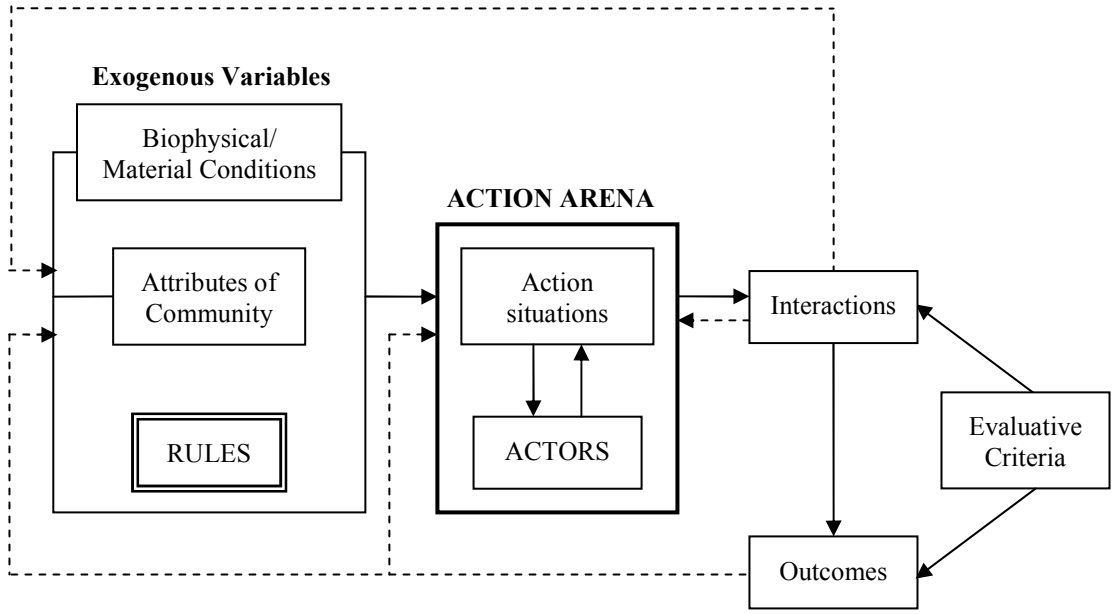
The general schema of the IAD framework is presented in Figure 3.6, from Ostrom (2005). As indicated, in the center of the analytical schema there is the *action arena*. Assuming the methodological individualism and a subject-object perspective, the IAD framework subdivides the action arena into two fundamental components: (1) an *action situation* component (the object) and (2) an *actor* component (the subject). As Ostrom, Gardner and Walker (2002, p. 28) put it, "Action situations refer to the social space where individuals interact, exchange good and services, engage in appropriation and provision activities, solve problems, or fight (among the many things that individuals do in action situations)". Yet, as pointed out, what actors engaged in instrumental action cannot do is partake in the subject-subject form of interactions, such as communicative action, for instance. Notwithstanding, the IAD framework puts forward that a minimal *action situation component* is characterized using seven clusters or vectors of variables: (1) participants, (2) positions, (3) actions, (4) potential outcomes, (5) transformation functions (that maps action into realized outcomes), (6) information, and (7) payoffs (costs and benefits assigned to actions and outcomes).

The proponents point out that many of these components are themselves relatively complex, as they are further constituted by combinations of physical, cultural and rule-ordered attributes, which are usually assumed as exogenous variables, for the purpose of a specific level of analysis (Figure 3.6). Yet, as the authors point out, many institutional analysis focus precisely on the factors that affect the action arenas rather than with the arenas themselves. Therefore, the set of elements compounding the action situations can describe an immense variety of action arenas including markets, CPRs, hierarchies, and legislatures. Ostrom et al (op. cited) point that, when seen as fundamentally different "pure types" of

situations, markets, hierarchies, and collective action situations each is presumed to require its own language and explanatory theory. Accordingly, scholars who attempt to explain behavior within markets may rely exclusively on neoclassical microeconomics theory, whereas scholars who attempt to explain behavior within hierarchies may rely exclusively on political and sociological theory. Finally, scholars who attempt to explain behavior in collective action environment may rely exclusively on noncooperative game theory. The authors recognize that such a view "precludes the development and use of a more general explanatory framework that, together with its constituent theories, could help analysts make institutional comparisons and evaluations" (OSTROM et al., 2002, p. 28). However, the game theoretical categories, the instrumental conception of rationality, and the methodological individualism that comprise the IAD framework, despite its openness for theory comparison and experimentation, still remains by and large speaking a very particular language. In particular, that of the subject-object perspective and the situational imperative I have insisted about.

Accordingly, respecting the *actor component*, the framework distinguishes four cluster of variables about which assumptions must be made in order to produce a model of predictable behavior in the action situation, namely: (1) the *preferences valuation* that actors assign to potential action and outcomes; (2) the individual *information-processing capabilities* mapping the way individuals acquire, process, retain, and use knowledge contingencies and information; (3) the *selection criteria* actor use for deciding upon particular course of action; and (4) the *resources* that an actor brings to a situation. As indicated, the IAD framework is *not* compromised with any particular set of assumptions regarding each of the above dimensions describing the actor component, while they are presumed to be separable or independent. We have already disused this issue in connection with Heiner's (1983) contribution and my concern on the "methodological imperative of the circumstances". I would like to reiterate that my point is neither to negate the pressure of the circumstances nor to suggest the useless of the careful consideration of the

situational variables in order to envisage how the actors should be modeled. The only point, once again, is that this way of proceeding excludes the behavioral restrictions imparted by the dynamics of interior growth. In any case, the interplay of all these elements compounding the action arena defines a particular pattern of interactions (including ecological and social interactions) that, in turn, produce final outcomes. Once more, I remind the significance of Habermas's typology of social interactions, so that the "interactions box" in the Figure 3.6 might be empowered with *communicative action*.



Source: Adapted from Ostrom (2005).

Figure 3.6 – The general schema of the IAD framework.

Evaluative criteria are used to judge the performance of the system by examining the patterns of interactions and outcomes. The illustration (Figure 3.6) suggests that these evaluative criteria are settled exogenously, that is, by the analyst. As described above, the criteria of valuation guiding actor's selection of alternative actions are part of the actor component itself. Clearly, *these* are the relevant criteria to understand how the actors will possibly change their choices

over time, given the feedbacks from the actual outcomes (doted lines). I highlighted the "rule box" in order to indicate that the IAD framework is particularly focused on the inquiry on how such feedbacks and interaction patterns may foster the emergence of self-organized institutional solutions to regulate the collective action problems taking place in the action arena. In this way, the effort is direct toward the attempt to *predict* how the agents are supposed to behave, depending on the contingencies of action.

Note that here is the heart of the puzzle facing scholars interested in advancing predictive models of human behavior, once they have acknowledged that individuals may: (1) interpret the actions situation and value outcomes differently, and (2) attach intrinsic values to actions itself. As Ostrom (2005) puts it, "How can we understand choice processes when participants may each utilize different models of the situation?" Or still puzzler, as Frohlich and Oppenheimer (2001) ask over, "How can we understand person's choices if, when confronting with a given environment, she can have multiple representations, each of which is associated with different values?" The answer, as pointed out, passes thorough the recognition that human behavior ensues from *hierarchical structures of increasing flexible rules*, which we shall explore in the next chapters. Prior to that, let us see what the IAD framework has to offer without this.

3.9. Administrating nihilism beyond panaceas: a growing consensus toward eight design principles of successful institutions

One of the distinctive qualities of the IAD framework is how its careful scrutiny of the situational variables affecting the likelihood of successful initiatives to manage common-pool resources works as an antidote against the perverse and extensive use of policy panaceas in misguided efforts to address collective action problems and make social-ecological systems (SESs) sustainable over time.

Respecting the attributes of physical world (Figure 4.6), a number of ecological characteristic of the resource such as renewability, susceptibility to

depletion (fragility), the length of time for resource become mature, the distribution of valued flow unities throughout the resource, the visibility of the resource, and whether or not the resource is stationary or storable, to mention the most relevant, all these may influence the likelihood of a satisfactory solution in many complex ways.

In Ostrom et al. (2002) the interested reader find a detailed examination of the implications of stationarity and storability or the resource on both appropriation and provision dilemmas. These characteristics are closely related with the degree of information users may have about the availability of resource along time. Bardhan (2000), in turn, examine, among other variables, the significance of topography and the geographic position of the village on users' ability to cooperate to the maintenance of infrastructure and adherence to allocation rules in 48 irrigation communities is South India.

According to Ostrom (1999), a growing consensus on variables enhancing self-organization is emerging among scholars familiar with the results of field research. Concerning the attributes of the resources, she mentions the following as the most important (see also OSTROM, 1992b, p. 298-299; BALAND; PLATTEAU, 1996, p. 286-289):

1. *Feasible improvement* must be possible, that is, the resource is not at a point of deterioration such that it is useless to organize or so underutilized that little advantage results from organizing;
2. *Reliable and valid information* about the general condition of the resource must be available at reasonable costs;
3. Sufficient *predictability* as to the availability of resource units over time, and,
4. The resource should be sufficiently small, given the transportation and communication technology in use, so that users can develop accurate *knowledge of external boundaries and internal microenvironments*.

As for the attributes of the community, which necessarily connects with that of individual users (the actors) and help to define the patter of interactions among them, the main factors affecting chances of long-surviving, self-organized resources regimes, as organized by the IAD framework (OSTROM, 1999), are:

1. *Salience*, i.e., the users must be dependent on the resource for a major portion of their livelihood or other variables of importance to them;
2. The users must have a *shared image* (or a common understanding) of the resource and how their actions affect each other and the resource;

3. The users must have a sufficiently *low discount rate* in relation to future benefits to be achieved from the resource;
4. *Shared interest*, in the sense that the users with higher economic and political assets are similarly affected by a current pattern of use, so that the different interests are not too conflicting;
5. The users ought *trust* each other sufficiently to keep promises and relate to one another with reciprocity;
6. *Autonomy*, in the sense that the users should be able to determine access and harvesting rules without external authorities countermanding them, and finally,
7. Users have learned from *prior organizational experiences*.

From these insights the IAD framework proceeds by suggesting eight design principles of long-surviving, self-organized social arrangements (the *rules*, in Figure 4.6) addressing collective action dilemmas. The **first principle** (OSTROM, 2000), which in a sense reminds us of the traditional recommendation ensuing from the strict economic theory of externalities, is the presence of *clear boundary rules*. Evidently, the private ownership is possibly the clearest boundary rule. Yet, the significance of the new institutional schools of economics ensues precisely from its ability to reframe such a solution so that it turned out into a special case of a wide range of alternative social arrangements deduced from the core concept of transaction-costs, as well as from the assumption that individuals create institutions aiming to reduce those costs (COASE, 1937; WILLIAMSON, 1979; NORTH, 1991). Therefore, as Ostrom (2000, p. 149) explains, "Group boundaries are frequently marked by well-understood criteria, like everyone who lives in a particular community or has joined a specific local cooperative". Yet, "Membership may also be marked by symbolic boundaries and involve complex rituals and beliefs that help solidify beliefs about the trustworthiness of others". Ostrom et al. (2002) showed that the adoption of boundary rules specifying who can appropriate from the common-pool resource is accomplished in all the organized CPRs examined by them and collaborators. Tang (1991, 1992) explored the variety of boundaries rules (as well as authority) in farmer-owned and government-owned irrigation systems, while Schlager (1989, 1990, 1994) examined this issue in the context of organized inshore fisheries. The researches points to the existence of a great variety of alternatives consisted with the attributes of the resource, environment,

and communities. Ostrom et al. report that a similar diversity exists in regard to forest resources in India.

The **second design principle** is that local rules-in-use, addressing the core problem of how to restrict the appropriation from a common-pool resource, *allocate benefits proportional to required inputs* (as a solution for the second-order, public good problem of providing such inputs), and are crafted to *take local conditions into account*. In this regard Trawick (1999) showed how critical it is the ability to create rules that link user inputs to the benefits they obtain in ways such to configure what the users themselves shall regard as a fair allocation system. Again, there is an enormous variety of possible rules related to appropriation across organized CPRs. Tang (1991, 1992) found that the variety of allocation rules used by farmer-owned irrigation systems is greater than on government-owned systems. Most rules, however, focus principally on assigning the order and timing rights to withdraw water, rather than actual quantity of water to be appropriated. Martin and Yoder (1983a, 1983b), in turn, have illustrated how shared ownership can be more effective than land ownership as the basis for water rights in a comparison of two neighboring irrigation systems in Nepal.

The **third design principle** suggests that the most of *individuals affected by the resource regime should participate in making and modifying their rules*. As an example, Bardhan (1999) found that the quality of maintenance of irrigation canals, in a study of 48 irrigation systems in India, is significantly lower on those systems where farmers perceive the rules to be made by a local elite. Connecting the second and third principles, Bowles (1998) suggests that fair rules of distribution help build trusting relationships, since more individuals are willing to abide by these rules because they participated in their design and also because they meet shared concepts of fairness.

The forth and fifth design principles ensue for the acknowledgment that few long-surviving arrangements rely only on endogenous levels of trust and reciprocity and thus they often requires monitoring and sanctioning mechanisms. The **forth principle** thus connects monitoring with the third principle to indicate

that successful social agreements usually rely on *monitors selected by the community of users*, been either individuals accountable to the user or the users themselves. The **fifth principle** points out that efficient *sanctioning mechanism are graduated* according to the seriousness and context of the offence. Both field and lab evidences supporting these principles are put together in Ostrom et al. (2002).

The complexities involving the autonomy for self-organized solutions within the broader context of the official governmental structures (national, regional and local), lead to the sixth and seventh design principles. The **sixth** one points to the importance of *access to rapid, low cost, local arenas to resolve conflict* among users or between users and officials. In this regard, Ostrom highlights that "Rules, unlike physical constraints, have to be understood to be effective"; and that "There are always situations in which participants can interpret a rule that they have jointly made in different ways". So, "By devising simple, local mechanisms to get conflicts aired immediately and resolutions that are generally known in the community, the number of conflicts that reduce trust can be reduced" (OSTROM, 2000, p. 152). At same time, the **seventh principle** set forth that local users' *rights to organize* and devise their own appropriation rules have to be minimally legitimated by the official governmental agencies. Johnson and Libecap (1982) have illustrated the point for the case of fishery. Finally, the **eighth design principle** relates with the size or scale of the CPR. Accordingly, the structuration of governance activities in *multiple layers of nested enterprises* tends to characterize successful systems when the common pool resources are somewhat larger. As Ostrom (2000) reports, based on Yoder (1992), it is not unusual to find a large, farmer-governed irrigation system, for example, with five layers of organization each with its own distinct set of rules.

Note this principle entails the same idea conveyed in Russell Hardin's (1982) critic on Olson's text. While Olson's perspective minimize the importance of agents' attribution of value but to the object of transaction (a public good, in the case), Hardin analysis begin by acknowledging that people usually derive some rewarding value from the solidarity of the group itself (intrinsic value),

together with a non-additive model for the distribution of the benefits accrued from the provision of the public good. Accordingly, grounded on these ideas, Hardin reframes Olson's typology of groups, distinguishing them by the subgroups that barely benefits when it brings about a collective good. Akin to the ideas of nested enterprises and proportional sharing rules, a subgroup is not motivated by the amount of benefit divided among all members, but by the ratio of benefit to cost for the subgroup. Accordingly, the subgroup will act as long as it barely benefits by bring the collective good.

Using Olson's terms, Hardin defines a *privileged group* as having a subgroup of one member, an *intermediate group* as having a subgroup of a few members, and a *latent group* as having a subgroup of many members (HOWARD, 2002). As suggested, to this n-person prisoner's dilemma-like situation, Hardin adds three mitigating factor associated with a broader view of human psychology, namely: political entrepreneurship, selective incentives, and extrarational behavior. These factors are argued to provide reward values which are usually distinct from the provision of the public good itself and possibly related with unselfish motives.

It is not to say that Olson's account has simply overlooked the complexities of human psychology. Indeed, in his typology, a *privileged group* is defined precisely on the basis that, when the number of individuals involved is sufficiently small, no formal organization is needed because many different psychological, social, and economic incentive can motivate member to provide the public good. Also, the role of institutions is not only the creation of economic but also social and psychological incentives. Finally, Olson explicitly indicates that self-interest is *not* a required assumption for the general provision failure he is concerned with. In fact, he suggests that even "rational altruists" individuals will not act to further group interests if they realize that their effort, as one of many, will have no perceptive effect on the outcome (OLSON, 1965, p. 64). What we have here is a clear appearance of the stance of moral skepticism, upholding both the view of instrumental action as the quintessence of rationality, as well as the situational imperative, here in the form of group size. Such a

widespread perspective is likewise evident in Howard's critique of Olson's approach, as he assumes that what is not explained by instrumental reason should be explained by some non-rational force, such as emotional appeal. In his words:

> Although his study helps explain some of the successes and failings of different organizations, the weaknesses of Olson's approach is like that of other overly rationalist approaches. He acknowledges that organizations can have an ideological or emotional appeal to its members, but he holds that this appeal is still rational [i.e. instrumentally rational] because people make a rational choice to satisfy a personal need. However, ideologies can overpower the rational individual, and while Olson's typology of groups may well explain the workings of established groups like firms, unions and some political parties, it may not help us understand groups that arise out of broad yet vibrant social movements. Furthermore, Olson's model only allows for collective benefits that can be broken down and distributed to members who make an informed rational choice to associate with the group. Since the environment provides indissoluble benefits that can accrue without the conscious knowledge or decision of an individual (i.e. future generations), Olson's model would not explain why we have collective action on environmental issues (HOWARD, 2002).

As we saw, such widespread stance toward rationality is pointed by Habermas as one of the key factors under the long drawn out dispute between economists and sociologists regarding the meaning of rule-governed behavior and the so-called problem of order. On the one hand, Health (2001) observer that the seemingly intractability of this dispute has led many theorists to reject abstract social theories entirely in favor of local, interpretative studies. On the other hand, the same widespread acceptance of instrumental rationality as the only form of rational action helps to explain the gradual encroachment of the "imperialistic" New Institutional schools on territory that has typically been reserved for the Institutionalists. The same attitude toward rationality and supposedly rational intractability of moral issues foster the behaviorist approach to cooperation as a complex system of social feedbacks linking thus, reciprocity and reputation. As we shall see in the sequence, while such core relationship provides a much better account on communication than the rough and ready perspective of standard game theory, it still lacks the depth provided by a direct account of speech acts (Habermas) and the nature of the hermeneutic circle of mutual understanding (Wilber). Both contributions will be important to explain why, despite of the significant and robust positive impact of nonbinding communication in fostering cooperation, these opportunities present uneven

strength to persuade different individuals as to the desirability of committing with a joint strategy on behalf of the collective good.

3.10. The outer face of cooperation: the IAD framework's move toward a behavioral theory linking trust, reciprocity, and reputation

In her "Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation", Ostrom suggests that "Behavior in social dilemmas can be better understood if boundedly rational individuals are assumed to enter situations with an initial probability of using reciprocity based on their own prior training, and experience" (OSTROM, 2003, p. 49). As we saw, Mantzavinos et al. (2003), stress that the issue is not whether agents are perfectly or boundedly rational, but rather how human beings actually reason and choose, individually and in collective settings.

The seeming divergence is just superficial, once we acknowledge that in order to understand how human beings actually reason and choose, we should begin by recognizing that we do those things supported by both limited computational capabilities and information, so that genuine uncertainty (i.e. uncertainty which is uneducable to risk analysis) often hinders actual human beings in distinguishing preferred from less-preferred behavior (HEINER, 1983). This fact, together with individuals' prior training and experience, originates the so-called "behavioral heuristics" (Simon) or rule-governed behavior (Heiner). As already indicated, the significance of Heiner's contribution regarding the meaning of uncertainly in producing predictable behavior is that it puts the decision theory in line with the findings of the developmental psychology. If we put this together with Habermas's onto/phylo parallel, to be discussed in the chapter four, we have all the ingredients of the Integral conception of the development/evolution of the Human Holon (Figure 5.1), which is the hallmark of Graves's ECLET conception.

In this vein, we can put forward that in order to understand how human beings actually reason and choose, as we develop on the basis of our training and

experiences, we should acknowledge that those processes are supported by holarchical structures of increasing flexible mental models and capacities, which mix together – in highly complex and variable patterns – about two dozen of developmental lines. As I have insisted upon, the significance of the stage developmental theory is that it provides the awfully needed empirical information to construct the family of assumption regarding the way individuals acquire, process, represent, retain, and use *information*; how individuals assign value to outcomes and actions; and about the processes (maximizing, satisficing, or using diverse heuristics) they use for selecting particular actions or strategic chains of actions in light of their resources.

However, even lacking the behavioral structuration and restrictions that developmental psychology is ripe to provide, Ostrom (2003, p. 43-44) points out that a theory of bounded rationality and norm-using behavior consistent with the findings of *evolutionary* game theory and psychology assumes that humans tend to:

1. learn from interactions with others about how frequently others use norms such as reciprocity,
2. learn to recognize and remember who are both trustworthy and untrustworthy individuals,
3. cooperate with individuals who are expected to be trustworthy reciprocators in those risky transactions expected to generate benefits (these expectations are assumed to be formed on the basis of prior interactions, from information about social history, or from visual and verbal cues),
4. build a reputation for being trustworthy by trying to resist temptations to gain short-term gains at the expense of losing opportunities for major long-term benefits,
5. punish those who have cheat or simply not reciprocate in the past, and
6. use time frame that extends beyond the immediate present.

Note that, as evolutionary traits, such behavioral trends can certainly be assumed to be shared by all adult human beings, and this is what confers substance to Ostrom's assumption regarding individuals' capabilities to reason, as far as we remain within the instrumental conception of rationality. This is why we are supposed to step forward, toward Habermas's typology (chapter four), in order to appreciate the significance of the stage developmental psychology. Notwithstanding, the trends above provide the foundation for the core relationship liking reciprocity, reputation, and trust, as illustrate in the Figure 3.7.

The figure indicates, moreover, that the triangle of indirect reciprocity is acknowledged not enough to explain the stability of and the large-scale cooperation in humans. Accordingly, the importance of the mechanism of social transmission and other key aspects of human social organization, as well as of the physical environment, that may support cooperation are represented by the exogenous variables that configure the general schema of the IAD framework (compare Figures 3.7 and 3.6). As such, the main point of the behavioral approach is that, instead of trying to explain the level of cooperation directly, the approach dismember the action arena and leads one to link the structural variables to the inner triangle of indirect reciprocity. Therefore, the *situational imperative* articulates the explanation as different levels of cooperation should ensue from different structural scenarios affecting the emergence of indirect reciprocity. Evidently, what I call "situational imperative" is but IAD framework's objectivity, since cooperation is being approached only as an empirical phenomena (Right-Hand), rather than a subjective (internal) disposition or value (Left-Hand), associated with the development and adult human sociocognitive capacities and moral reasoning.

In this vein, the Ostrom puts forward three basic behavioral heuristics imparting reciprocity norms, which research suggests as been likely present in any population of individuals:

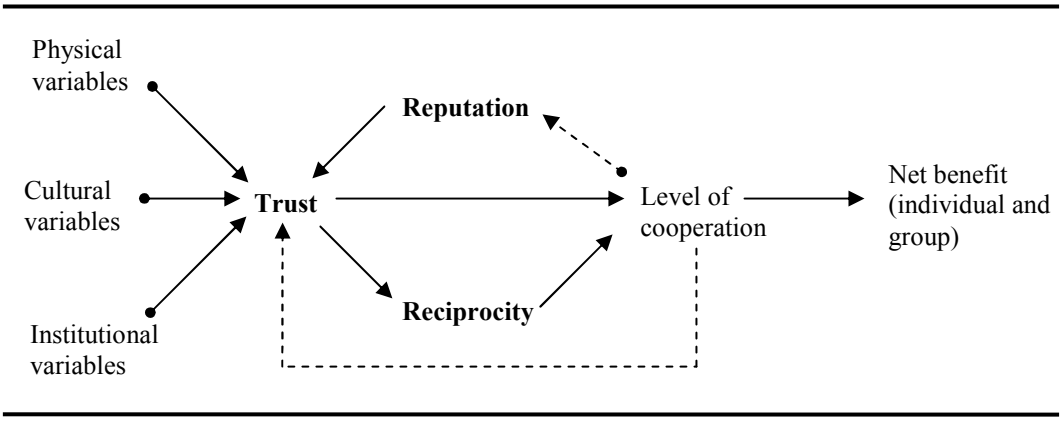
1. Always cooperate first; stop cooperating if others do not reciprocate; and punish noncooperators is feasible.
2. Cooperate immediately only if you judge others to be trustworthy; stop cooperating if others do not reciprocate; punish noncooperators if feasible.
3. Once cooperation has been established by others, cooperate yourself; stop cooperate if others not reciprocate; punish noncooperators is feasible.

In addition to these reciprocity norms, she suggests one can expect to find at least three other norms.

4. Always cooperate in some contexts.
5. Never cooperate.
6. Mimic norms 1 or 2, but stop cooperating if you can successfully free ride on others.

Ostrom (2003) then explains the functioning of the instrumental rationality that links the trust that an individual has on others, the investment others make in trustworthy reputations, and the probability of using reciprocity norms. Accordingly, given the fundamental payoff structure of the dilemma, the more benefit the individuals have received in the past from other reciprocators, the higher their own initial inclination to cooperate. The more often they have faced retribution, the less likely will they be to see free riding as an attractive option. Their trust that others will reciprocate is highly correlated with their own norm but affected by the information they glean about the reputation of other players and their estimate of the risk of extending trust.

Once again, what the behavioral approach lacks is a developmental theory to inform on the probability with which each of the above heuristics may be elicited, not only as instrumentally rational responses to the rewarding feedbacks connected with different structural scenarios, but also in the measure that they ensue from the internal laws of mind and the formation of the character, configuring holarchical subjective structures of increasing flexible rules.



Source: Adapted from Ostrom (2003, p. 51).

Figure 3.7 – The core relationship and feedback in repeated social dilemma.

As we shall see in the next chapter, while the basic three rules of reciprocity are likely to most of Graves's "levels of existence", the rule (6) is typical of the fifth level, whereas the rule "never cooperate" (unless coerced to do so) is the hallmark of the egocentric level of existence (third level). In turn, the contexts which may elicit the rule "always cooperate in some contexts" can be distinguished according to the characteristics of the second, fourth, and sixth levels of existence. In particular, we shall see that the sixth level presents an especial distaste for punishing (but not because of the eventual costs involved in the sanctioning mechanism), while the forth level is partially prone to do so (but not as a genuine expression of "strong reciprocity"). Further, since such levels comprise a holarchy of mental models, each higher level transcend-and-include all the antecessors. This is what produces the "hierarchy of increasing flexible rules" governing behavior, as Heiner (1983) put it. In the limit, at the farther reaches of human potential, Wilber's spectrum of consciousness development suggests that behavior can be hardly inferred from the structural scenario, as the enlightened individual does not act under the pressure of the rewards and gains increasing control over his or her (increasingly free) will. As an illustration, the seventh rule of such numinous sage might be something like:

> 7. Always cooperate on behalf of the collective good, regardless of what others do; never punish even if feasible, for punishment is never costless (it costs *karma*), and it will come anyway; do your duty in this lifetime and pay for your accumulated karma (or sins), so as you can return faster to your true Self (attain the Dao, or Heaven); meantime, witness how people perform their characters in this "divine comedy", right next to the entrance of the Hell (or different sort of humanly devised nightmares); when feasible, advise sentient beings to cultivate and elevate their moral character, for no institutional arrangement will ever suffice to rise above overindulgence, intemperance, self-centeredness, and lack of awareness; nor humans should live under the sword, but move intend toward increasing autonomy, conscientiousness, and divine freedom.

Note that, apart form some arcane concepts I built into it, the above rule governing behavior is possibly the most widespread one, as it appears with variable tones in the so-called perennial philosophy and in virtually all spiritual traditions. The difficulty is that, first, it cannot be by far accommodated into the instrumental conception of rationality – at least not without voiding it of meaning

– and, second, the behavioral approach, the moral skepticism, and radical relativism all of them obfuscates the vision of the vertical position of this seventh rule in the holarchical structure of increasing flexible rules. As clear, in the present context, the flexibility is not expressed in *actual* behavior, which in the appearance is as obstinate as in the rule "never cooperate". Instead, the flexibility ensuing from its higher position in the holarchy of internal rules is a property of the *potential* behaviors, whereas the behavioral firmness is not an expression of rigidity but of the trustworthiness of the character, and *self-control* over the behavioral alternatives that are transcended-and-included along the gradual awakening of the human consciousness.

But putting aside, for a moment, the significance of the stage developmental and Integral psychology (WILBER, 2000b), the behaviorist approach proceeds explaining the levels of cooperation by assuming that, if initial levels of cooperation are somehow moderately high, instrumental rational individuals are expected to learn to trust others, increasing their wiliness to adopt reciprocity norms (as norm 1, 2, 3), or simply cooperate with others. As Ostrom (2003, p. 50) explains, "when more individuals use reciprocity norms, gaining a reputation for being trustworthy is a better investment". This relationship suggests both, that levels of trust, reciprocity and reputations for being trustworthy are positive reinforcing, as well as that a decrease in any one of these can lead to a downward cascade.

Ostrom then points out that, with these core relationships, one can begin to explain why repeated face-to-face communication changes the structure of the situation, as consistently indicated by experimental observations that contradict the prediction of standard game theory. In cooperative game theory, it is assumed that players can communicate and make enforceable agreements. Since agreements are enforceable, standard self-interested rational players are able to change the payoff structure of the situation in order to seize the Pareto improvement resting on the collective action dilemma. In noncooperative game theory, however, players are assumed to be unable to make enforceable agreements (HARSANYI; SELTEN, 1988, p. 3). Therefore, as far as the players'

obey the rules of the standard model of self-interested "perfectly rational" individuals, whether or not they can communicate in considered irrelevant (cf. BARRY; HARDIN, 1982; MILLER, 1992). Ostrom explains that, from this theoretical perspective, face-to-face communication should make no difference in the outcomes achieved in social dilemmas, whether they are one-shot, finitely repeated, or indefinitely repeated (OSTROM, 2003, p. 29).

Such theoretical perspective comprises an extremely unnatural view of the nature of communication. As we shall explore further in the chapter four, the institution of human language through the pragmatic fixation of meaning to utterances could not have emerged out of strategic interactions (HABERMAS, 1984/1987; HEATH, 2001). Therefore, and diametrically opposed to the fundamental assumption of game theory, and Hardin's "double bind" perspective to characterize communication as a strategic interaction and in effect misguiding act, "*the veridicality of speech is more basic than any particular desire that the agents may posses*" and attempt to satisfy through strategic communication (HEATH, 2001, p. 22). Seeing as the obvious and natural *telos* of actual communication (i.e. truthfully inform to get to mutual understanding) has been corrupted on behalf of an abstraction (i.e. to preserve the preeminence of the standard model of individual), no wonder that consistent, strong, and replicable findings indicates that substantial improvement in the level of cooperation are achieved when individuals are allowed to communicate on a face-to-face basis. Ostrom observed that this hold true across all types of social dilemmas studied in laboratory settings and in both one-shot and finitely repeated experiments. In fact, based on Cason and Khan's (1996) contribution, she adds that communication has been found to have robust and positive impact on cooperation levels even when individuals are not provided with feedback on group decisions after every round.

As indicated, a more complete account of the nature of the communicative acts is provided by both Habermas's theory of communicative action and Wilber's discussion on the properties of the hermeneutic circle (Lower-Left). We shall go into this in the chapter four. Nonetheless, out of

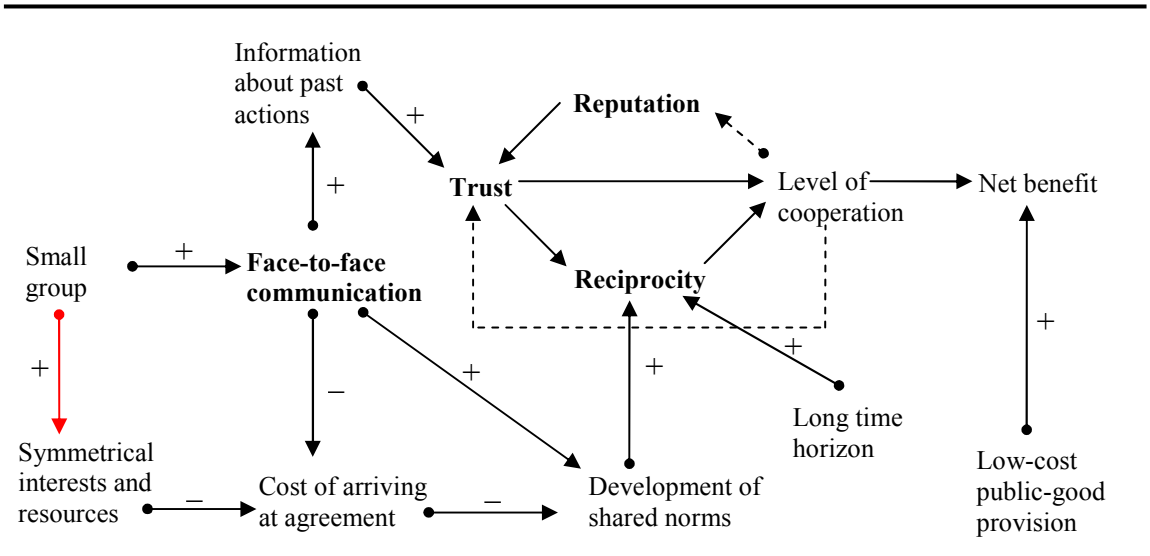
experimental results showing that the importance of communication is not primarily related to the provision of information about optimal strategies (c.f. MOIR, 1995; ISAAC et al., 1994), Ostrom put forward that unless mutual trust in the promises that are typically exchanged during the communicative opportunities in the experimental settings were to increase, expectations of the behaviors of others would not change. "Thus", she concludes, "building trust appears to be a key link in the communication-cooperation connection" (OSTROM, 2003, p. 34). In this vein, the posited link advances that with a chance to see and talk with others repeatedly, an participant into a action situation can assess (somehow) whether he or she trust others sufficiently to try to reach a simple contingent agreement regarding the level of joint effort and its allocation. We shall see that the cognitive science (Upper-Right) helps to explain how people infer the intention of others not only through verbal communication but also from facial expressions and other cues from body language. At the same time, successful communication (i.e. communication that build trust and produce mutual understanding) will be connected also with similar horizons and depth of consciousness, so as that people engaged in communicative acts may have solidarity of signification and concerns (chapter four).

To complete the behavioral perspective on cooperation, the design principles of the IAD framework come to assist the evaluation on how conductive the institutional setting might be in supporting instrumental rational actors to devise successful initiatives to handle their social dilemma. To remind, the likelihood of such initiative has been found to be higher as:

1. Rules on group boundaries are clearly defined.
2. Rules governing the use of collective goods are well matched to local needs and conditions.
3. Most individuals affected by these rules can participate in modifying the rules.
4. The rights of community members to devise their own rules are respected by external authorities.
5. A system for monitoring member's behavior exists; the community members themselves undertake this monitoring.
6. A graduated system of sanctions is used.
7. Community members have access to low-cost conflict resolution mechanisms.

8. For CPRs that are parts of larger systems: appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

Following this general schema (Figure 3.7), Ostrom offers two illustrations portraying first a simple and conductive scenario, and then a complex and difficult scenario vis-à-vis the likelihood of self-organized institutions to support the provision of a public good (e.g. the construction of a bridge). In the conductive scenario (Figure 3.8) the attributes of the community which facilitates the attainment of self-organization are represented by small group size, symmetrical interests and resources, and long time horizon (i.e. the group is assumed to stay together for long). As for the physical attributes, in the conductive scenario the array of variables must contribute to reduce the costs of the provision of the public good (e.g. easy access to supplies and equipments, slight border's length), thus augmenting the net benefit accruable from its provision. These four variables are assumed exogenously, as indicated in the Figure 3.8.



Source: Adapted from Ostrom (2003, p. 57), with the addition explained in the text.

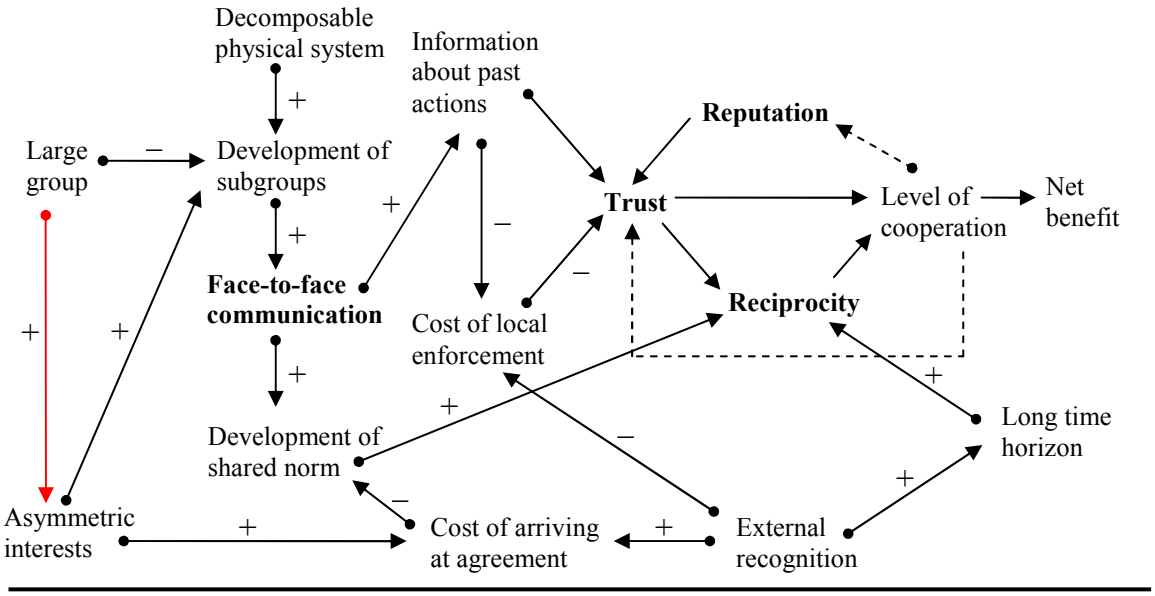
Figure 3.8 – A simple scenario of collective for provision of a public good.

According to the posited relationship, the small group size facilitates (+) face-to-face communication, which is expected to reduce (–) the cost of arriving at agreement, thus facilitating the development of shared norms (the negative sign indicate that the higher the costs the lower the likelihood of the development of shared norms). This trend is reinforced by the symmetry of interests and resources, which is an exogenous variable in the original illustration, but possibly connected with the group size. I thus add by my own an arrow connecting positively this two attributes of the community in order to suggest that the likelihood of symmetric interests and resources is expected to be higher in smaller groups. The illustration assumes that the higher the frequency of face-to-face communication, more information about the past actions circulates and is shared among the participants. As the development of shared norms positively reinforces reciprocity, and then the level of cooperation, the continuous flow of encouraging information helps to built reputation and trust. Reciprocity is increasingly reinforced by long time horizon of interactions, which is presented as an exogenous attribute of the participants, as indicated. We shall see, in this regard, as the time horizon varies with the level of awareness and contributes to compound the distinctive characteristics of Graves's constructs. Finally, low-cost of provision is assumed to have a positive (and indirect) impact in facilitating the level of cooperation by independently raising the net benefit of the provision of the public-good. If sustained, the described virtuous cycle fosters the required cooperation for the provision of the public good, so that the group can enjoy from its facilities.

Ostrom observes that this is a rough but coherent causal model that connects a model of the situation and a model of the individual to offer an explanation for observed levels of cooperation and that allow the identification of the conditions most likely to facilitate collective action. Since the example illustrates a simple and very conductive scenario, changes in any of the exogenous variables (group size, distribution of interests and resources, time horizon, or attributes of the physical resources), tend to increase the difficulty of achieving successful collective actions. Indeed, given the positive reinforcing

structure of the core relationships (indirect reciprocity), disruptive changes in one of the exogenous variable can lead to a cascade of changes in the set of endogenous variables. Therefore, Ostrom (2003, p. 58) stresses out "how difficult it is to make simple bivariate hypotheses about the effect of one variable on the level of cooperation".

In this vein, the importance of group size is illustrated in the difficult scenario (Figure 3.9). To begin with, the regular positive effect of face-to-face communication on cooperation levels is clearly hazarded as the group size increase from a few members up to, say, fifteen thousand individuals. Ostrom's illustration indicates that, is this circumstances, face-to-face communication must be facilitated through the development of subgroups. Such development, however, may or may not be facilitated by the attributes of physical system, such as the topography of an irrigation system or the stationarity/storability of the resource, in helping to decompose the collective action problem in relatively independent pieces to be handled separately by different subgroups. Accordingly, the schema in Figure 3.9 indicates that, if the physical system is decomposable, then face-to-face communication can be fostered through the development of subgroups. This possibility is nevertheless continuously challenged as the number of individuals increase. As before, I added a positive connection between group size and asymmetry of interests, with makes it an endogenous variable. Ostrom suggests that the asymmetry of interests and resources at once makes higher the costs of arriving at agreement and induces the development of subgroups. In each subgroup, face-to-face communication fosters the development of shared norms, as before. However, this effect is now opposed by the increased cost of arising at agreement out of the asymmetry of interest and resources. Moreover, this asymmetry is posited to be reflected in the very development of subgroups, so that this possibly will create difficulties for the development of shared norms at a higher-level of analysis (involving communication between subgroups). In any case, at the level of each subgroup, communication facilitates the flow of information about past actions, as before.



Source: Adapted from Ostrom (2003, p. 59), with the addition explained in the text.

Figure 3.9 – A complex scenario of collective action for provision of a public good.

In a large group like this, the social movement toward self-organization is assumed to require more explicit legitimation from the external authorities. The illustration suggests that this need increases the costs of arriving at agreement, thus opposing the likelihood of local development of shared norms, which is the foundation for reciprocity. At the same time, though, reciprocity may be reinforced by the external recognition in the measure that it is a key factor boosting participants' confidence as to the length of the time horizon for the development of their local rules. As we saw, empirical findings suggest that self-governance in large groups hardly can be sustained without local enforcement. Other things equal, the costs of enforcement are evidently associated with the level of noncooperation, so that the higher these costs the more difficult it is to establish trustworthy relationships. Yet, these costs may be reduced by the influx of information about past actions, as nurtured by face-to-face communication. As before, communication and the resulting flux of

information are connected with the enforcement of trustworthy interactions. However, depending on what is communicated, the formation of reputation either endorses or dampens the core relationship linking trust and reciprocity toward cooperation for the provision of the public good.

3.11. On the partialness of the evolutionary approach to cooperation

In the introduction of this chapter I briefly summarized Heath's objections to the evolutionary game theoretical approach to cooperation. Fundamentally, Heath sustains that, the lack of an independent and intelligible criterion to estimate the functional fit of plural intentional states enforced by a cultural trait or deontic constraint (e.g. concern for others) renders inappropriate to treat "cooperate in prisoners' dilemmas" as a behavioral disposition that could be acted upon directly by the forces of biological selection. As he puts it (Heath, forthcoming):

> Those who have tried to apply evolutionary models to the analysis of cooperation as a cultural pattern have assumed that cooperation will have low fitness in the cultural domain for roughly the same reason that altruism in other species has low fitness in the biological domain. This is what justifies their use of the same replicator dynamics to analyze the conditions for the emergence and reproduction of cooperation among humans. They begin by assuming that the morality meme is unfit, because of the utility payoffs it generates for its host (forgetting that utility for the host is not the same as "enhanced fitness" for the selfish meme). They then go on to look for external structures – "outside" the black box of culture [and consciousness] – that will sustain it (such as group selection). The unstated assumption is that we are able to make judgments, with some degree of confidence, about what will turn out to be fit and unfit in the cultural realm.

However, as we saw, this is just what we hardly could do so; short of a general theory of cultural reproduction that could perhaps provide the problematic criteria to evaluate the cultural fitness of plural intentional states and morality memes. Accordingly, as Heath says, "To set up an evolutionary game with "Cooperate" and "Defect" as the strategies is therefore tacitly to commit oneself to a sociobiological framework that is empirically inappropriate for the analysis of human behavior" (Heath, forthcoming).

I already commented that Heath anticipates the game theorists' objection to such a claim by exposing the strict formal character of the replicator dynamic rule when applied to "memes", and how this procedure ultimately "turns the old 'survival of the fittest' slogan into a tautology". Heath discusses a series of examples around the emergence of language and social learning to clarify that what counts as "fitness" in models of cultural reproduction remains a black box, containing dozens of different factors, all of them interacting in unknown and highly unpredictable ways. According to him, lacking a general theory of cultural reproduction, which must account, moreover, for the nonrandom variation of the cultural traits, we can simply not state with any confidence what is "fit" and "unfit" as a cultural replicator. In conclusion, Health stresses out that, "If we are interested in understanding morality [as we should in order to step forward toward an Integral approach of the collective action dilemmas], the primary question is whether what goes on "inside" that black box is more important that what goes on "outside" of it".

However, it is just here that the "collapse of the Kosmos" (Wilber), i.e. the exclusion of the interior (Left-Hand) side of reality, and the resulting monological-objectivist mode of explanation, level whatever moral distinctions and make it impossible to see, or at least to rationally decide amongst different degrees of moral virtue. As Wilber explains, the collapse of the Kosmos meant a reduction of a rich multidimensional network of values, degrees of consciousness, and qualitative distinctions into their sensory-empirical and thus measurable correlates. Quality and values were not absolutely ignored but they should be translated and measured in objective terms. As a result, this "*flatland* world means a *homogenization of motives* for all beings (…). And this means as well that the homogenized motive would have to be the *lowest common denominator* for all beings: namely, sensory pleasure/pain, hedonic happiness" (WILBER, 1995, p. 435).

Note that such a reduction is the hallmark of the *standard* instrumental conception of rationality, as set forth by Tomas Hobbes ([1651] 1991). Heath (op. cited) explains that, unlike his predecessors, who conceived of practical

philosophy as involving a sometimes bewildering array of virtues, vices, and other such substantive dispositions, Hobbes offers a much simpler framework. Out from a *representationalist* theory of cognition – constituting what Kant would later call "theoretical rationality" – Hobbes put forward that practical deliberation involves simply considering some action and thinking through the representational chain of future consequences in order to evaluate either the *attraction* or *aversion* (or eventually indifference) that one feels toward these imagined consequences ensuing from alternative courses of action. This evaluation is then communicated back through the chain of associations, so that one begins to feel either attracted or averse to the action. Choosing an action will then involve simply selecting the one that is most attractive, or the least aversive. Heath adds that, in order to show that there is nothing more to it, Hobbes goes on to argue that all of the traditional virtues and emotions – the distinct qualities that were normally thought to be associated with actions – can be reduced to some combination of one of these basic attitudes with some belief. Heath then shows how Hobbes goes on to draw out of this framework, which still provides the foundation of mainstream microeconomics, the implications that moral judgments are not going to count as real thoughts – an attitude known as *moral noncognitivism*, which is the fist step toward moral skepticism. Accordingly, as Heath summarizes it, "Moral properties are not generally regarded as having causal efficacy, and so are incapable of interacting with our senses. Thus moral judgments must reflect our passions, not our thoughts".

Note that this moral noncognitivism is quite clear in the way how modern evolutionary game theory approaches cooperation and conformity to social norm, by incorporating emotional reactions such as spite, shame, and guilty into the preference function of rational actors (cf. BOWLES; GINTIS, 2003). Fessler and Haley (2003) explain that different emotions affect the cooperative behavior in different ways. Some emotions lead actors to forego the temptation to defect, some lead them to reciprocate harm suffered or benefit provided, and some lead them to repair damaged relationship. Moreover, the current understanding is that emotions are inseparable and adaptive parts of

human decision-making, not forces which necessarily lead us astray. In fact, Damasio, Bechara and colleagues (DAMASIO, 1994, 2004; BECHARA et al., 1994, 1997, 2000) have showed how emotions may be eminently cognitive, weighting probabilities in so-called "multi-arm bandit" tasks (McELREATH et al., 2003). Another key feature of emotions is that they are sometimes not penetrable by other parts of cognition (ROZIN et al., 1986)[43].

This shows that moral noncognitivism is *not* an impediment for modeling the adherence to social norms or deontic constraints within the framework of instrumental rationality. As we shall see in the chapter four, the limitation is that, on the one hand, the instrumental conception of rationality does not provide a sufficient basis for a general theory of social interactions, especially to address the effect of verbal communication and commitment. Further, the instrumental conception of rationality possibly cannot buttress a consistent theory of interdependent choice that acknowledges both "bounded rationality" and plural mental models of the action situation, for once agents' heterogeneous subjective weighting of the payoff matrix has been recognized, the deduction of game theoretical solutions will demand on the parts of the rational agents' implausible cognitive capacities, so that they could optimally chose among mutually interdependent alternatives of action under the condition of irreducible uncertainty regarding other's expectations and information. Once more, I would like to remark the significance of Heiner's approach, in this regard, as well as its agreement with the stage developmental psychology (HEINER, 1983). However, on the other hand, moral noncognitivism and the said flatland perspective on epistemology easily work in concert to catapult the analyst toward a general moral skeptic attitude. In so far as altruism and concern for others connects with moral issues, we can notice the effect of this flatland view quite clear in Crabtree's (2006) announcement of what he sees as a theory of "universal egoism":

[43] For discussion on how emotional mechanisms provide a new avenue to support cooperation see e.g. Fessler and Haley (2003).

> Biologists, sociologists, philosophers and above all, psychologists, have held to the "universal egoism" theory; that all apparent altruism is really selfishness in disguise. Most arguments for altruism are based on ignorance of the underlying reasons for behaving good towards others or are purely semantic in nature, not logical.
> People behave altruistically for a number of selfish reasons. We are programmed genetically to behave in a way conducive to the sociability of the species: This unconscious species-instinct is the closest thing we have to true selfless altruism. In nearly every other conscious sense, altruism is an illusion. We behave well because social good behaviour fires off pleasant neurochemicals in our brains (the pleasure reward), because consciously or unconsciously we want others to see us as a good person (the social reward) or to feel good about ourselves (for pride and self-esteem). Altruism is image and illusion.

However, I would say, the delusion is not that altruism seems like "selfishness in disguise" when scrutinized under the lens of functional fit, as the researcher's aim is to elucidate what goes on "outside" of the black box of consciousness and culture. The illusion ensues from researcher's epistemological belief that "all the truth that is fit to know" (Wilber), must be supported by sensory-empirical data, exclusively. Once we start to explore the origin of intentional states such as altruism and compassion, just like greediness and self-centeredness, as well as the reasoning under propositional attitudes like concern for others and ideas of morality from the perspectives of the methodological paradigms which are appropriate to inquiry on what goes on "inside" of it, we may realize that there is nothing necessarily ruling out the individual authenticity (phenomenology) or cultural legitimacy (hermeneutics) of the moral act following from the fact that it can fires off pleasant neurochemicals in the brain (Upper-Right), or eventually revert into social benefits (Lower-Right) to the organism who has initiated such action, or both, for the authenticity/legitimacy of the motivation underlining the observed behaviors and physiological correlates is a interior phenomenon – not a fake phenomenon. It cannot be neither reduced to nor deduced from the images that either the neuroscientist or the social engineer sees on the surfaces of the appearances. Likewise, it doesn't matter if the organism is conscious enough to forbear in cultivating truthfulness and compassion, as well as in promoting them, purposefully, which the clear mind

intent of fostering large scale structural and functional fit, so that this planet may last a bit more to nourish our genes.

In order to highlight the issue, let us briefly clarify how the uniqueness of the human large-scale cooperation or hypersociability challenges the biological explanation of altruism among other species. According to Hammerstein's (2003) report, the need for theoretical sophistication in this regard was recognized early on by William Hamilton and Robert Trivers, who based their theories of cooperation on genetic relatedness (kin selection) and on the logic of repeated interactions (reciprocal altruism). Kin selection (HAMILTON, 1964) involves no conscious calculation of costs and benefits. The mechanism requires only that the genes encoding the so-called Hamilton's rule (a rule conveying the cooperative behavior) be favored by natural selection (KURZBAN, 2003). Hamilton's kin selection theory has had many successful applications and is now widely accepted in biology, as well as increasingly in social sciences. It is assumed to have played a major role in the evolution of a vast array of species, and mostly notable the social species, in which contact with conspecifics in general and kin in particular is frequent.

However, in its original formulation, the kin selection theory does not explain how design features that deliver benefits to genetically unrelated individuals or between members of different species (e.g. symbiosis, mutualism) can emerge. For this, Trivers (1971) suggested the mechanism of reciprocal altruism, based on the structure of the repeated prisoner's dilemma. In essence, the reciprocal altruism theory relies on one organism conferring a benefit on another organism *contingent* on the second organism having conferred a benefit on the first. As long as the benefit delivered by each is greater than the cost incurred by each, organism can mutually benefit one another by what is equivalent to the economic principle of gains on trade. By representing the reproductive successes of any given strategy by means of an algorithm which attach value to some function of the payoffs along repeated interactions, those strategies that have greatest reproductive success in comparison with other competing strategies will come to dominate the population. The most famous

simulations of this process is possibly the tournament ran by Axelold and Hamilton (1981), in which the well-known *tit-for-tat* strategy rose above the competitors and proved to be robust in a variety of environments. *Tit-for-tat* is the simplest and strict conditional strategy but it is clearly not the only possible reciprocal approach (see e.g. DUGATKIN, 1997). Moreover, while it showed to be robust in simple situations, the parameter of the environment can have important effects on which strategy are likely to be successful (see e.g. AXELROD, 1984; BOYD; RICHERSON, 1990; NOWAK; SIGMUND, 1992).

Generally speaking, for reciprocal altruism to work, at least four basic minimum requirements must be met – two referring to the structure of the environment (benefits to be conferred and opportunity for repeated interactions; in the Lower-Right) and two to attributes of the organism (sufficient information-processing abilities to distinguish cooperators from defectors and sufficient information-processing sophistication to behave flexibly or contingently; in the Upper-Right). In general, research has shown that human ancestral conditions are likely to have had the required structure for the evolution of reciprocal altruism. Also, the required cognitive capacities seem intuitively trivial to humans, although they involve intricate information processing systems that might pose important barriers to the evolution of cooperation.

Concerning the environment, Kurzban (2003) observers that, while obviously necessary, the possibility of coffering benefits might be relatively rare, given that the commodities valuable by one member of a given species are likely to be similarly valuable to another member, limiting the possibility for gains in trade. Accordingly, Cosmides and Tooby (1992) comment that adaptations for reciprocal altruism might be more likely in organisms that have more variance in their needs over time. Also, diminishing returns apply – in a periodical basis – for most of the basic commodities the organisms need in order to survive (e.g. food and shelter), and this fact favors the possibility of mutual gains. In this regard, it is worth mentioning the implication on this requirement ensuing from Abrahm Maslow's hierarchy of needs (MASLOW, 1943, 1962). The suggestion is simple: the hierarchy of needs is a source of variance in organisms' needs over

time. Organisms who have their basic needs fulfilled are in better conditions to exercise altruism involving commodities associated with the satisfaction of those needs. Additionally, as the satisfaction of those basic needs turn out to be relatively secure, due to the evolution of social structures and means, this social fact underpins the possibility of extending altruism to others, regardless the concern with reciprocity.

Nonetheless, in so far as reciprocity does provide the logic of cooperation, the emergence of mutual aid is less likely as the probability that one will encounter the same partner in the future decreases (MICHOD; SANDERSON, 1985). This is the reason for the significance of Axelrod's "shadow of the future" (AXELROD, 1984). At this point, if we turn the focus for a moment to look at what may be going on inside the black box of culture and consciousness, it should not go unnoticed the implication of the theory of reincarnation on reciprocity. Indeed, in the core of such worldview it is the very issue of karmic relationship and karmic retribution, so that no debt would ever become unpaid. Of course, one is not required to share this view in other to be able to extend the "shadow of the future" beyond one lifetime, as the records on multi-generation inter-family fight, the notorious Israeli-Palestinian conflict, and many others conflicts of this kind make fairly clear. In these cases, moreover, it seems rather marked as well that what goes on inside the black box of conscience and culture is much more significant than what goes on outside of it.

Notwithstanding, and returning to the outer face of cooperation, the inquiry proceeds by examining the neurophysiological and behavioral adaptations which are necessary to support *indirect reciprocity* in nonsimultaneous exchange settings. In particular, cognitive scientists and evolutionary psychologists have been investigating the issues on memory and pattern recognition capacities that support the processes of bookkeeping and reputation formation, as well as the emotional involvement implied in the willing to punish defectors. As Kurzban (2003) explains, in simultaneous exchange one does not have to worry about defection and cheating. In nonsimultaneous exchange, on the contrary, an organism needs to incur into the costs to benefit

another organism in the *hope* that this organism will reciprocate in the feature. In other words, nonsimultaneous exchange requires *trust* (i.e. positive expectation regarding the behavior of the order organism) on the part of the organism that moves first in the transaction, as we saw in Ostrom's core relationship linking reciprocity, trust, and reputation (Figure 3.7). In turn, adaptations to trust, or to move first in a social exchange, require an additional set of adaptations designed to seek retribution for unfulfilled reciprocal exchange. This is, roughly speaking, the hypothesized reason for the emotion of *spite*, as the desire to punish (or at least to avoid) those who defect or fail to cooperate, which works as a critical auxiliary system for trust, according to the canon of the current explanation of cooperation based on reciprocity (see CLUTTON-BROCK; PARKER, 1995; and BOYD; RICHERSON, 1992 for discussion of punishment).

Both the ability to detect cheating and the possibility of building reputation involve the capacity of bookkeeping information regarding outcomes of past interactions, as well as the ability to infer the intention of others. While the "puzzle of friendship" (SILK, 2003) suggest that humans may cooperate without counting, research provide evidence that humans keep rough internal accounts both in regard of goodwill (McCAE; SMITH, 2003) and treats (BOULDING, 1963). As for the ability to read intention, Eckel and Wilson (2003) argue that this capacity has evolved along with mechanism for displaying intention through facial expression, tone of voice, and other bodily languages. Following the leading of Paul Ekman's (1972, 1982) research on the universality of facial expressions, the aforesaid authors have examined the effect of these clues on strategic play in two-person bargaining games. The results indicated that subjects easily (and readily) draw inferences about the trustworthiness of others based on their facial expressions. In particular, smiling and trustworthiness was showed to the perceived as related. However, when it comes to the strategic action contingent on such inferences, behavior and attitudes appeared to be only weak correlated.

Respecting the ability to detect and willing to punish defectors, Cosmides and Tooby (1989, 1992) argue that human evolutionary history has

favored the adaptation of specialized cognitive mechanisms for detecting cheaters. In fact, evolutionary psychologists who study the cognitive structure of the human brain are converging toward the conclusion that humans do not develop general analytical skills that are applied to a variety of specific problems, as implied by the unitary cognitive foundation that characterizes the traditional model of individual in neoclassic economic theory. Rather, the human brain appears to have evolved a domain-specific, human reasoning architecture (CLARK; KARMILOFF-SMITH, 1991). More generally, the authors report that growing evidence indicates that humans have *many* reasoning, learning, and preferences circuits that: (i) are complexly specialized for solving the specific adaptive problems our hominid ancestors regularly encountered; (ii) reliably develop in all normal human beings; (iii) develop without any conscious effort; (iv) develop without any formal instruction; (v) are applied without any awareness of their underlining logic; (vi) are distinct from more general abilities to process information or behave intelligently. Indeed, the authors add, these reasoning, learning, and preference circuits have all the hallmarks of what people usually think of as "instincts" (for discussion see TOOBY; COSMIDES, 1992). The authors comments that "this view implies [furthermore] that cultural differences are vastly overstated, because beneath existing surface variability all humans share the same set of preferences-generating and decision-making devices" (COSMIDES; TOOBY, 1994, p. 330; see also PINKER, 1994). Grounded on such findings, Axelrod (1986) has investigated the emergence and stability of norms on the basis of a double indirect reciprocity mechanism he labeled "metanorms", which specifies the agents' willingness to punish someone who did not *enforce* the said norms.

Please note that, despite all such developments and relative theoretical convergence toward the recognition that humans present plural and specialized reasoning, learning, and preferences circuits supporting adaptive responses to varying circumstances, the still widespread belief respecting the set nature of adult human psychological apparatus – as clearly expressed in Ostrom's methodological discipline, quoted above – ultimately renders such complex

neurophysiological “architecture” into a horizontal (flatland) interlocking system, equally assessable to one and all adult human and that works basically as instrumental reactions to the environmental circumstances. The vertical (developmental/ontological) dimension of such structuration of mind, while acknowledged during the individual history started on birth, or even since the neurological embryogenesis, is simply assumed to come to an end in the adulthood. As we shall introduce in the next section and extend along the next two chapters, while this is *not* an unprovoked assumption, out of both ordinary observation and prevalent moral noncognitivism, it provides a narrow and superficial conception of the human nature and our potential for interior growth and consciousness awakening.

Notwithstanding, as the indirect reciprocity account of altruism is recognized to be insufficient to explain the unlikeness of human hypersociability, the inspection on what goes on outside the black box of human consciousness and culture moves father outwardly toward the inquiry on the social circumstances (Lower-Right quadrant in Figure 3.5) that might explain a gene-cultural co-evolution of human cooperative dispositions. Therefore, theoretical work by Boyd and Richerson (1988) suggests that reciprocal altruism in unlikely to evolve in large and even moderately sized groups. Using multi-agent simulation models, the authors show that reciprocal altruists only do well when they are paired with other reciprocators. In all other cases, nasty strategies do better. Accordingly, cooperation which is conditional on cooperation of all other group member is unlikely to be an effective mechanism for cooperation in large groups. The improbability of emergence and the vulnerability of cooperation in large groups have to do with the quality of information, the extension of the connectedness among the individuals, and costs of committing mistakes, i.e., to trust a cheater (see Figure 3.5). Henrich et al. (2003) reports that, if individuals do not know the “standing” of some fraction of the individuals they may interact with, the basin of attraction of cooperation in the evolutionary simulation models shrinks dramatically. Similarly, if “standing ranking” is even moderately noise, cooperation is unlikely to evolve or be maintained (PANCHANATHAN; BOYD,

2002; BOYD; RICHERSON, 1989; LEIMAR; HAMMERSTEIN, 2001; WU; AXELROD, 1995). As for the payoff structure, Mohtashemi and Mui (2003) show – using a model of in which information is selectively retrieved from and propagated trough dynamically evolving networks of friends and acquaintances – that for indirect reciprocity be evolutionary stable, the differential probability of trusting and helping a reputable individual over a disreputable individual, at a point in time, must exceed the cost-to-benefit ratio of the altruistic act. In other words, in so far as altruism does approximate "selfish in disguise", the benefit received by the trustworthy must out-weight the cost of helping the untrustworthy.

Therefore, as indicated, additional theoretical directions for explaining large-scale cooperation in humans have began to explore the key features of the human sociability, such as economic specialization, social segmentation, communal decision making, and social learning. This is just what is under Ostrom's comparative scenario analysis discussed above. In a reasoning that parallels Russell Hardin's (1982) critique on Olson's (1965) approach to collective action, and is as well incorporated in the IAD frameworks' eighth design principle, Richerson and Boyd (1998, 2000) and Richerson et al. (2003) argue that by subdividing larger groups (e.g. military organizations, lineage groups, and corporations), culturally evolved institutions create small-groups environments that make effective use of our ancient social instincts by embedding them in a large context. Here is the significance of the group selection argument raised by the gene-culture co-evolutionary approach to cooperation, commented by Heath, above. In this vein, Bohem (1996) has argued that, because many small-scale human societies likely engaged in various forms of communal decision making in emergence situations, group members partake in a shared fate that levels the fitness variation within groups and enhances between-group selection processes, thereby generating a species-specific circumstance that favors genetic group selection, provided some conformist transmission is involved. Henrich et al. (2003) stress out that the understanding the details of human social learning may be critically important for studying the

evolution of cooperation. In particular, the authors point to the cognitive adaptations for acquiring adaptive behaviors in information-poor environments (BLOOM, 2000; SIMON, 1990; RICHERSON; BOYD, 1998; BOWELS; GINTIS, 2003). In this regard, they suggest that one of the key areas for future research is to figure out what is actually transmitted between individuals and what is inferred from limited information.

Young (1998), Boyd and Richerson (1985), and Calvalli-Sforza and Feldman (1981) have begun to develop theories for how social norms and rules are transmitted within and between social groups by learning and imitation. Both theory and evidence suggest that, instead of merely imitating a random individual from population or learning from one's parents, individuals uses both model-based cues and integrative algorithms to extract adaptive information from their social world. The issue relaters with the selective focus of attention (a scarce resource) so that learning usually focus on individuals most likely to have the information that is worth acquiring. Cues may be numerous and vary according the domain of behavior. The authors single out the conformist character embedded in basic rules such as "copy the majority" or "ignore outliers" preferentially, which has receiving substantial support from empirical work in psychology, economics, and anthropology (summarized in HENRICH; GIL-WHITE, 2001). Signaling provides a related pathway to explain cooperation based in the provision of clues supporting selective social interaction. The signaling approach suggests that individuals may be willing to provide collective goods in the absence of reciprocity or punishment when the provision of such goods can act as an *honest* signal to other individuals of the "provisioner's high quality" (HENRICH et al., 2003). The reasoning is that, as a consequence of receiving this advertisement, other individuals are more likely to provide fitness-enchasing benefits to the signaler in the form of coalitional alliances and matting opportunities. The theory links the honesty of the signals with the costs of signaling, so that quality-dependent cost serves to ensure that the signal honestly advertises the relevant underlining qualities of the signaler. Finally, because a

"good signal" has high broadcast value (lots of people receive the signal), the provision of collective goods is posited to provide one effective type of signal.

In summary, the human distinct form of sociability creates a variety of transmission mechanism other than genetic. In particular, social learning mechanism creates a second system of inheritance that allows the transmission of many cooperative phenotypes that cannot evolve genetically in a given population, because this would require group selection, while selection is often too weak at the group level (RICHERSON et al., 2003). Nothing in such overall interpretation is being denied here. The problem, as we saw along with Heath's remarks, is that the just described awful complexity and subtleness of human sociocultural life renders basically ineffective the methodological resolution in explaining the emergence of human sociability and moral concerns only under the perspective of functional fit. Most important, as the naturalistic framework distorts the perspective on the genuineness of the moral act – as in saw in Crabtree's (2006) defense of the "universal egoism theory", and Hardin's (1993) announcement on the long-run demise of "conscience", together with its short-run "pathogenic effects" as well – and turns indeed difficult to see how a defense of morality and genuine concern for other could be assembled out of these elements, Heath (forthcoming) points out that it is in fact "easy to imagine morality being dismissed, along with mythology and religion, as merely an immature stage in the development of human understanding".

Clearly, in so far as mythology, religion, as well as ideas of morality entail and typify different systems of representation, which are continually legitimated within the hermeneutic circles formed by the new individuals passing through such stages of development of human understanding, they shall never be dismissed from the sphere of human representations, even if the contents of the myths, the objects of the religious devotion, as well as the ideas of morality themselves may change beyond recognition. Indeed, as far as I am concerned, these are all merely different ways of describing the same phenomenon of consciousness awakening, each one proper and typical of each stage of development of understanding, including the current instrumentally rational,

system-objectivistic approach to altruism and cooperation. Some important remarks in this regard are, first, that each developmental stage may evolve either in a sound or pathological fashion, as we saw in Wilber's characterization of the Ego and the Eco camps. Second, as culture and society evolves and submit their own past to new forms of representation and practices, the mythological or religious social order, for instance, shall no longer provide the collective and average mode of understanding, but only be recapitulated, for the better or worse, into the ontological development of new comers, as in the modern mythological figure of Santa Claus and the reconfiguration of the religious spirit of the celebration of Jesus's birth according to logic of material consumption and economic growth. The issues on sociocultural evolution will be further discussed in the chapter four. Respecting the epistemology of morality, Heath (forthcoming) points out that:

> [The] mere fact that normative conclusions cannot be derived from factual premises does not mean that a theoretical reconstruction of the normative structure of morality need not cohere with an enormous amount of other scientific knowledge, in fields ranging from evolutionary biology and cultural anthropology to developmental psychology and sociological theory. After all, doing normative theory is not the same as legislating the way things ought to be. The moral philosopher is offering a reconstructive articulation of the normative ideas that are already implicit in everyday social interactions. Whether or not morality is a product of socialization is an empirical question. If it is a product of socialization, then we need an empirical account of how it gets reproduced across successive generations, how it arose, and why it might change over time. These are all empirical questions, yet also questions whose answers impose constraints on the range of plausible metaethical theories.

Heath notices, moreover, that these empirical accounts are naturally not generated in the inductive style that early positivists proposed as the model of valid scientific reasoning. Indeed, he says, scientific researchers interested in morality already rely heavily upon normative theories as a guide to hypothesis-formation in their work. Heaths illustrates the point by reminding us on how Kant's moral philosophy, for instance, "has explicitly informed an important tradition of empirical research, which starts with Emile Durkheim and runs through Jean Piaget to Lawrence Kohlberg". Finally, he adds that the results of this research have in turn affected the direction of thinking in Kantian moral philosophy, and that this reasoning provides the basis for his claim that the

normative theories must only "cohere" with the results in various empirical disciplines.

In short, Heath seems just to be proposing a balance respecting the significance of what goes on outside and inside the "black box" of consciousness and culture, as well as with regard to the methodological attitude for the inquiry into those domains of the Human Holon. Nonetheless, out from Habermas's transcendental pragmatics of discourse ethics, Heath does *not* endorse the traditional interpretation of morality as a *system of virtues*. According to him, if morality really were a system of virtues, then it should be possible to detect systematic differences in the behavior of individuals who do and do not possess such virtues, and these virtues should be a major determinant of behavior. Yet, he says, empirical studies have shown, time and again, that there are, in general, no such stable traits of character, and that even in cases where there are, they are easily overwhelmed by situational factors.

This is a terrible subtle issue, since it is plenty of potential for semantic confusion. On the one hand, the term "transcendental" has acquired somewhat pejorative connotation in contemporary philosophy, as Heath (2001, p. 283) explains, "in part owing to an unfortunate tendency to ignore the technical meaning of the term, and to use it in a way that makes it roughly equivalent to "metaphysical" (a tendency that is exacerbated by the way that the word is used in popular culture, in expressions like "transcendental meditation", and so forth). On the other hand, the term "metaphysics" itself is likewise commonly downgraded, as the scientific revolution turned natural philosophy, which used to be a part of metaphysics, into an empirical and experimental activity and unlike the rest of philosophy. Metaphysics therefore became the philosophical inquiry into subjects "beyond the physical world". Following the Kantian critique, and in part due to the obscurity of most philosophical *tractatus*, metaphysics is seen as mere speculative and often useless thought or rather to include a wide range of controversial phenomena believed by many people to exist beyond the physical. In this last sense, a "metaphysical bookstore", is not one that sells books on ontology, for instance, but rather one that sells books on spirits, faith healing,

crystal power, occultism, and other such topics. According, as Heath (forthcoming) explains, a "transcendental" claim, in Kant's sense of the term, denotes the opposite of a "metaphysical" claim (especially in the popular sense of the term), so that the idea of transcendental necessity (the Kantian *a priori* point of view) arises directly from the recognition of the *embodied* character of human rationality and cognition. Therefore, as Heath (2001, p. 285) puts it, "It is precisely because our knowledge always arise in specific biopsychosocial contexts that aspects of this context acquire a transcendental status with the respect to the world as the object of this knowledge". In this vein, the need for a transcendental argumentation strategy, in Heath's view, is essentially imposed by the willingness to understand moral motivation in its empirical context, taking into account the important facts known to us from the study of human psychology, child development, sociology, and evolutionary biology.

Note that such transcendental pragmatics, as simply the *conditio sine qua non* of knowledge and morality, is no more than the enactment principle of Wilber's Integral Methodological Pluralism. Therefore, while what Heath means by "system of virtues" is not that clear, it seems he fails to notice the full implication of the stage developmental point of view, as an expression of the enfoldment principle, which provides, in my view, a better response to radical relativism than the mere claim of the transcendental necessity *per se*. Accordingly, the stage developmental psychology puts forward that the emergence and structuration of plural intentional states, mental models, and worldviews, inside, in connection with those various reasoning, learning, and preferences circuits in both the human brain and the social practices, outside, obey their own internal laws of transformation, and are not an universal function of the situational factors. As all these such *biopsychosocial* systems (the very term Graves uses) comprise holons, and holons evolve holarchically in direction to increasing autonomy and *telos*, no wonder that the agents' moral reasoning and decisions are (potentially) highly responsive to the circumstances of action, which mind may assess increasing critically. Furthermore, that people often "fail" to react as expected to the contingencies of action is the central focus of the

behavioral experimental psychology and discrimination theory of rule-governed behavior. Indeed, contrary to Heath's claim, rule-governed behavior has been currently shown to be systematically associated also with personality traits, not only with the away of discriminative stimuli arranged in different reinforcing schemas. In addition, the overall psychology of humans entails about two dozen relatively independent developmental lines or streams, among which the moral line is but one of them. In order to morality, as a clear distinctive trait of character, be likely to dominate the overall psychology and be a major determinant of behavior it would be required that the individual was either stuck at the conventional-absolutist stage of moral development or have already get to father reaches of human potential, so as to rise above (transcending-and-including) the heteronomy of the circumstances and social approval, and take hold of the autonomy and freedom pointed out by moral philosophies as distinct as those of Kant and Nietzsche, or Fiche and Spinoza.

Therefore, despite the difficulties inherent to the complex, subtle, flowing, and waving nature of the interior development and how it affects the quality of the behavioral responsiveness to specific situational variables, we shall see in the chapter six that the results of our empirical tests did detect systematic differences in the participants' behavior, in diverse experimental conditions and institutional settings, associated with Clare Graves's stages of *biopsychosocial* (i.e. embodied) development. While I do *not* claim that these evolving intentional states are the major determinants of behavior, and am ready to acknowledge that they *may* indeed be easily overwhelmed by *certain* situational factors (e.g. a burning cinema, unbearable humiliation), the significance of such systematic differences ensues not from they presumed supremacy to determine behavior. Instead, as pointed out, the significance of a serious consideration of the internal rules of interior transformation is that it opens a new avenue to deal with social dilemmas and other issues calling for a deeper account of deontic constraints. Accordingly, grounded on a clearer understanding of the laws of mind and the formation of the character, in which John Stuart Mills struggled to found his utilitarianism, we can handle some key situational factors in order to foster the

natural awakening of human consciousness, eventually toward the awareness of the profound meaning of morality beyond all conventions and social practices. Morality as "systems of virtues", in this sense, comes up as a *post hoc* normative gift, ensuing from similarly embodied experiential injunctions and social practices. In order to formulate such standpoint more cogently, I must expand a bit on the difficulties of a direct (non sensory-empirical) apprehension of reality.

3.12. On the difficulties of the immediate apprehension of reality, the elusiveness of its communication, and the validity claim of mysticism

As commented in the first chapter, Kant's critique of metaphysics brought forward an epistemological strangeness, following the previous cosmological and ontological strangeness resultants, respectively, from the Copernican revolution and the Cartesian dualism, in the course of the formation of the Western worldviews (TARNAS, 1999). We just saw, along with Heath's justification of the transcendental argument, that Kant's perspective ensues from his re-embodiment, say, of the human mind and reason, so as that human knowledge is now seen as inexorably reliant upon our sensorial machinery.

Granting the understanding that, in the Kantian sense, the transcendental inquiry denotes noting more than an attempt to investigate the *conditions of possibility* of some particular (and typically mysterious) phenomenon, such as human factual knowledge, Kant ascertained that the only way to avoid the conclusion that certain facts are transcendentally necessary is to posit a "God's-eye point of view", or some form of intelligence that is able to grasp things as they are in themselves (HEATH, 2001, p. 285). Evidently, Kant's critique presumes human mind is devoid of such a form of intelligence or apprehension. Therefore, accordingly to the Kantian "philosophy of consciousness", as Habermas calls it, there would be no transcendentally necessary truths, such as the pure reason's "rules of understanding", only if the world were not in any way constituted by our own cognitive faculties. Heath (op. cited) then explains that, with the so-called linguistic turn, transcendental inquiry become less interested in

structures of consciousness, characterized in terms of the dominant perceptual paradigm, and becomes more interested in the structure of language. Seen as philosophers had for a long time noted that both consciousness and language share the property of intentionality, or "aboutness", the linguistic turn reversed the early modern viewpoint, which used to assume that the intentional structure of consciousness was to be taken as primitive and the structure of language as derived. As a result, Heath explains, since all of the "representations" that epistemologists have traditionally been interested in are *propositional attitudes*, the linguistic medium in which these attitudes are formed will occupy a transcendental status with respect to the content of any representation.

Heath (2001, p. 287) clarifies, further, that one way to understand the shift in transcendental philosophy effected by the linguistic turn is to see it as a disagreement over which possible worlds are cognitively accessible from our own. As indicated, the broadly verificationist constraint, in the heart of Kant's critique of metaphysics, ensues directly from the transcendental character of his postulation that all cognition had to arise through sensibility, and so the only worlds cognitively accessible from our own were those that satisfied the conditions of possible sensorial experience, i.e., experiences mediated by ordinary sight, hearing, tact, etc., and their technological expansions. Following Wittgenstein's unconvincing attempt, in Heath's view, to reformulate this Kantian idea by eliminating the distinction between sensibility and understanding, so that language itself would constitute the only transcendentally necessary structure (cf. WITTGENSTEIN, 1961), a renewed interest in Kantian verificationism appears in the work of theorists like Dummett and Ayer, for instance, now within the linguistic paradigm. According to Heath's (*op. cit.*) account, both Dummett and Ayer retain from Wittgenstein the idea that language determines the cognitive accessibility relation, but suggests instead that conditions of possible experience impose a verificationist constraint on the expressive power of language. "They agree, in effect, that the set of cognitively accessible possible worlds is limited only to 'that which can be said,' but they argue that 'that which can be said' is in some ways limited by 'that which can be

experienced'". Heath then explain that Habermas takes this shift to the philosophy of language a step further, holding that language occupies a transcendental position with respect to the world it is used to represent, but that language itself in grounded in certain types of social practices. Accordingly, if a particular set of social practices, such as argumentation, confer semantic content upon our utterances, and if language is a vehicle of thought, then these practices are just as constitutive of our cognitive capacities as the symbolic systems that they support. "This means that not only will certain features of language be transcendentally necessary but certain social practices will as well" (HEATH, 2001, p. 288). This is, in short, Habermas's "transcendental pragmatic claim", in which Heath grounds his refutation of morality as a "system of virtues"[44].

As indicated, I am *not* committed to the claim that the normative structure of morality involves any metaphysical basis, seen as the transcendental pragmatics in which actual moral systems are grounded in and evolve just amounts to the tetra-enactment principle of the Integral Theory. The tricky point is that the Integral Theory suggests that the Kantian epistemological stance is limited by an insufficient account of human actual cognitive and perceptual capabilities. Note that this is not an original story, as Henri Bergson, for instance, grounds his *Creative Evolution* ([1911] 1998) in an explicit refutation not of Kant's verificationism on "that which can be cognitively accessible", but of his sensorial restraints and *a priori* status of his "laws of understanding". Indeed, Bergson put forward that the epistemological strangeness of Kant ensues from such "haven fallen" character of his posited laws of theoretical reason. Actually, by regarding the general frames of the understanding as absolute, that is, as no longer dependent on anything, Kant have made everything else dependent on them. Bergson, on the contrary, regards that the nature of human *intellect* or intelligence is explained by the *needs of action*. Precisely because of this, the intellect would be specially adapted to deal with the inert, seeing things *as if* they

[44] We shall see in the chapter four that this is as well the basis in which Habermas point to the insufficiency of the instrumental conception of rationality. Heath (2001) then provides an additional support to Habermas's point of view based on a negative proof ensuing from straightforward game theoretical reasoning.

rest immobile "out there", and exerting over it an unlimited power of decomposing its object according to any law and of recomposing it into any system (BERGSON, 1998).

Note that Bergson's formulation helps to understand what Quine (1960) calls our "tiresome bias" in the treatment of time and the corresponding object ontology which is reproduced by the subject-predicate structure of language, not dictated by it, as suggested along the linguistic turn[45]. This would mean that a philosophy of consciousness is primary to a philosophy of language. In any case, according to Integral Theory, the transcendental argument is located neither in the individual (consciousness/brain functioning) nor in the collective (language, culture/society). Rather, the conditions of possibility of every *holon of being and knowing* should be searched for in the tetra-enactive nature of evolution in the four quadrants. Therefore, while Kant denied any other faculty of knowing but the intellectual (reason), Bergson recognized *intuition* as the faculty which allows us to apprehend reality immediately, or reality seen *from within*, as an unbroken, single, and creative fruition which is *time* or *duration itself* (BERGSON, 1998, 2002). Yet, the potential for experiencing the intellectually unmediated reality – as duration as such – using the power of intuition, is restricted by the push of intellectual mind, turned upon the ordinary need of acting on the environment in order to survive. Accordingly, the intellect does indeed provide our *usual* way of being in relation with the world, while not all which can be experienced and is cognitively accessible. As Bergson explains:

> Everyone can have noticed that it is more difficult to make progress in the knowledge of oneself than in knowledge of the external world. Outside oneself, the effort to learn is natural; one makes it with increasingly facility; one applies rules. Within, attention must remain tense and progress become more and more painful; it is as though one were going against the natural bent. Is there not something surprising in this? We are internal to ourselves, and our personality is what we should know best.

[45] As Heath (2001) explains, Quine's suggestion that one could adopt an "event" or "process" ontology instead of our usual "object" ontology was rejected by some philosopher along the linguistic turn owing to the fact that the object ontology is dictated by the subject-predicate structure of language, and that a language lacking this structure would be defective in various ways (c.f. BRANDOM, 1994, p. 376-384). My point in the text is only that the object ontology in actually dictated by the nature of the intellect, as put forward by Bergson. As such object ontology constitutes our ordinary way of perceiving the world using power of intellect; it is communicated to the structure of language.

> Yet, such is not the case; our mind is as if it were in a strange land, whereas matter is familiar to it and in it the mind is at home. But that is because a certain ignorance of self is perhaps useful to a being which must exteriorize itself in order to act; it answers a necessity of life. […] Thus nature turns mind away from mind, turns mind toward matter. But in that way we see how we can, if we like, indefinitely widen, deepen, and intensify the vision of the mind which has been granted to us. Since the insufficiency of this vision is due in the first place to the fact that it is directed upon the mind already "spatialized" and divided in mental compartments where matter can be inserted [in memory], let us separate the mind from the space in which it is so at home, from materiality which it takes to itself in order to rest upon matter. In so doing we shall restore it to itself and be able to comprehend it immediately. This direct vision of the mind by the mind is the chief function of intuition, as I understand it (BERGSON, 2002, p. 41-42).

Bergson's account on the nature of intellect and intuition, as well as on the qualities of a possible introspective apprehension of duration or time in itself, has been currently refined by David Bohm's cosmological distinction between an *implicate* and an *explicate* order (BOHM, 1980). Accordingly, Bohm's cosmology put forward that, in the enfolded [or implicate] order, space and time are no longer the dominant factors determining the relationships of dependence or independence of different elements. Rather, in a delineation quite close to Bergson's narrative, and in a sense also answering Quine's beg for going beyond the "tiresome bias" in the treatment of time (i.e. our ordinary object ontology), Bohm claims that an entirely different sort of basic connection of elements is possible, from which our ordinary notions of space and time, along with those of separately existent material particles, are abstracted as forms derived from the deeper order. These ordinary notions in fact appear in what he calls the "explicate" or "unfolded" order, which is a special and distinguished form contained within the general totality of all the implicate orders (BOHM, 1980, p. xv). Then, concerning the philosophy of consciousness, and how in ordinary consciousness [i.e. intellect, for Bergson, or gross cognition, in Wilber's saying] the explicate order is what is manifest, Bohm (1980, p. 205) goes on to say:

> The manifest content of consciousness is based essentially on memory, which is what allows such content to be held in a fairly constant form. Of course, to make possible such constancy it is also necessary that this content be organized, not only through relatively fixed association but also with the aid of the rules of logic, and of our basic categories of space, time causality, universality, etc. ... there will be a strong background of recurrent stable, and separable features, against which the transitory and changing aspects of the unbroken flow of

> experience will be seen as fleeting impressions that tend to be arranged and ordered mainly in terms of the vast totality of the relatively static and fragmented content of [memories].

In the chapter four, after I have introduced Graves's theory, we shall see how Wilber connects the cognitive possibility of grasping such unbroken flow of experience, by factoring cognition according to the three natural *states* of consciousness, namely, of being awake (gross or intellectual cognition), dreaming (subtle cognition), and in deep sleep (causal or intuitive cognition). Moreover, the possibility of expanding consciousness in order to cognitively apprehend the implicate order, and communicatively create an hermeneutic circle within which knowledge is constructed and legitimated, rigorously follows the three fundamental strands on any valid knowledge, as proposed by Thomas Kuhn, namely, a series of practical injunctions or exemplars; apprehension, illumination, or experience of accessible worlds from our own; and communal checking, leading to either rejection or confirmation of the evidences thus produced (WILBER, 2001, p. 75).

Such experiences pertain evidently to the domain of spiritually, and bring about explicit moral precepts which look like that which Heath possibly means by a "system of virtues". Here is the tie I took from the Integral Theory to bound the already manifest connection between collective action dilemmas and morality, with the findings of the stage developmental psychology and the *contemplative sciences*. As we shall explore in more detail along the two following chapters, the reason why the spiritual response to moral noncognitivism ("why be moral?") and to the general skeptic doubt about the status of human knowledge is associated with the *higher levels* of the development of consciousness, and are *introspective* instead of outwardly focused, can be summarized, in the terms of Bergson's distinction between the nature of intellect and intuition, thus: (i) the spiritual awakening relies upon the capacity and strength of intuitive mind instead of those of intellectual mind, (ii) the intuitive capacity of the mind is the proper faculty for inwardly contemplation that illuminates immediate data in consciousness, while the

intellectual faculty is adapted to act upon the world which is perceived through the fleeting impressions upon our senses, and then represented or stored in fragmented content of memories, following the ordinary object ontology, and (iii) while both faculties (or the three kinds of cognition, in Wilber's construction) comprise our innate perceptual equipment, the powers of intellect (or gross cognition) develop first as the ordinary form of awareness because it is required to secure biological survival (it answers a need of life, as Bergson put it).

Accordingly, the radical validity claim of Wilber's *postmetaphysical mysticism*, which is deep spirituality, ascertains that it involves in part a *broad science of the higher levels of human development*. In other words: deep spirituality involves the *direct investigation of the experiential evidence disclosed in the higher stages of consciousness development* (WILBER, 1998). Once again, these deep-spiritual investigations follow the three strands of all good science: they rely on specific social practice or injunctions (such as meditative contemplation); they rest their claims on data and experiential evidence; and they constantly refine and check these data in a community of adequate – which is why they are referred to as *contemplative sciences*.

In Bergson's speech, and watching against terminological confusion, such immediate experience that illuminates the inner side of reality is the proper method of *true metaphysics*. For him, "we should call it science, did not science prefer to limit itself the other [the exterior] part of reality". Indeed, according to Bergson's epistemology:

> Metaphysics, then, is not the superior of positive science; it does not come, after science, to consider the same object in order to obtain a higher knowledge of it. To suppose such a connection between then, as is the almost invariable custom among philosophers, is to wrong both of them: science, which one condemns to relativity; metaphysics, which will never be anything more than a hypothetical and vague knowledge, since science will necessarily have taken to itself in advance everything precise and certain that can be known of its objects. Quite different is the relation I establish between metaphysics and science. It is my belief that they are, or that at least they can become, equally precise and certain. They both bear upon reality itself. But each one of them retains only half of it so that one might see in them, if one whished, two subdivisions of science or two departments of metaphysics, if they did not mark divergent directions of the activity of thought.

> Precisely because they are both on the same level, they have points in common and each can, upon these points, be verified by the other. To establish between metaphysics and science a difference in dignity [or validity], to assigning to them the same object, that is to say, the totality of things, stipulating that the one shall look at from below and the other from above, is to exclude this mutual aid an reciprocal verification: in that case, metaphysics is, of necessity – unless it loses all contact with the real – a condensed extract or hypothetical extension of science. Instead of this, let us allot to them different objects; to science let us leave matter, and to metaphysics, mind: as mind and matter touch one another, metaphysics and science, all along they common surface will be able to test one another until contact becomes fecundation (BERGSON, 2002, p. 43-44).

In other words, Bergson is just suggestion something like the Integral Methodological Pluralism, in connection with the acknowledgement of the interior and exterior domains of reality that configures the vital postmetaphysical step in Integral Theory. To be sure, Bergson ascertains, further, that the results obtained on either side will of necessity be linked, because matter links up with mind. "If the insertion is not perfect, it will be because here in something to rectify in our science, or in our metaphysics, or in both". Bergson explains that true metaphysics (or post mere intellectual metaphysical speculation), as any young science, is always quick to dogmatize, out of having only a limited experience at its disposal, which forces it to work less upon facts than upon simple ideas, suggested by facts or not, that it then treats deductively. Given that the faculty of intuition, while existing in each one of us, is nonetheless covered over by functions more useful to life, Bergson says, a whole labor of clearing away is necessary in order to open up the way to inner experience. That such labor in part involves the development of consciousness, comprising a number of major interior transformations toward higher levels of awareness, as put forward by the central thesis of this study, is quite clear in Bergson's account that "the higher the consciousness, the stronger is [the] tension [that intuition must bear in order to sustain] its own duration in relation to that of things" (p. 89), which is necessary to disclose the nature of an entire half of the real.

However, as indicated, Bergson is at that point involved with the issue on the development of a more precise language to foster the progress of valid knowledge regarding both the phenomena being described at the fringe of contact between the brain (by conventional science) and the mind (by true metaphysics),

and the communal checking on the subject of the depths of the worlds cognitively accessible trough immediate inquiry. If only because the extremities of the events apprehended through the strength of intuition "will have to fit exactly upon those of positive science", Bergson (2002, p. 44) says, "our metaphysics will be that of the world in which we live, and not of all possible worlds. It will embrace realities". Accordingly, "science and metaphysics will differ in objects and method, but will communicate in experience. Both of them will have to put away vague knowledge stored up in the usual concepts and transmitted by means of worlds", such as the "low" and the "high", the "heavy" and the "light", the "dry" and the "moist", once used in explaining the phenomena of nature. In this regard, Bergson suggests, the true metaphysics is in a much less convenient position compared with the mathematical form with modern physics has taken. For while it also must begin by eliminating ready-made concepts, as it also rely upon experience, that inner experience which is constitutive of true metaphysics will nowhere, in Bergson's view, find a strictly appropriated language.

Bergson explains that this would be so because, in order to be communicated, the inner experience will be necessarily compelled to return to the concept, with at most the addition of image, but then "it will have to enlarge the concept, make it more flexible, and indicate, by the colored shading around the edges, that it does not contains the whole of experience" (BERGSON, 2002, p. 45). It seems worth mentioning, at this point, that what kind of language may emerge to communicate metaphysical experiences and knowledge will depend not only on the depth and stability with which such practices are experienced (the strength of intuitive mind to sustain it own duration and illuminate more and better data), but on the depth or level of consciousness development of those to whom one intends to communicate with. The communication can thus take different forms depending on the public it must address. It can take the form of seemingly simpler anthropomorphic tales and mythologies, or the obscure langue of alchemy, the form of poetry, metaphors, parables, and paradoxes, as seen in the Bhagavad Gita, the Dao De Jing or the Christian Evangel, for instance. It can

as well take the form of body language, as the hand signs performed in Buddhism and Daoism. In any case, those who have no cognitive access to that which can be experienced through contemplative cannot grasp what true meaning of that whit is being communicated, that is, the meaning that the communication has within the hermeneutic circle of the initiated in those social practices. In any case, as true (i.e. experiential) metaphysics moves toward conceptual precision, it will have accomplished in its domain, Bergson claims, the same kind of reform that modern physics has brought about in its own.

The critical point to bear in mind is that one should *not* expect of such metaphysics simple conclusions or radical solutions. As Bergson explains, this would be lead us back to the region of the pure possible and would be no more than manipulation of concepts. At this point, Bergson denounces, as the initial vice of philosophical systems and the source of the dogmatism of modern philosophy, the trend to increase the extensions of a term or definition to the point where it designates everything that exists. As he puts it, "no matter what name you give to the "thing itself", whether you make of it the Substance of Spinoza, the Ego of Fiche, the Absolute of Schelling, the Idea of Hegel, or the Will of Schopenhauer, it will be useless for the word to present itself with its well-defined signification: it will lose it; it will be empty of meaning moment it is applied to the totality of things". Such a manipulation, Bergson concedes, may have the beauty of the definite, but it is suspended in the air, in the region of pure possible. In other words, it is subordinated to the thesis; it is hypothetical. Of course, the same would happen to Bergson's concept of *duration* was not it made explicit it represents his attempt to put into words the content of a metaphysical experience. Here is what he says about the method and the development of his contributions to experimental metaphysics:

> Tension, concentration, these are the words by which I characterize a method with required of the mind, for each new problem, a completely new effort. I should never have been able to extract from my book Matter and Memory, which preceded Creative Evolution, a true doctrine of evolution (it would have been one in only appearance); nor could I have extracted from my Essay on the immediate Data of Consciousness a theory of the relation of the soul and the body like the one I set forth latter in Matter and Memory (I should have had only a hypothetical construction); nor from the pseudo-philosophy to which I was devoted before the Immediate Data – the is to say from the general notions

> stored up in language could I have extracted the conclusions on duration and the inner life which I presented in this first work. My initiation in the true philosophical method began the moment I threw overboard verbal solutions, having found in the inner life an important filed of experiment. After that, all progress was an enlarging of this field.

According, contrasting with the beauty of the definite, while empty all-inclusive concept, the possible answers ensuing from experiential metaphysics, are unfinished, but they push strong roots down into the real. To sum it up, "The truth is that an existence can be given only in an experience. This experience will be called vision or contact, exterior perception in general, if it is a question of a material object; it will take the name intuition when it has to do with the mind". As for how far intuition might go, Bergson consents that it alone will be able to say.

> It catches hold of a thread: it is for it to see whether this thread goes as far as heaven or stops at some distance from the earth. In the first case, metaphysical experience will be bound up with that of the great mystics: I think I can state for my part that the truth lies there. In the second case, these two metaphysical experiences will remain isolated from one another without being mutually repugnant on that account. However one looks at it, philosophy will have raised us above the human state (BERGSON, 2002, p. 50).

How contrasting is the outlook of a philosophy of consciousness which insists in reducing the intuition of the mind and our intentional states into the physics of the brain and the sociobiology of functional fit. We already saw how Crabtree (2006) condemned the authenticity of the moral act, to affirm the only authority of the pleasant neurochemicals, whereas Hardin (1993, p. 15) denounced any attempt to conjure up a moral conscience in others, for the matter of addressing our commons dilemmas, as "tempting to anyone who wishes to extend his control beyond the legal limits". But in going more deeply into this point, Pinker inform us that "The computational theory of mind does more than explain the existence of knowing, thinking, and trying without invoking a ghost in the machine [i.e., that that you call you]. It also explains how those processes can be *intelligent* – how rationality can emerge from the mindless physical process" (PINKER, 2003, p. 32-33). He adds that, "Cognitive neuroscientists have not only exorcized the ghost but have shown that the brain does not even

have a part that does exactly what the ghost is supposed to do: review all the facts and make decisions for the rest of the brain carry out. Each of us *feels* that there is a single "I" in control. But that is an illusion that the brain works hard to produce" (p. 42-43, italics in the original).

I cannot stop asking why it would be so that the brain works so hard. Does the brain have intention of producing such illusion, or any other intention from which the "I" emergences as an unintended result? What is the point in negating reality to our intuition of self and its intentional states in order to donate intentionality to the Brain? Of course, this is not what Pinker means by "the brain works hard to…". This should be regarded as a mere "habit of speech", produced along the mindless evolutionary process that, likewise the brain, worked hard to produce the brain itself, which works hard to produce the illusion of a single "I" that, in turn, works even harder to dispel the illusion of being himself. It is worth of notice, incidentally, that while some neuroscientists and mathematical physicists posit that the explanation for such ultimate illusion of the self and consciousness may rest somewhere in level of the quantum phenomena in the brain, some quantum physicists put forward that conscious awareness provides the condition of possibility of all that we perceive as real. More specifically, a living observer is posited to be required in order to collapse the superposition of wavefunction that characterizes events in the quantum domain, so that matter (i.e. form) emerges truly in consciousness, not as a reflection of some independent world (e.g. RUSSELL, 1992; GOSWAMI, 2003, 2005; see also McTAGGART, 2002).

Please note how much in line this proposition goes with Bergson's account of the nature of intellect, and how the elucidation of the current mystery of consciousness passes first through the recovering of the status of reality to interiority. As Wilber (2000b, p. 174) puts it, the mind-body puzzle "is a peculiar lesion in the modern and postmodern consciousness, concomitant with the collapse of the Kosmos into the flatland". He notices that the *differentiation* of mind and body, which is at least as old as civilization, has never bothered anybody before. According to him, the mind-body dilemma (the "world-knot")

arises from the *dissociation* of mind and body, owing to the subtle reductionist attempt to translate all interior realities and qualitative distinctions into objective observables and quantitative measures. As already indicated, the interior realities, such as that there is an "I" in control of the body, as still felt and strongly intuited. However, Wilber explains, these realities are faced with the objective world, as described by science and thought to be ultimately real, and where the brain is indeed just a part of the biological organism. Consciousness is thus presumed to somehow be a function of that brain (the ultimate real). But as Pinker and other authorities inform us, there is absolutely nothing in the brain that even vaguely corresponds to the *qualia* or experience or realities of the mind and consciousness. We must then either reduce consciousness to brain (and thus deny consciousness in its own terms, resulting is the downgrading of the "I" and of our intentional states to the condition of illusions that both the Brain and the Society work hard to produce, nobody knows why), or accept the dualism as real, whereupon we can't even explain how I can move my arm (or how one reality affects the other). What Wilber suggests is that both those solutions occur within the flatland paradigm (WILBER, 2002, p. 179).

I will not go along Wilber's technical details on how to untie such world-knot. In quite general terms I must simply indicate that, according to Wilber's account, the mind-body problem involves three dilemmas, namely: (1) how to relate Mind (interiors, including the modes of apprehension of every holon or sentient being) and Body (exteriors, including brain); (2) how to relate mind (interior conceptual consciousness) and body (interior feelings); and (3) how to see the final relation of Mind and Body (subject and object). To put things schematically, Wilber suggests that those three items can be approached, respectively, like this: (1) acknowledge that every exterior has an interior (all the way up, all the way down), with binds Mind and Body; (2) acknowledge that there are interior stages of consciousness development (i.e. formal operational stage transcends and includes concrete operational, which transcends and includes vital feelings and sensorimotor awareness) which binds mind and body, so that the mind can causally operate on the body (i.e. formal operational can

operate on concrete operational, which operates in sensorimotor, as every developmentalist knows); and (3) acknowledge that there are higher levels of consciousness development, which finally unites Mind and Body (does preventing any form of dualism) (WILBER, 2002, p. 179, see also note 15, ibid, p. 276).

As indicated, to proceed along those three steps we are invited to rise above flatland and embrace a more integral approach. If we limit ourselves to the level of the Human Holon, the first step may not be a major problem, even if the physicalist (or materialist) will insist about the illusory character of human interiority and intentionality. Likewise, the second step does not command any paradigm change, as the stage developmental psychology provides abundant evidence on the holarchical structuration of a number of developmental lines, including the development of the moral reasoning (Kohlberg) and value systems (Graves). However, here a major difficulty comes from moral noncognitivism, which is moreover directly associated with the instrumental conception of rationality. I have already indicated that this is *not* a problem to acknowledge interiority (which is step one), as the existence of intrinsic values and other-regarding preferences, even if some researches may see them as "image and illusion", as they most conform with the "universal egoism" theory. The problem is that, under the canon of moral noncognitivism and instrumental conception of rationality, colored with the tones of political correctness and cultured relativist reasoning, the stage developmental structuration of value systems, moral reasoning and worldviews, seen just crude and unacceptable. Respecting such concerns, I will in the next two chapters do my best to show that the stage developmental psychology and the paradigm of adequate structuralism are firmly grounded in empirical evidences. In addition, I shall add to them my own findings, to be presented and discussed in the chapter six. Finally, concerning the third step (the existence of higher levels of consciousness development), Bergson's report on his own path may suffice, on the one had, to prevent me from attempting to provide whatever conceptual definition of such levels. My own experiential limitations, on the other hand, impede me to provide interesting

descriptions or characterizations, which can be found in a number of other sources anyway (cf. WILBER, 2000a). Notwithstanding, I would like to point out that the logical necessity of the existence of such higher levels is inherent to the holarchical structure of reality, and predicted by the Incomplete or Uncertain (IOU) tenet of Integral Theory, which despite of be a theme found in virtually all of the mystically or contemplatively oriented philosopher-sages the world over, as well as a general theme running throughout the Idealist writers, gained much more respect when it shows up in mathematics as Tarski's Theorem and Gödel's Incompleteness Theorem.

As Wilber (2006) explains, the introduction of the IOU principle in mathematics "initially caused an uproar, almost panic, because it meant that set theory and arithmetic (and by implication, the whole of mathematics) were on very shaky ground – that is to say, on self-contradictory ground". While the truth value of such self-contradictory ground cannot be denied, since it was mathematically proved, ironically, mathematics "escaped" the paradoxes by postulating unendingly expanding sets ("transfinite"). "Put rather simplistically", Wilber explains, "the only way for mathematics to avoid profound self-contradiction is to postulate a yet higher level of inclusion, which avoids the paradoxes of one level – but then faces the same paradoxes on its own level. Another yet higher level is thus postulated, and this continues endlessly". In other words, both mathematical paradoxes of some level and the incompleteness of whatever level of consciousness – and indeed of every holon "all the way up, all the way down", in the four quadrants – can be superseded *only* at the next higher level (the next more-encompassing set). But that set then faces its own IOU, that is, it is either incomplete or inconsistent, and that continues... forever. "The sets must be postulated to expand forever, because the moment they stop, mathematics becomes self-contradictory". Wilber observes, further, that "Nobody ever actually sees all of these transfinite sets: they are just a promissory note that allows mathematics both to keep going and to get going in the first place" (WILBER, 2006).

We shall see in the next chapter how such transfinite postulate confers one of the uniqueness of the Clare Graves's Emergent-Cyclical Double-Helix Levels of Existence Conception of adult psychosocial development.

4. THE DEVELOPMENTAL STRUCTURE OF VALUE SYSTEMS AND WORLDVIEWS: CLARE GRAVES AND THE SPIRAL DYNAMICS

In this chapter I present Clare Graves's Emergent Cyclical Double-Helix Levels of Existence Theory (ECLET) of development of adult personality and cultural institutions. The ECLET conception supplies us with the theoretical constructs (six selected out of eight *biopsychosocial* levels or waves of existence) we assessed in our participants during the experimental phase of this study. The issues related with the assessment tool and procedures will be discussed in the next chapter. Here I begin by clarifying on the meaning of the structural constructs associated with stage developmental approach in general and the importance of the paradigm of *adequate* structuralism. This is important in view of some misunderstandings ensuing from early identification of such structures as "wholes", so that together which the recognition that every "whole" is at same time a "part" of something else the "poststructuralist" movement alleged having gone beyond structuralism altogether. I shall then rely on Wilber's account that this is not the case, even if so-called poststructuralists have indeed pointed out some missteps of an *inadequate* early structuralism.

This is made an issue of course for the sake of ascertaining the reliability underlining Graves's conception, which is presented here in fairly detail. It is also vital for the sake of understanding the limits and possibilities of our

generalizations, as well as those of Graves's himself, ensuing from both the relative independent development of different *lines* of the interior milieu, and the way as structuralism and phenomenology naturally go together. In this regard, I claim that the ECLET conception can work as *proxy* for the most general trends associated with the awakening of human consciousness, and connects with the farther reaches of human potential, as reported by the spiritual traditions of all the times. Notwithstanding, the focus goes upon the extrication of the usefulness of Graves's ECLET conception for the matter of addressing collective action dilemmas – an endeavor which we will complete when specifying our testable hypothesis, in the next chapter.

4.1. Entering ZONE # 2: the meaning of structure and the importance of post/structuralism

Following Wilber (forthcoming, excerpt D), let's start by recalling that zone #1, or the interior seen from within, is a first-person experience of a first-person reality, whether singular (I) or plural (WE). In Figure 3.7, this means anything seen from inside or within the boundaries of a holon in the Upper-Left and Lower-Left quadrants. In Figure 3.8, the major methodologies enacting zone #1 are given as interior phenomenology and hermeneutics, respectively. This is, in fact, a first-person experience, or knowing by acquaintance, of that which confers to the holon its wholeness aspect, its internal consistence and coherence, its identity and agency capacity.

Zone #2 is simply those same holons seen from the outside (or seen from without). Since all of these Left-Hand holons are interior realities, one cannot see their insides or outsiders in the exterior, sensorimotor or empirical world. As Wilber explains, "interior" classically means first-person, and "outside" classically means third-person. Thus, zone #2, or an "outside view of the interior", means *a third-person approach to first-person realities*. As such, the methodologies addressing zone #2 supply an objectivistic description of that same realities known by acquaintance.

This is the distinctive hallmark of all zone #2 paradigms: they are third-person approaches to realities that I have some sort of access to in first-person modes (as such, this approach is quite different from third-person approaches to third-person realities, which is typical of most forms of system theory). In Figure 3.8, "structuralism" is listed for the outside of the individual interior, and "cultural anthropology" for the outside of collective interiors. Wilber explains that structuralism can be, and is, used in both, but that the complexities of collective holons render structuralism just one of the many useful tools in cultural anthropology, whereas for the outsides of individual interiors over time it has no viable competitors and thus is listed as the exemplar paradigm of zone #2 in the Upper_Left.

Wilber (excerpt D, p. 14) underlies the place of zone #2 among the indigenous perspectives thus: "Zone #2 is a wonderfully important event horizon because, in an AQAL matrix of indigenous perspectives, this zone highlight, enacts, and brings forth those occasions that help me to reconstruct the interiors of another sentient being so that yet further forms of mutual understanding and compassionate embrace can stand forth in a Kosmos of radiant regard". Structuralism, as a third-person view of first-person holon attempts to offer a *reconstructive* account of the pattern or agency of that holon's interior. That is, it attempts to discover, describe, or elucidate the "internality codes" of a holon, or the rules and patterns that the subholons internal to that holon are following; in this case, the internality codes of the interiors (I or WE) of the holon. As Wilber puts it, "the simplest definition of structuralism is *interior holism*.

A structure or internally pattern of a holon will almost always be some sort of holistic configuration, for the simple reason that the holon must hang together in order to endure, that is, it must have some sort of unity or wholeness in order to exist as an entity. But a holon is a whole/part, and the "structure" of a holon refers to the "whole-ness" or unity aspect of the holon. Then, the patters that the structuralist attempt to elucidate are patterns of things in constant dynamic flux, and what remains stable and unchanged is indeed the pattern of the change itself. The key point is that, precisely because the things stuffing such

dynamic flux are whole/parts, the dynamic itself presents a directionality which is imparted by the very nested structure of holons within holons within holons, indefinitely. Thus, alongside the permanence of the dynamic pattern itself, the flux reserves to the compound holons either transcendence or disintegration.

In a very broad sense, Wilber usually refers to such patterns as *probability spaces*. Thus defined, a "structure" refers to the enduring pattern or regime of any holon in any quadrant – such as the structure of a molecule, the structure of a town, and so on. In this sense, the researcher would be asking, for example, how can some bacteria remain essentially the same for a billion of years when all their components change ceaselessly? In this case, the paradigm of autopoiesis is the specific methodology for elucidating the internality codes of the holons which are registered, by our sensorial means, as pertaining the objective world (Right-Hand quadrants).

In the narrower sense, structure means as interior (subjective) pattern, which is elucidate by the paradigmatic practice of adequate structuralism. Accordingly, the "structure" of an individual agency (Upper-Left) and/or a cultural nexus-agency (Lower-Left) is simply the probability of finding, in a particular locale of the interior dimensions of the AQAL matrix of indigenous perspectives, the corresponding behavior that is described or defined as "within the structure". In this last sense, structures are themselves a coherent whole that help to enact and bring forth a world that is a co-creation of those structures doing the perceiving, knowing, and felling. Accordingly, as Wilber explains, unlike early structuralists, the paradigm of adequate structuralism (which many regard as post-structuralism) recognize that "structures are con-structed", which means that "the word is not given, but constructed". It is worth noticing, at this point, that this view is but the enactment principle in heart of Wilber's Integral Methodological Pluralism. Moreover, he adds that such understanding was already present in Piaget's perception that different cognitive structures enact and bring forth different worlds, and that this "was much more than the mere rhetorical assertion, offered by other postmodernists, that intersubjectivity creates worlds and hence knowledge in socially constructed". Instead, "this was a highly

meticulous research into exactly why and how that construction of reality occurs. In particular, facing the radical relativist position, Graves (2005) asserted that, as a contribution to developmental stage theoretical psychology, his system deals with at least five of the major criticisms generally directed toward it. First, it offers an explanation of how the constructs (i.e. the subjective structures) develop. Second, it present a picture of what the process of development is like. Third, it hypothesizes what factors determine the hierarchical order of the constructs. Fourth, it explains what determines the particular characteristics of constructs. And, fifth, it suggests how the constructs operate.

Thus, structuralism, as social practice or paradigm, highlights those dimensions and perspectives of interior holons that involve the patterns, rules, or regularities that they display. When done adequately, Wilber says, structuralism does not impose these rules but truly disclose them. Moreover, he continues, "these pattern and regularities cannot be spotted by phenomenology, hermeneutics, or systems theory, which becomes particularly obvious when we look at complex social interactions, such as those embodied in language, because part of their existence involves indigenous perspectives not activated by those other inquiries" (Wilber, excerpt D, p. 19). Wilber adds that, it is by using the paradigm or social practice of adequate structuralism that we can determine, for example, the steps necessary to develop ecological consciousness, given that ecosystems themselves do not produce ecological consciousness nor explain it.

But how does structuralism do this? How does it disclose these invisible regularities? As Wilber explains, structuralism is basically *phenomenology plus history*. That is, it starts with phenomenology (and hermeneutics) – or any first-person interior realities – but then follows the phenomena over long periods of time and attempts to spot the patterns that the phenomena may eventually follow. These patterns represent the *internality codes* of the interior holons being engaged. As we saw, every holon or stable entity has some sort of identity or agency, that is, it has some sort of wholeness or some sort of coherence. Structuralists thus attempt to identify or disclose the nature of that wholeness in

the interior domain. That is, they attempt to elucidate the wholeness aspect of an interior whole/part or holon.

Wilber (excerpt D, p. 22) clarifies that early pioneering structuralist such as Levi-Strauss, Ronald Barthes, early Foucault, and Lacan, among others, unfortunately, as often happens owing to their pioneering but less-than-adequate paradigm and theories, came to define "structuralism" as a whole. As a result, when the "post-structuralists" came along, they interpreted poststructuralism as going beyond structuralism altogether, whereas it was simply trying to go beyond *inadequate* structuralism (and ended up beneath adequate structuralism).

Finally, Wilber lists Clare Graves among some of who he grants as truly brilliant structuralists, including Jean Gebser, James Mark Baldwin, Jean Piaget, Lawrence Kohlberg, Abraham Maslow, Erik Erikson, Robert Kegan and Jane Loevinger. We already gave a quick look into the contributions of most of them when discussing the developmental streams of consciousness in the last chapter. Let's now see the basic steps of adequate structuralism before going through the specifics of Clare Graves's ECLET conception.

4.2. Basic steps in the paradigm of adequate structuralism

We must start by recognizing that, even when structuralists are focusing on individuals, they usually begin by studding large *groups* or aggregate of individuals. Wilber explains that they do so for several reasons. Among the most important, the first has to do with the time required to observe the behavioral *transformations* associated with the dynamic process the structuralist is attempting to disclose. Since the majority of the observable personal transformations generally take many years to develop, a longitudinal study following the transformations in only one or a few individuals would last for decades in order to actually reveal any significant transformation or development. Thus, by studding large groups of individuals, the researcher will catch many of them undergoing transformations, and hence s/he can study the development of structures more easily. In this regard, Clare Graves collected his

basic data during nine years, until he have worked with a total of 1,065 subjects (HURLBUT, 1979).

Also, even if acquaintance with the interior structures being described are essential in the paradigm of adequate structuralism, as a third-person approach to first-person realities a fair amount of the descriptive undertaking can be engaged in *without personal transformation on the part of the researcher*. In other words, the third-person or descriptive component of structuralism, by *temporarily* removing the structuralist from the burden of first-person transformation, allows the researcher to follow and observe various outward aspects of the development of large number of structures and stages that he or she would never be able to observe if confined to the necessities of only first-person methodologies (such as introspective phenomenology).

Yet, since the structuralist, unlike systems theorists, is *not* studding merely the *exterior* behavior in order to observe any *exterior* patterns (objective or interobjective), but *exterior* behavior in order to deduce *interior* patterns (subjective or intersubjective), the structuralist must at some point work *within* a hermeneutic space. That is, she or he must have some degree of first-person acquaintance concerning the structures being described.

With those qualifications in mind, the first step in adequate structuralism is simply to pose a dilemma (or working questions) to a very large group of individuals and then note, say, their verbal and cognitive behavior in response to that dilemma, and see if those responses show any pattern (or fall into any types or classes). If the structuralist notices any general classes of response, she then might follow that same group over a period of a year or more in order to observe whether there are any sort o patterned transformation on those responses. If it is a large group, and if the responses that the structuralist noticed are actually stages, the researcher must be able to identify *directionality* in the way her or his subjects change the responses along time. Again, if it is a large group, the directionality found by the structuralist should suffice to establish a true stage sequence, that is, one that cannot be altered by social or environmental conditions.

Wilber (excerpt D, p. 55) explains that, "the reason [real stages] cannot be altered by social conditioning is the same reason that the sequence 'atoms to molecules to cells' cannot be altered by environmental conditioning", that is, one can't have cells first and then atoms, because cells are composed of atoms. True stages, in any realm, are compound individuals that become ingredients, elements, or subholons is succeeding compound individuals, and the sequence cannot be altered without destroying it. In the paradigm of adequate structuralism, *this* is what bestows hierarchical structure to both the psychosocial development of the individuals and sociocultural evolution, and that bothers so much the radical relativists. So, after having followed and watched very carefully the sequential relation of the classes of responses, and if they do indeed emerge in a sequence that does not seem alterable by environmental conditioning, then the structuralist provisionally accepts that these classes of responses are stages in a developmental sequence of some sort, at least for this group.

At the next step, structuralists generally attempt to extend their studies to larger groups in an attempt to determinate how "local" or how "universal" these stages might be. Wilber (excerpt D., p. 56) notes that "This is a purely reconstructive inquiry after the fact" and it is an empirical inquiry in that sense. He adds that some *stages* apply only to a few people, some to small subcultures, some to cultures, and some to humans in general, but this is a matter of *actual research* by those versed in social practice of adequate structuralism (coupled, moreover, by adequate hermeneutics). Wilber adds further that "No competent structuralist has ever implied stage sequence for individuals without appropriate evidence".

So, if the responses the researcher has found continue appear to be stages – whether local or universal – then at some point the structuralist will very likely attempt to zero-in on the actual *structure* of each stage itself, which is indeed the heart of structuralism and, also, when it must directly reconnect with hermeneutics, so that in fact structuralism begins and ends with hermeneutics. Otherwise put, in order to finally and fully describe the distinctive capabilities, typical contents, mindsets, corresponding intentional states, and other traits that

outline the internal codes of any given stage of any particular line or stream of interior development, the research must know that structure both from within (by acquaintance) and from without. This requirement can impose obvious limitations for the adequate elucidation of the higher stages of consciousness development, which may indeed present quite distinctive traits, whereas being ungraspable from the structuralist's perspective. When this is the case, these levels may be reworked and refined according the contributions of other researcher. That is precisely the idea under Wilber's additions presented in Figure 3.15. In this regard, if one goes by the levels described in the Buddha School, there are at least three great enlightenment attainments status which are truly above the heads of ordinary people, namely: Arhat, Bodhisattva, and Tathagata. Pretty much like official structuralism, the Buddha Scholl elucidates and distinguishes the internality codes corresponding to such enlightenment stages.

Notwithstanding, since what is being elucidated are the interior patterns of a living and creative dynamic flow of consciousness awakening, which co-enacts worlds in the AQAL matrix of indigenous perspectives, it may well be the case that the internal codes of such farther reaches of human potential were actually not quite well defined – for it would be necessary have more individuals operating from those such perspectives so as to co-enact new worlds and better delineate the corresponding internality codes. So, keeping in mind that all holons are agency-in-communion, or structures-in-exchange – where "structure" means the defining agency, deep features, enduring patterns of any self-organizing holon, and "in exchange" refers to the fact that all holons possess not just autonomous agency, but also exist in networks of open communication, relationship, and embeddedness – and that the Gravesian psychosocial value stream is only one of the many developmental streams of consciousness, which takes its rightful place with the Piagetian cognitive stream, Loevinger self stream, Kohlberg moral stream, and Marlow needs stream (see Figure 3.15), lets' see how he went through the basic steps of adequate structuralism and the stuff of his

constructs, in which we grounded our experimental approach to the commons dilemma.

4.3. The phases of Graves's research throughout the steps of adequate structuralism

As we saw, the first step of adequate structuralism is to pose a dilemma to a large group and notice any class of responses, the second is to make sure that those classes are actually stages, the third is to perform a cross-group studies to see how widespread those stages are, and the forth is the attempting to elucidate the internal codes of those stages, which is indeed the core and major step in the paradigm of structuralism. We also saw that this last step reconnects the structuralism with hermeneutics, that is, the third-person description of first-person realities with the first-person acquaintance of those same realities.

After consider various alternatives, Graves ([1971] 2004) decided that the conception of psychological health would work as a proper axis for a systematic investigation whose ultimate purpose was to find reasoning for the fact that people have so different views about almost everything. With this concept as vehicle and having decided to work with adult population (he mentions his subjects were 18 to 61 years of age, both male and female), Graves initiated what he called "a piece of scientific research" by stating the following six primary specific questions:

1. How do biologically mature human beings conceive of what is the healthy personality? And, secondly, he asked himself:
2. Do biological mature human beings have, basically, one major identifiably conception of what is the psychologically healthy adult?
3. Do biological mature human beings have more than one conception of what is the healthy personality?
4. If adults have several conceptions of healthy personality, can the conceptions be classified into groups of similar conceptions?
5. If the various conceptions are classifiable how can they be classified?

a. Can they be classified content wise? – If so, how?
b. Can they be classified structurally? – If so, how?
c. Can they be classified functionally? – If so, how?

6. Will there be evidence that some one conception of healthy personality will stand out as superior to other conceptions of healthy personality? (At least according to the accepted psychological theories of personality).

In order do collect his data, Graves explains that he "didn't want people who knew anything about personality theory" but also he "didn't want people who were very sophisticated in psychological knowledge". Since he could not conduct any kind of research outside of the classroom (Graves was carrying 27 hours at that time), he decided to initiate the research with his students in a class entitled "The Theory of Normal Personality". However, instead of provide lectures about the topic he went into the classroom on the first day and said to the students:

> *I do not want you to read a thing during the first five weeks of this class. I want you to be no further along than you are at the moment you came in here in any kind of study of the filed of psychology or the field of personality.*
> *We will take the first five weeks to talk in class about what people would put into a conception of a healthy personality. What are the things you would get in such conception? And you people will share this information with one another. I will just be in here seeing to it that you people stay on the subject of talking about what is a healthy personality; what one ought to include when one talks about health personality in operation.*
> *As we proceed though this five weeks I want you to think about your own personal conception of what health personality is; and at the end of these five weeks present to me your paper on your personal conception of what is a health human being, the psychological health human being, in operation.*

With the individual students' papers in hand, Graves explains that he was able to set up a "complete intercommunication system" in other to investigate:

1. What will happen to a person's conception of healthy human being behavior when he is confronted with criticism of his point of view by his peers who have developed their own conception of psychologically healthy behavior?
2. What will happen to a person's conception of healthy human being behavior when he is confronted with the task of comparing his conception of healthy human behavior to conceptions which have been developed by authorities in the field?

3. How will subjects behave when under peer and authority criticism?

Such strategy resulted three different sets of data: (a) the very basic data consisting of the original conceptions, (b) the modification and defense under peer pressure and (c) the modifications and defense under influence of authority.

Graves thus took the original conceptions and handed them to a group of people who were entirely independent of the study and ask them to classify the papers in any way the person judge possible. He explains that he repeated the whole process with eight different groups and used different people as judges each time. But the classification process was cumulative, so that Graves as able to check the consistence of the classifications according the amount of data was increasing[46]. Besides the classification by independent judges, the data processing involved: (a) the study of the classifications in relation to peer criticism, (b) study of the classifications in relation to authority criticism, (c) study of peer interaction data, and (d) cross checking results of (a), (b) and (c), with randomly selected through interviews.

At his point, after all runs, the diversity of original conceptions of psychological healthy behavior was so great the only 60% of the conceptions fell into two major categories of two subtypes each. The main categories clearly opposed the conceptions that the healthy personality is able to *deny/sacrifice the self* to the conceptions that the healthy personality *affirms/expresses the self*. The deny/sacrifice self category presented two different rationales related to the time of rewarding: *deny/sacrifice for reward latter* and *deny/sacrifice for reward now*. The affirm/express self category presented two subtypes defined by the positioning regarding the well fair of others: *affirm/express self in a calculating fashion and at the expense of others* (without shame or guilt) and *affirm/express self but not at expense of others*.

As suggested since the Introduction, one can easily figure out the promise of such basic structure on the inquiry on collective action. Notwithstanding, by studying the relations of such categories with the peer and

[46] Graves continue to gather the same type of primary data year after year until he had collected 1,065 conceptions (HURLBUT, 1979).

authority criticism Graves observed that the *deny/sacrifice self system for reward latter* change to the *express self as self desires calculatingly/rationally without shame or guilt*, which is turn change to the *deny/sacrifice self to get acceptance now*, which finally change to the *express self but not as the expense of others*.

This structure of change provided the basics for Graves's conceptualization. He conceived that, rather than a definite state, the psychological healthy describe better a process and that the students' conceptions of psychological healthy were actually reflecting their on personality systems. In a latter stage of this second phase (early 1960), Graves reports that, as things changed in the colleges and people who were previously deprived of the opportunity to enter into college began to appear in his classes, he started to get to a new conception – one the he have not yet previously had. Such conception started to appear enough for it to be sorted out by the independent judges a new category, which then clustered together with some similar examples scattered among the previously (40%) unclassified conceptions. This conception was *express self impulsively at any cost*. Graves reports that, when this conception changed, low and behold, it changed into the *deny self for reward latter* category. Also, in 1959, Graves says he "got the shock of [his] life when this *express self but not at the expense of others* began to change (GRAVES, 2004, p. 25). This is because the *express self but not at the expresses of others* system seemed to Graves as corresponding to Maslow's self-actualizing stage and Rogers's fully functioning person. As we shall see, this finding imparted as distinctive character in Graves's conception, as he began to see the psychosocial development of human beings as an essentially open-ended process, so that is would not exist, even in theory, any such a thing as a stage of psychological maturity. The said conception seemed to Graves as *devy/sacrifice self to existential realities*.

So, at the final the second phase of his research Graves's had the following sequence of change of the conceptions of personality health, which he was seen as a representing the patterns of an open ended hierarchical process of psychosocial development of adult human beings (GRAVES, 2004, p. 28):

> *Express-self impulsively at any cost* (latter labeled CP)
> change to

Deny/sacrifice self now for reward later (future DQ)
change to
Express-self for self gain but calculatedly/rationally (future ER)
change to
Deny/sacrifice self now to get acceptance now (future FS)
change to
Express self with concern for others, but not at the expense of others (future FS)
change to
Deny/sacrifice self to existential realities (future HU, appearing after 1959)

Graves then initiate a third phase in his research consisting in grouping the subjects with similar conceptions and submit them to batteries of standardized psychological tests and planned problem solving situation. He thus systematically observed how subjects with similar conceptions interacted with each other, how they worked toward to solution of the problems, how long, on average, it took for each group to find answers, how many solutions each group found, and the quality of the solutions.

In the total Graves had a minimum of 36 dimensions correlating with the initial four categories of personality systems. The results of phase three produced seven kinds of data (GRAVES, 2005, p. 48):

1. How subjects who had similar conceptions of mature personality operated in certain problem situations.
2. How subjects who had similar conceptions organize to solve problems.
3. How subjects who had similar conceptions interacted with one another in course of attempting to solve problems.
4. How subjects who had similar conceptions worked toward the solution of problems.
5. How long it took subjects who had similar conceptions to solve problems.
6. How well subjects who had similar conceptions solved problems.
7. How subjects who had similar conceptions of mature personality performed on certain standard psychological tests.

In view of the above, I want once more reinforce the suitability of Graves's procedures relative to out present undertaking, since our participants in the experiments must similarly organize themselves in order to unravel their collective action dilemmas. Unfortunately, owing to logistic restrictions, we have

not the opportunity of grouping the individuals which similar conceptions. Notwithstanding, the fact is that in real life situations we shall typically find people with dissimilar mindsets embattling possibly incommensurate perspectives respecting the way things should be done. And this is even more so in situations in which deontic constraints surface and bring about diverse attitudes toward the conflict between the individual and the collective interests.

As a special finding in Graves's data from problem solving, he reports that he became particularly impressed with a dramatic jump observed in the system *express self not at the expense of others*. He tells that the *express self not as expense of others* (GT) system, when presented with diverse open-ended problems situations, found more solutions than all the other systems put together! Moreover, the average period of time required to arrive at a solution was much shorter than the other groups. Further, the quality of the solutions arrived at was better than the other groups. Yet, this result could not be attributed to superior intelligence or brilliantness. Indeed, Graves explains that the people in this system were no different than people in the other systems by any psychometric measure. He thus reports that: "I had to conceptualize that a sudden and almost unbelievable change in human behavior takes place when the individual begins to believe that psychological health should be both expressive of self and taking care of other human being at same time". Besides, he continues, "When a human being starts to think is this manner, that thinking evidences a personality reorganization that is almost unbelievable in character and it must be represented in this conceptualization".

So, in the effort to understand what his data were ultimately "saying", in the course the fourth phase of his research Graves promoted a sort of marriage between the cognitive-developmental and the existential systems of thought, in order to conceive his Emergent-Cyclical Double-Helix Levels of Existence point of view, settled in thee-part premise (GRAVES, 2005, p. 29).

1. human nature in not a set thing: it is a ever-emergent, an open systems, not a closed system;

2. human nature evolves saccadic, quantum-like jumps from one steady state to another, and
3. human psychology changes as the systems emerges in new forms which each quantum-like jump to new stead state of being.

Grounding on such understanding and the extensive empirical data, Graves completed the final organization of his theory in accordance with the seven criteria listed below:

1. Conceptualize adult behavior so as to allow for no variation in certain psychological dimension, such as intelligence and temperament.
2. Conceptualize adult behavior so as to allow for quantitative variation in some dimensions, such as authoritarianism and dogmatism.
3. Conceptualize adult behavior in an altering wave like fashion, allowing for repetition of the theme, such as in change and organizational data.
4. Conceptualize adult behavior so that every other system is similar to but a at same time different from its alternate; the change data.
5. Conceptualize adult behavior so that each system has its system specificness, so that each system has a quality all its own; the interaction data.
6. Conceptualize adult behavior so that certain systems are more external oriented, so that other systems are internally oriented; the basic deny self, express self data.
7. Conceptualize adult behavior so as to show increased degree of freedom in each successive system; the freedom to behave and problem solving data.

Let us then give a glance at the outlines of the ECLET conception, as we resume the main characteristics of Graves's constructs that will support our assessment of the commons dilemma.

4.4. The emergent-cyclical, double-helix, open-ended and two-tiered aspects of Clare Graves's model

As introduced, rather than purely a psychological study, Graves's orientation integrates "bio-", "psycho-", and "socio-", and postulates that the

biopsychosocial development of human beings arises from the interaction of a double-helix complex of two sets of determining forces: the environmental social determinants (the existential problems of living), and the neuropsychological equipment of the organism for living (the biological coping system). As we saw, the starting point in the conceptualization of Graves's data was the two initial contrasting categories of healthy psychological behavior, as classified by the independent judges: the *sacrifice self category* and the *express self category*.

Yet, Graves was seeing different sorts of patterns according the various dimensions he tested along the original four categories (the two main classes and their two corresponding types). Graves had also regular socioeconomic data about his subjects and noted that the lower the socioeconomic status of the individual the more his conception of healthy personality tended to be low in the hierarchy of conceptions that he had set up. That is, the lower it was the conception of healthy personality, according to the sequence of change described above, the higher the likelihood of the individual be confronting or at least concerned with vital sort of problems.

The last notion goes in line with Maslow's Hierarchy of Necessities conception (MASLOW, 1962), and plays a decisive role in Graves's model as well. Yet, Maslow *original* conception accounts neither for the cyclical aspect nor for the multi-system and open-ended nature of Graves's point of view. In fact, Graves reports that after fighting with Maslow over these differences for about eight years they got to a final agreement along this line, in which Maslow came around to his point of view (GRAVES, 2004).

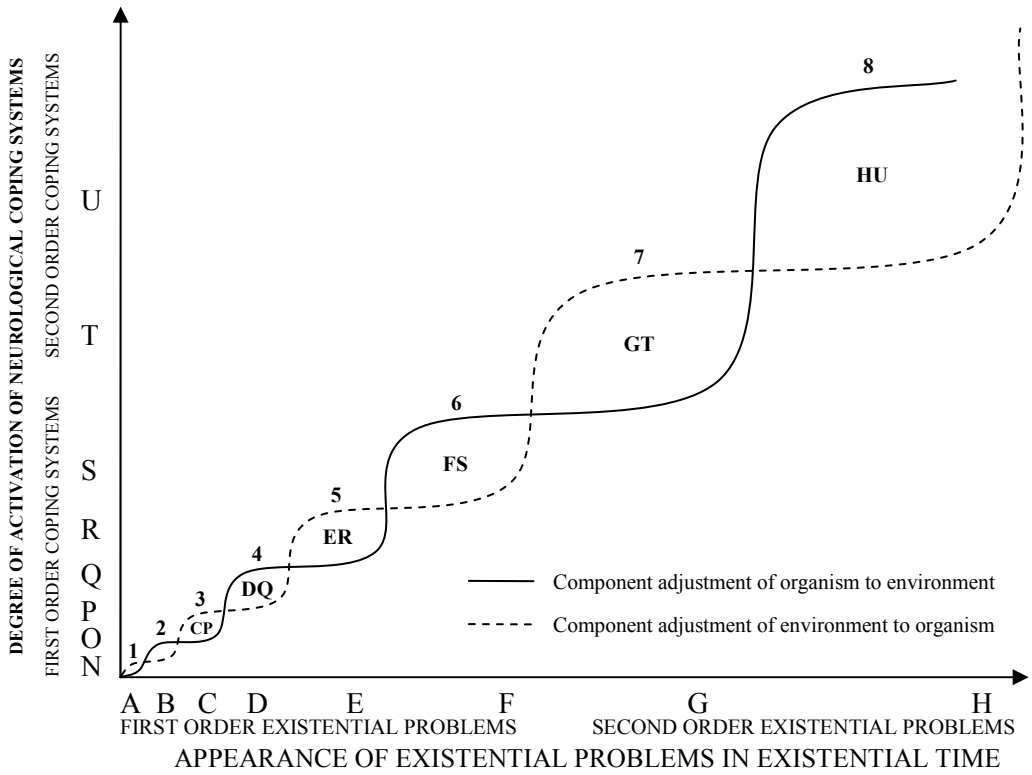
The Figure 4.1 illustrates the emergent-cyclical, double-helix, open-ended, two-tired aspect of Clare Graves's ECLET model of adult biopsychosocial systems development. The two basic opposing conceptions of healthy personality are represented by the two lines in the diagram. In this generalization, the main classes of *deny/sacrifice self systems* and *affirm/express self system* provide the fundamental tension between the *communion* and the *agency* capacities of the self, respectively. The communion capacity is presented as the component *adjustment of the organism to environment* (full line), whereas

the agency capacity appears as the component *adjustment of environment to organism* (dotted line). Accordingly, though each system must be viewed with a different premise, out of their own specific intentional states and means, the agency capacity has an *outward focus*, whereas the *locus of control is within* the individual. This trait brings forth the family of *express-self* systems: 1 – 3 – 5 and 7 (or AN, CP, ER, and GT, or yet Beige, Red, Orange and Yellow, according to Spiral Dynamics' popularization of the ECLET conception). Conversely, the communion capacity has an *inward focus*, whereas the *locus of control is outside* the individual. This produces the family on of *sacrifice-self* systems numbered as 2 – 4 – 6 and 8 (BO or Purple, DQ or Blue, FS or Green, and HU or Turquoise). The oscillating pattern comes from the posited alternation between the two basic capacities of the self along, the sequence of systems development. As Cowan and Todorovic (2005) puts it, Graves's theory puts forward that people tend to oscillate back and forth between these two fundamental stances, much like the relative position of a pendulum in its arc between "me" (agency) and "we" (communion) orientations.

At this point, it is opportune the mention again how significant the eventual actuality of such a developmental pattern is for the inquiry on collective action dilemmas, for the implications ensue not only from the structural involvement of the fundamental tension between agency and communion capacities of the self, as well as of any holon – which grants to Graves's basic conceptualization an extremely high level of generalization, universal indeed, and which no wonder reapers in the conflict between the individual and the collective interests that characterizes collective action problems – but also from the fact that such structural tension underlines a specified and hierarchical sequence of biopsychosocial systems' development. Wasn't that enough, the nested structure of the systems, as we shall see (Figure 4.2), is likewise an interior (Left-Hand) expression of an universal developmental pattern, as set forth in the Integral Theory. Finally, Graves's conception is itself an integral approach to the human holon, so that the interior pattern is posited to ensue in

accordance with the objective structure and functionality of the human brain (Upper-Right) as well as the social life conditions (Lower-Right).

Accordingly, in the Figure 4.1, the driving force of the development of the self was conceived as the *tension* (harmony-dissonance) between the neurological coping systems (vertical axis) and the existential problems (horizontal axis). When the full line spurts this represents the effort of the organism to adjust to environment, which requires deny/sacrifice self in some extend. But the organism is also active to adjust the environment in order to better affirm/express itself. This is represented by the plateaus of the full line alternated with the spurts of the dotted line in the Figure 4.1. The mechanism, hypothesized by Graves, accounting for the dynamic interaction of the coping systems and the existential problems was threefold: structural, biochemical and behavioral. The brain's structure ought to support the potential for the psychological development. The behavior would act as the main mechanism of interaction with the exterior world and the major source of information in perceiving harmony and dissonance of the organism in relation to the life conditions. The biochemistry of the brain would be the bridge linking behavior and structure. Indeed, Graves suggested that the whole physiology of the organism, in its interaction with the environment (including social), would be involved in the process of psychological changing; hence *biopsychosocial* systems.



Source: Adapted from Graves (2005).

Figure 4.1 – The double-helix, oscillating and spiraling aspect of adult psychological systems.

Finally, the increasing areas inside the lines intend to portray the increasing of the behavioral freedom observed by Graves during the fourth phase of his research (problem solving data and psychometrics assessments). Clearly, this trait ensues also from the Maslowian hierarchy of needs, underlining the ECLET conception as well. Accordingly, the nearer the diagram's origin the more vital are the existential problems one faces and the more limited are organism's behavioral freedom, owing to the burden set upon the coping systems required to handle them. As already reported, Graves data unmistakably pointed to a discontinuity involving the behavioral freedom of the individuals, when they start to conceive that the mature personality of humans beings should *express the self with concern for others, but not at the expense of others* (level 7 or GT in Figure 4.1).

As we saw, in order to portray this trait in the illustration above, it would be necessary that the area outlining the GT system be greater than the area of the six previous systems altogether. Graves connects the explanation for this finding which his view of the organism as integrated bioenergetic system. So, the dramatic jump in the effectiveness of problem solving observes for the individuals at the GT level of existence would ensue from a profound reorganization of the whole organism bioenergetic functioning, as it would no longer have to focus attention and expend significant energy on the struggles of either biological or social existence.

Such empirically grounded discontinuity accounts for the two-tiered aspect of Graves's model, where the first six levels of existence (AN to FS) pertain to the *first tier* or spiral of existence, whereas the seventh (GT) and eighth (HU) levels pertain to the *second tier* of the spiral. As we will see during the characterization of the value systems and intentional states of each system, the first order existential problems are all related with the *subsistence* (broad sense) of the organism, ranging from vital requirements to socioeconomic achievements and acceptation. On the other hand, the second order existential problems, as deduced from Graves's data, refer to the deeper problem of being – not at all departed from the psychological and moral dramas played by the deeper among Shakespeare's characters, and the views of Keats (*Mansion of Many Apartments*) Thoreau (*Correspondence*) about human nature, from where Graves went through in search of illumination to interpret his data, once convinced that they could not be rationalized within any then existing psychological or behavioral science theoretical framework (GRAVES, 2005, p. 134-140).

Finally, in a latter phase of his conceptualization effort, after he came across the HU system, Graves put forward the view that both GT and HU were indeed reviving the original AN and BO existential problems in a higher (second tier), advanced version. He then relabeled those systems as A'N' and B'O' respectively, and advanced that the overall structure of the consciousness awakening would comprise an open-ended series of harmonics of the six more basic structures. Accordingly, the full meaning of the cyclical aspect of the

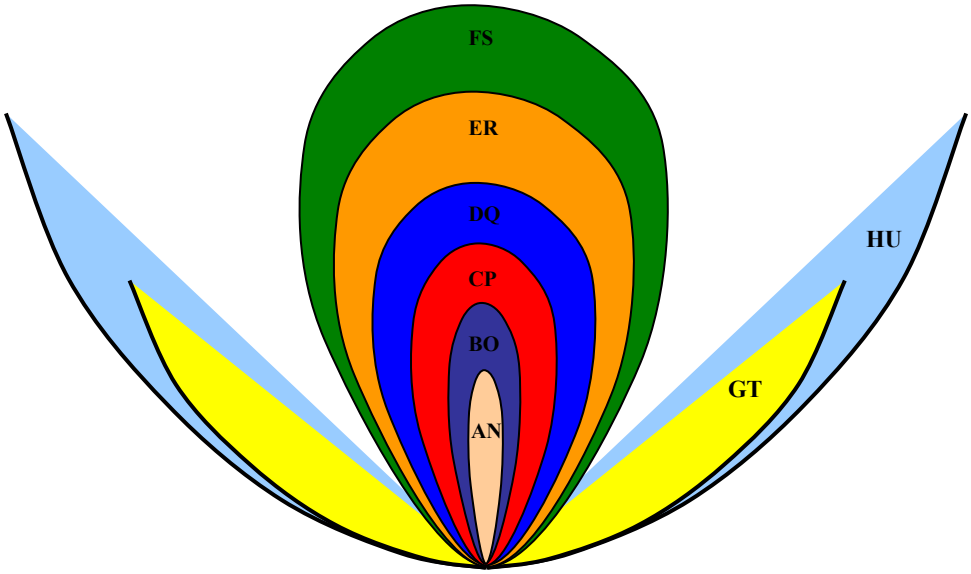
ECLET conception comprises not just the cycling of the *sacrifice-self* and *express-self* systems but also the cycling of the overall system from AN to FS, infinitely on (GRAVES, 2005, p. 368).

While the elegance of this conception of harmonics is tempting, from a theoretical point of view, I will follow the most common usage of the ECLET conception, as hold in the Spiral Dynamics (BECK; COWAN, 1996) version of it, which preserves the previous representation of both *being* levels (second tier) of existence as GT and HU, whereas keeping the significance of the change itself.

4.5. The nested or holarchical aspect of Clare Graves's model

Despite the significant change demarcating the passage to the second tier, the property of increasing behavioral freedom is part of the overall arrangement of the ECLET conception. As posited by Graves, this is associated with the nested structure of psychosocial systems development (Figure 4.2).

As we saw, the Integral Theory put forward that such structuration is indeed a universal pattern, which is then characteristic in all the four quadrants. Moreover, this pattern follows immediately from the first tenet of the Integral Theory, namely: that "reality as a whole is not composed of things or processes, but of holons" (WILBER, 2000a, p. 43). Accordingly, the four quadrants and the corresponding eight indigenous perspectives are but aspects of holons, which rigorously defined must be located in very center of the AQAL map and do not pertain to this or that quadrant (even if such way of putting things often renders expression more straightforward). This means that, properly speaking, there is *not* causality involving the interior and exterior of a holon, for both are complementary aspects of one and same thing.



Source: Adapted from Graves (2005).

Figure 4.2 – The nested aspect of adult psychological systems.

As noticed, this insight is present in Graves's conception of the organisms as an integral biopsychosocial system, so that it was lead to see a *correspondence* (not causality or parallelism) between the nested structure of the psychosocial systems development and the nested structure of the human triune brain and specificities of the biochemical neurological activation. So, Graves saw his physiological measurements as mutually confirming correlates, which do not run kind-of-parallel but instead configure the very integral systems themselves – hence *biopsychosocial systems*.

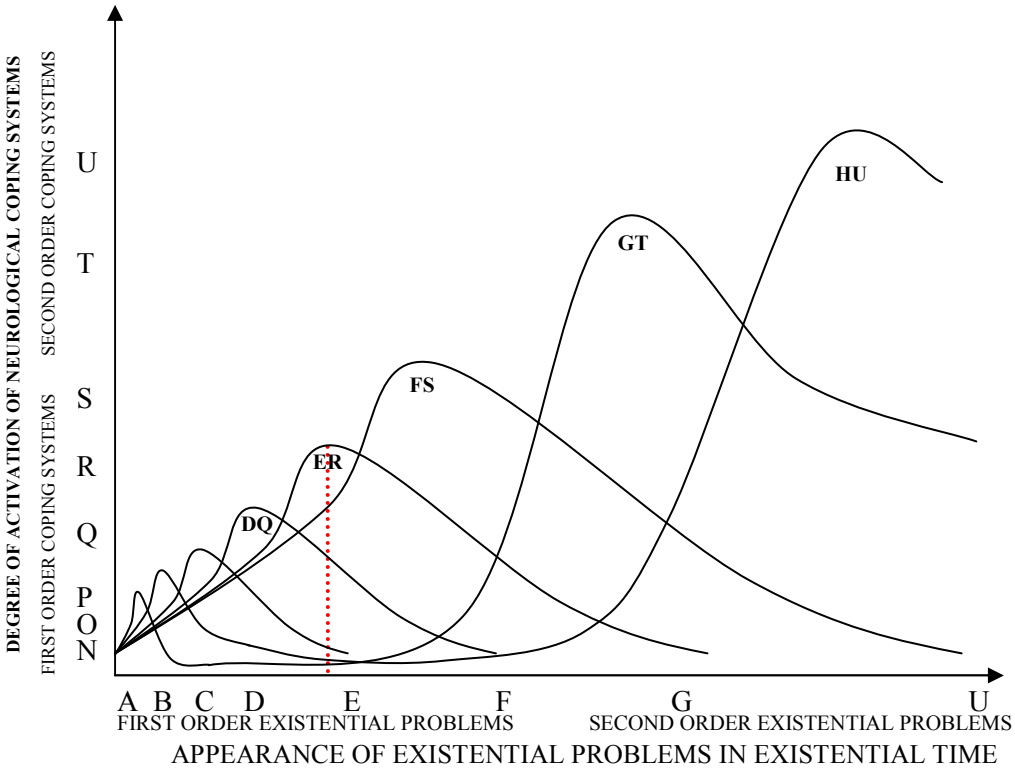
As examples, Graves reports that the adrenaline/noradrenalin ratio shifts in the passage from CP to DQ (in the CP system the noradrenaline is higher than adrenaline. This ratio reverses in the DQ system). Graves also associated the discontinuity between first and second tiers with the biochemical factor involved in the absence of impulsiveness and fear associated with the GT system. Finally, he found that the electrical resistance of the skin changes significantly at the HU level. Graves mentions that the electrical resistance of the skin in the subjects classified at the UH level was between 2 to 4 standard deviations above the mean

of the others groups (GRAVES, 2004, p. 60-68). Again, these objective correlates are seen as the exterior aspect of the interior side of the holon being elucidated. The causality or driving force of the systems dynamic would be neither in the subjective nor in the objective side but in the *permanent tension* involving the fourfold aspects of the holon, as in Figure 5.1 (i.e. the agency and communion capacities of the self, in the interior, and the neurological coping systems and environmental pressures, on the exterior).

As reviewed, most of Graves's insights and even pure speculative thought based on limited knowledge of brains functioning at the time of his research agree with contemporary neuroscience and evolutionary psychology. In particular, as hypothesized by Graves, we saw that evolutionary psychologist suggests that the human brain appears to have evolved specialized cognitive mechanisms and architecture (CLARK; KARMILOFF-SMITH, 1991; COSMIDES; TOOBY, 1989 and 1992). As supposed by Graves, such architecture would be nested or holarchical (Figure 4.2), so that, when a new system takes over, the lower level is subordinated in the newer system, not substituted for it. This means that the lower systems remain perfectly active, whereas their activities are now under the internal codes of the higher system (i.e. the same general pattern of transcend-and-include). This is why the behavioral freedom increases.

4.6. The wave-like or multilevel aspect of Graves's model and the change dynamics

Further than the existence of multiple and relatively independently *lines* of subjective development, the complexity and subtleness of the interior life is captured in the ECLET conception by its wave-like aspect, illustrated in Figure 4.3.



Source: Adapted from Graves (2005).

Figure 4.3 – The wave-like aspect of adult psychological systems.

The undulating pattern of systems' emergence and subordination imply that, at any given moment, more then one system may be relevant in determining different kind of behaviors. Nonetheless, the developmental sequence and the nested structure of progressive subordination of the lower level systems impart restrictions to the range of the possible combination of systems that characterizes, at least, a normally integrated or health personality. As well, when changes ensue, old elements take on new subjective meaning in new systems. These restrictions lead to the notion of *center of gravity* to indicate a state of biopsychosocial equilibrium which entails a limited and theoretically consistent combination of systems. For example, in Figure 4.3, the vertical line at the ER system indicates that the individual *center of gravity* is characteristic of that level. Observe that the DQ system is decreasing and that the FS system is already unfolding. In such case, the ER system is posited to dominate individual's

behavior, *provided* the action situation elicits the sort (or line) of psychosocial system and intentional states captured by the ECLET conception. So, according to the Figure 4.3, ER would be the primary systems or center of gravity, while FS, still emerging, would operate in the service of the ER as *secondary* system. DQ would still contribute to some extent for the overall personality, while likewise subordinated primarily to ER and secondarily to FS perspectives. The lower systems (AN, BO and CP) continue to play their roles and may momentarily assume the organism's control, under extraordinary circumstances (e.g. when one is before unbearable humiliation and then impulsive (CP-like) responses subordinate more reflexive and strategic (ER-like) thinking).

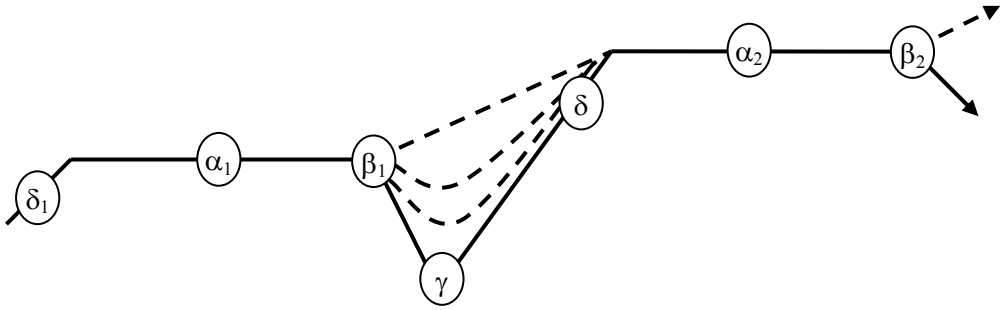
Of course, Figure 4.3 imparts a quite idealized sequence of envelopment. In reality, both the particularities of each personal history and distinctive cultural traits may bestow prominence to some special aspect of certain psychosocial systems (or of a system pertaining to any other developmental stream). According, depending on these factors, such a system may resist subordination and becomes more enduring, or even be stuck as a permanent trait. If so, along the developmental path such systems shall remain in a secondary position longer than it would be theoretically expected. As part of the development of the accessing tool that we apply in empirical part this research, Hurlbut (1979) tested and accepted the hypothesis that a person primary level of existence and secondary level of existence would be within three levels of each other. This limit range was expect to be true for 80 percent of the individuals tested (n 216). I was actually found to be true for 98.9 percent.

Yet, a persistent system may just arrest the overall personality at some proximate level of existence. This may be so individually and it is clearly more likely collectively, as a lock-in effect ensuing from positive feedbacks of mutual understanding (which we call "culture"). This is why Graves's ECLET conception is not only theory of individual adult personality development (which already integrates the "bio-", the "psycho-", and the "socio-") but also provides the foundation, at least, for a theory of the cultural evolution.

Graves went further in theorizing about the finer change dynamic comprising *horizontal*, *oblique* and proper *vertical* change. As he explains, each way-stage of adult psychology has, stylistically, its way and time integrating the whole. It is characterized by a period of preparation, a period of achievement of relative equilibrium, and a period of disintegration as preparation takes place for movement to higher stage. According to him, to understand personality we must comprehend the totality of one's systems. This totality is a totality is the sense of the momentary total state of the organism. Graves suggests that this totality of the moment, that is the organization or the internal codes around which the individual is centralized, operates by the minimaxing principle of Von Neuman (NEUMAN et al., 1947). At any moment of time (in the developmental path), the whole (whole/part, indeed) may be dominated by minimizing the growth or change tendency and by maximizing the conservation tendency. At another moment in time it will be the maximizing of growth that rules while the conservative tendency is minimized. As we saw, this is one more way of seeing the cyclic turn of agency and communion trends (GRAVES, 2005, p. 483).

Very briefly, the horizontal changes portray the dynamic of improving the capabilities within a given systems (more of the same). The oblique dynamic represents the experimenting phase with select elements of a more complex system of thinking, perceiving, evaluating, information processing, etc. Figure 4.4 illustrates the phase transitions in a generic movement upward (vertical change). The *Alphas* represent the centers of gravity of any two levels of existence, in the sequence proposed by Graves. As such, they are times of stability and congruence when solutions match problems. So, in general, the individual has no incentive to change his psychosocial being during such times. But circumstances changes and, eventually, the degrees of behavioral freedom that the individual has at a given level of development are no longer sufficient to manage problems. Alternatively, the environment may remain quite stable, whereas the interior pressure for the awakening of human potencies creates a sort of tedious discomfort, sense of purposelessness and futility. Whatever the reason, and typically out of some combination of both, *Beta* (Figure 4.4) is a time of

dissonance and uncertainty. As the theory holds, the individual will first engage in horizontal and then oblique moves, as well as in regressive searches for old solutions, but eventually take on a qualitative change, upward.



Source: Adapted from Graves (2005).

Figure 4.4 – The changing states and the alternatives paths of change.

Graves discusses six basic conditions for a vertical change: (1) structural and genetic *potential*; (2) quality and number of *solutions* (possible actions) already developed for deal with problems at a given level, (3) degree of *dissonance* between current solutions and actual outcomes, (4) *insight* into alternative means to handle problems (influx of new information and communication), (5) ability to identify and manage *barriers* to change, and (6) existence of *support* to consolidate transition (institutions included). Favorable condition shall make path smoother. In adverse conditions the depression through *Gamma* is more profound (Figure 4.4). *Gamma* is thus a state of anxiety, fear and frustration where the problems are apparent but the solutions cannot be found.

More important for our present purpose of bringing the stage developmental theory to the forefront, however, is that all the six basic conditions aforementioned (including the genetic potential) will interact with the particular change of levels at sake. The fundamental implication being is that the effect of working on those conditions will depend upon which system the individual is going from to the next system. Otherwise put, what counts as dissonance or supporting conditions, the possible actions and insights the

individual envisages, all these depend on the internal codes of the particular systems she or he is centralized in the moment.

Graves (2005, p. 486) suggests that, earlier in development, the dynamic fostered by the mutual interactions of the conditions *present in the total organism-environment* produce systems with are more automatic or mechanistic (AN, BO, CP and DQ). Later, owing to the nested (holarchical) arrangement (Figure 4.2), the dynamic interplay produces systems which are more fluid or organic (ER, FS, GT and HU). The systems tend toward openness but can, to a certain extent, congeal into machine-like systems, that is, to closed systems. Note that this is what confers to the ordinary personality – whatever the level of centralization, but especially at the first tier – its character of *delusion* a propos of which the spiritual tradition so regularly talk about, for when the human overall biopsychosocial system becomes arrested, the individual do not in reality have the agency upon either her mindset or behavior. This is likewise pretty much the same Marxist notion of *alienation*, even if the explanation is quite reduced to the exterior conditions of existence (modes of production).

Yet, Graves stresses out that neither change nor lack of change is the rule. If there are no dissonances or tensions, no change can *appear* to be the rule. If there is dissonance, change may be seen to be the rule. Further, in the evaluation whether there are disturbances or not, likewise for the other change conditions, one must remind that there are both general and specific factors which propel the organism from one stage to another. As a quite general conditions to make possible a movement up in the hierarchy, a supply of energy is necessary, and *the necessary supply of energy comes from the resolution of existential problems at a level*. Also, as the theory holds, one cannot see or conceive the possibility of higher levels until the person has reached the degree of control over current problem that hinders other possibilities. As mentioned, development may become arrested due to the formation of unquestioned assumptions and concepts for living amid the milieu of common problems, common mental models for approach to problems, and the very trend to seek optimal comfort and stability that define every level. Finally, structural potential,

such as the capacity for enduring the distress associated with repeating dissonance periods (*Gamma*) and exterior barriers to change, may let development goes up to certain level but no farther.

I regard as going unstated the way how all this stuff runs right into our conceptions of how science should be done and what counts as valid knowledge. Notwithstanding, we must now go ahead and introduce the substance or content of Graves's constructs, in which we ground the testable hypothesis advanced in the next chapter.

4.7. The waves of agency and communion in the spiral dynamics: characterizing Graves's constructs

The conceptual model of Clare Graves was popularized by Don Beck and Christopher Cowan's book entitled *Spiral Dynamics* (BECKER; COWAN, 1996). The authors, who collaborated with Professor Graves for many years, have designed a series of learning experiences around the principles of his theory to apply it directly in business and organizational life, as well as in societies at large. These include applications in leadership systems and styles, as a technology for aligning and integrating entities, community development, large-scale systems transformations, heath and wellness, sport values, education, public safety and management supervision.

The Table 4.1 resumes the main characteristics of Graves's proposed systems as popularized by Beck and Cowan's *Spiral Dynamic* (BECK; COWAN, 1996), where the eight developmental stages were pictorially stacked in a spiral-wise fashion and color-coded for convenience.

Table 4.1 – The eight stages (or living strata) of the spiral dynamic model

Stage/ Wave	Color Code	Popular Name	Thinking	Cultural manifestations and personal displays
H-U	**Turquoise**	WholeView	Holistic	*collective individualism; cosmic spirituality; earth changes*
G-T	**Yellow**	FlexFlow	Ecological	*natural systems; self-principle; multiple realities; knowledge*
F-S	**Green**	HumanBond	Consensus	*egalitarian; feelings; authentic; sharing; caring; community*
E-R	**Orange**	StriveDrive	Strategic	*materialistic; consumerism; success; image; status; growth*
D-Q	**Blue**	TruthForce	Authority	*meaning; discipline; traditions; morality; rules; lives for later*
C-P	**Red**	PowerGods	Egocentric	*gratification; glitz; conquest; action; impulsive; lives for now*
B-O	**Purple**	KinSpirits	Animistic	*rites; rituals; taboos; superstitions; tribes; folk ways & lore*
A-N	**Beige**	SurvivalSense	Instinctive	*food; water; procreation; warmth; protection; stays alive*

The first order existential problems are suggested to be all related with surviving issues, in a decreasing vital character (Figure 4.2). Graves defined the "A" problems as those involving the operation of the imperative periodic physiological needs of the organism, such as nutrition. Those problems would be handled by functional system in the brain, "N", specifically suited to the task of running vital and periodic in nature processes in the organism. The system AN (Beige) thus portrays the fundamental agency-wave, where the impulse is for sheer survival and procreation. Roughly speaking, it would correspond to the social stage of development in which people live in small hunter/gatherer bands or are forced for other reasons to spend all their vital energy in the search for physical surviving (e.g. starving masses around the planet).

The "B" sort of problems would also relate to physiological needs but in a non-periodic fashion, as in the problem of finding shelter and other security needs. Socially, the BO (Purple) system would be represented by the evolution of the hunter/gatherer bands into complex communities of tribal order. It portrays the fundamental communion-wave, where authority is unchallenged, customs are

scrupulously observed, and the individual is secure in his niche. As belief system it is associated to voodoo courses, blood oaths, ancient grudges, good-luck charms, family rituals, magic ethnic beliefs, etc. (WILBER, 2001). Thus, at this system the forces driving decisions direct to solve social dilemmas do not need to be rational, in the conventional sense, and the effect of externally imposed institutions can produced surprisingly results.

The "C" life condition is often associated with urbanization. Physiologically, as mentioned, Graves encountered high noradrenalin/adrenaline ratio in the "P" coping system, which correlates with a new agency-wave in order to assert individual survival in a new and dangerous world. Behaviorally, the rebellious, impulsive, egocentric, and heroic individual breaks away from the constraints of the tribal order. Wilber (2001) mentions as examples the "terrible twos", the rebellious youth and their wild rock stars, the epic heroes, the gang leaders, and the soldiers of fortune, among others. In terms of learning, Graves associates the CP system (Red) to the Operant or instrumental conditioning, so that cooperation in the context of commons dilemma would be expected to be rather conditional to instant satisfaction.

The environment co-created by the emergence of the CP system tends to be amoral and anarchical. The next system DQ (Blue) appears as an adaptive reaction which in turn actively creates a new rule-based order environment (often religious) in which codes of behavior are strictly set and observed. Physiologically, the noradrenalin/adrenaline ratio is reversed in relation to the CP (Red) system, as already commented. This is another communion-wave, resonating with the BO (Purple) system, and that corresponds to the original *deny/sacrifice self now for reward later* category, as yet noted. Reward depends entirely on the strictly observation of the rules and may assume a typical after-death character. The very rule-based order at the DQ stands for an institutional solution to the social dilemma set out during the previous stage. Imposition of different institutions in contradiction with the prevalent belief system would just destabilize the order and be cause of intolerant and defensive reactions.

Again, resonating with the CP (Red) agency-wave opposing to the BO (Purple) tribal order, the ER (Orange) systems emerges and break free from the stultifying rules and regulations of the previous level (Blue). Historically, the great emergence of the ER system is associated with the Enlightenment or the revolution of reason (as scientific achievements) facing the dogmas of institutionalized religion. Individually, the ER typify the achiever self. As expressed by Linscott, the adventurous individual seeks to fully harness the forces of nature for profit and individual comfort. This is the mindset which drives entrepreneurial endeavor and which (in contrast with the rigidity of Blue) is comfortable with political and other trade-offs (LINSCOTT, 2002). So, here we found the classic materialistic and individualistic *homo economicus* as an expression of the original category *express self as desired in a calculating fashion and at the expenses of others*.

The materialistic, impersonal and mechanistic cosmology which characterizes the ER system is in the root of our current environmental threat, amoral hegemony of economic power and deep social inequalities. Without breaking free from the broad cosmology itself the reaction to such situation appears as a new spurt of the communion-wave at FS (Green) level (the original *deny/sacrifice self now for to get acceptance now* category). As such, the individuals and groups pursue egalitarian agendas of justice, fairness and a sharing of resources (LINSCOTT, 2002). The Green wave is characterized by ecological sensitivity, sense of community and network. Wilber (2001) suggests that people at Green stage trend to belief that human spirit must be free of greed, dogma, and divisiveness, so that feelings and caring supersede cold rationality of Orange (ER).

At this point, providing the individual does not return to face any significant surviving restriction, her existential problem changes of nature. We have already explained how Graves's data led him to posit a qualitative change demarcating a passage to a second tier or second spiral of existence at the GT level, where the individual now needs to cope more directly with the problem of *being*, not just *keeping existing* as both biological and a social being. We also

saw that Graves had a truly integral conception of human holon, so that he posited that the qualitative change observed in the passage to the second tire ought to present some physiological correlate. He then notices that the individual at the GT (Yellow) level of existence is not afraid of noting find food (AN), or shelter (BO); neither afraid of predatory man (CP) nor of God's ire (DQ). They are not afraid of not having status or not making it on their own in this world (ER). And they are not afraid of social rejection (FS). According to Graves it would be this afraid-free condition and its corresponding physiological correlates what liberate the vital energy of the individual and explain the tremendous jump of behavioral freedom exhibited by his subjects in the problem solving situations (*expresses self but not at expense of others* category). As such, the GT (Yellow) level expresses a qualitative shift to a higher order of integrated thinking. In Wilber's words, at this point "Life is a kaleidoscope of natural hierarchies (holarchies), systems, and forms. Flexibility, spontaneity, and functionality have the highest priority. Differences and pluralities can be integrated into interdependent, natural flows. Egalitarianism complemented with natural degrees of ranking and excellence. Knowledge and competence should supersede power, status or group sensitivity (WILBER, 2001, p. 12-13).

Finally, the UH (Turquoise) level describes a state of mind which is still evolving. It is defined as a holistic and essentially spiritual understanding of the cosmos and of man's place in it. Beck and Cowan (1996) mention Gandhi and Mandela as examples of person thinking at such stage. Wilber tells that the Turquoise thinking "sees multiple levels of interaction; detects harmonics, the mystical forces, and the pervasive flow-states that permeate any organization". Again, we saw that, it terms of the physiological correlates, Graves found that the electrical resistance of the skin in the subjects classified at the UH level was between 2 to 4 standard deviations above the mean of the others groups. Besides, according to Graves's behavioral observations, people operating at UH (Turquoise) "can turn off other levels of consciousness at will", they "can go out of this world and go off into other levels of consciousness and come back at will".

Clearly, Graves's characterization of the HU biopsychosocial system provides an explicit point of contact with Wilber's validity claim for mysticism. To begin with, such "other levels of consciousness" are of course consciousness of something else. Something "out of this world". However, the trick point is that, how one interprets such "something else" depends on ones *regular* level of consciousness development. As discussed, a temporary altered *state* of consciousness can eventually place anyone "out of this world", while not at will. Typically, as already pointed out, the magical understanding of such experiences co-enacts a world as seen form the BO perspective. The DQ perspective is mythical; while ER's positivist rationality put forward a world where such "something else" is seen as mere superstition. Finally, the quite sophisticated mind still co-enacting the postmodern turn, in some measure represented by the FS system, and after having deconstructed the whole structure of psychosocial development, eventually grants a kind of incommensurable validity to such reports from "other levels of consciousness" as pure relative "social construction".

Still, what Graves's theory indicates is that HU is getting to a postmetaphysical, and certainly no magical understanding of spirituality, which Graves himself grasped out of the very symmetric elegance of the ECLET conception. Accordingly, that the individuals at the HU level of existence can go out of "this world" ought to be understood, first, as meaning that they go out of the worlds that are ordinarily shared at the lower levels of consciousness. After all, "this world" is but the world as it appears in consciousness, and they are so much undisputed as more basic are the (biopsychosocial) systems involved in subjects' awareness. This is why no one disputes the concrete reality of a rock over there, as far as everyone is able to touch it. Yet, one may have a hard time deciding whether something is green or brown with a color blind fellow who is unaware of his sight variance. Well, maybe we are unaware of our "other worlds" blindness. As well, it is noteworthy that Graves's remark on the electrical resistance of the skin of the individuals at the HU level connects with the physiological transformations reported in the *qigong* practices.

In particular, the Falun Dafa discipline (HONGZHI, 2003) is a cultivation way with cultivates both the mind and body. *Gong* is the "cultivation energy" which transforms the practitioners' body by performing *qigong* exercises. As the Falun Dafa's Master, Li Hongzhi, explains, advanced students, who have their Third Eye opened at a high-level,

> "can see that *gong* develops on the surface of a practitioner's skin and is then absorbed into his or her body. This process of *gong* generation and absorption keeps repeating itself, going level after level, sometimes really rapidly. This is first round *gong*. After the first round, the body of the practitioner is no longer a regular one" (HONGZHI, 1999, p. 51).

Of course, if one has no such Third Eye opened at a high-level he or she cannot *see* those things. Yet, there are plenty of scientific reports of physical manifestations of energy emission from *qigong* masters (cf. among many others MO et al., 2003; YAN et al., 2002; CHEN, 2002; CHEN et al., 2001; JOHNSON, 2000; YAMAMOTO et al., 1996; see also HONGZHI, 2003, p. 69-71)

Finally, as already claimed, it is not by chance that the enlightenment statuses described in the Buddha Scholl are hierarchically structured in levels of consciousness awakening. Clearly, while Graves's research methodology may not be suited to disclose and characterize such farther realms, I just suggest that the open-ended character of the ECLET conception allows its interpretation as a *proxy* of more general process of consciousness development. I don't mean that such general process follows the specificities of the ECLET conception, just presented, for these specificities picture the consciousness development as seen from a particular angle of observation – one with is of special interest for our present undertaking as it draws right into the agency and communion capacities of the self. Notwithstanding, I do imply that normative structure of morality entrenched into the Buddha Scholl, as well as in virtually all spiritual tradition, has been captured by the most general trend of *decentering of the self*, so that capacities of being trustworthy, exercise unbinding self-restraint and act cooperatively on behalf of the collective good are traits developed at the higher levels of consciousness awakening.

Let us thus move ahead to expose the research strategy we designed and procedures we followed in order to submit our propositions to empirical tests.

5. INTEGRATING PARTIALNESS

5.1. Introduction

Now we have seen how distributed theoretical developments fit in the Integral map and touched basis with the ECLET conception, we are in a better condition to attempt a more integral approach to the commons dilemma. I begin by presenting the Integral view of the human holon in the AQAL map, as put forward by Ken Wilber (WILBER, 2000b and 2001). As we shall see, although the ECLET conception imparts in fact just one perspective on the interior development of human beings, bearing in mind our present conceptual use it can work as a *proxy* of a more general developmental process. In this sense, we then bring the ECLET conception to join the IAD framework and the "cognitive institutionalism", in order to endow these approaches with an enhanced conception of the human actor. By doing so, we can broaden the conception of human rationality and disentangle it from the exclusionary instrumental approach. In turn, this move will allow us to take advantage of Habermas's theory of *communicative action*, so that we can deepen significantly our understanding on the repercussions of both communication and other institutional arrangements on the outcome of collective action dilemmas.

Yet, as indicated, a genuine Integral approach to our subject matter cannot get around the further implications of consciousness development research, since it is here that we find the bond which the interiority of our existence, whose significance for addressing our most stringent social dilemmas we lost in the course of the threefold strangeness pointed out by Tarnas (1999): the cosmological strangeness brought about by the Copernican revolution, the ontological strangeness of Descartes, and the epistemological strangeness initiated by Kant – and hopefully coming to an end with Ken Wilber.

Therefore, in order to grant to the ECLET conception the level of generality we propose, as a *proxy* for consciousness development in general, we must examine how the Integral Theory put forward a pattern with connects the evolution of the human holon in the four quadrants with the evolution of every known holon, all the way down, all the way up. Once more, this effort of generalization is not presented here as a luxury. On the contrary, as already indicated, it is essential to enhance collaborative inquiry in addressing concrete collective action dilemmas, for it is through this generalization that we may retrieve, by re-conceptualizing (i.e. transcending-and-including), the spiritual understanding of reality that is hold by most communities, instead of approaching them granting secularism as privileged epistemological stance, in order to look after some kind of social engendering in the vein of Skinner's *Walden II*.

Furthermore, an authentic integral view of the human holon is not just sustaining the principle according to which differences ought to be "respected", so as to charitably put up with superstitions and flawed beliefs. In fact, the Integral Theory is putting forward a conception of how such differences ought to be evolutionarily and developmentally structured, so that the said principle can be deepened and adequately weighted by a knowledge base which adequately supports the *prime directive*.

Clearly, the tricky point in such undertaking is how the spiritual worldview can be brought into such integral vision. Fortunately, Graves's ECLET conception already gave us a clue, as he saw in the HU (B'O') system a

cognitively higher understanding of the *spiritual nature of reality* – grounded, by the way, into quite concrete behavioral injunctions that elicit "other levels of consciousness" – an understanding which appeared earlier as *magic* at the BO level, then *magic-mythic* at CP and DQ, to be completely lost as *blind belief* and *superstition* at ER, and beginning to be retrieved in the form of *care and sensitivity* at FS level.

In the path of exposing the suggested generalization I shall add some brief comments on how other contributions of my knowledge fit pretty well with Wilber's account.

5.2. The integral view of the human holon

The Figure 5.1 offers a representation of human holon into AQAL matrix. The picture summarizes the undertaking carried out by Don Beck and Ken Wilber in refining the original ECLET conception to bring about the so-called Spiral Dynamics Integral (SDI). In this regard, it is important to emphasize that, apart from an improved characterization of the HU level of existence (holistic self – turquoise, in the Figure 5.1) and more detained scrutiny pertaining to the correspondences between the individual subjective development, in the Upper-Left, and the biological (Upper-Right) and sociocultural evolution (both Lower quadrants), the SDI refinement ultimately adds noting essential to Graves's original conception. To be sure, the four quadrants are pretty much represented in Figure 4.1, where the neurological coping systems (vertical axis in that figure) stand for the Upper-Left, and the appearance of the existential problems out of changing life conditions (horizontal axis) put across the changes in both lower quadrants. Naturally, the emergence of the psychosocial systems out the tension between agency and communion capacities of the self (in the interior of the chart) is what is being portrayed in the Upper-Left quadrant (compare Figures 4.1 and 5.1).

Of course that Figure 5.1 conveys a huge reconstructive linearization of the developmental/evolutionary processes in each quadrant—it implies neither

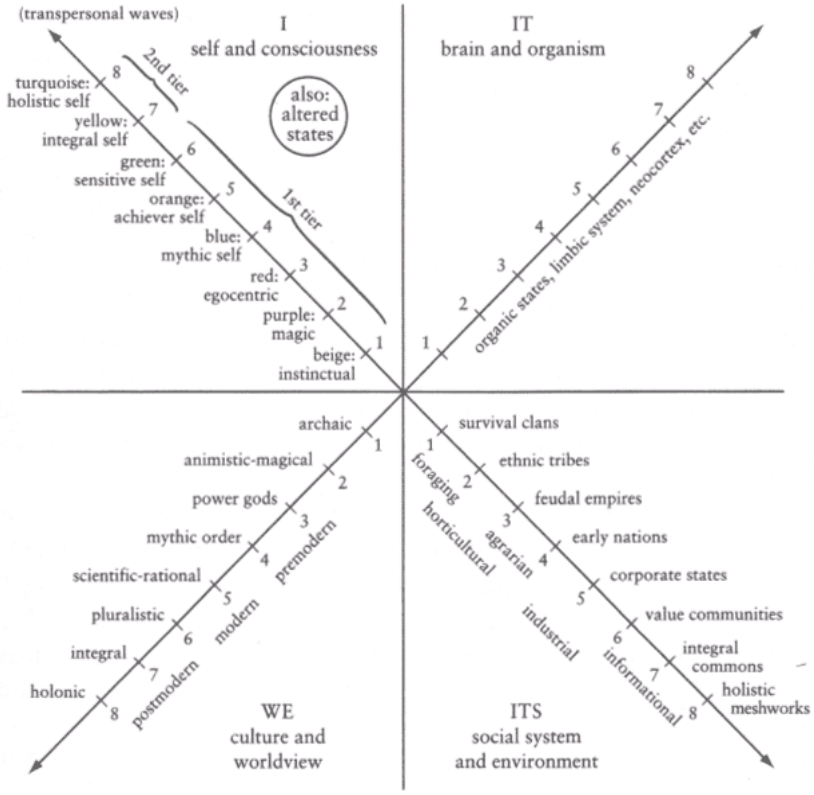
that the processes go smoothly (as they did not) nor that they go on in tandem along with the four quadrants. Besides, as Wilber (2000b, p. 146) explains, the correspondence between the individual and the collective aspects of the human holon does not mean, for example, that all individuals in a given culture or society are at the same level or wave of psychosocial existence. To say, for example, that a given society is at an animistic-magical level of development (Lower-Left) does not mean that everybody (Upper-Left) in that society is at that level. It only means that the *average* level of consciousness is generally magical, and that, more specifically, the defining laws, principles of cultural organization, and mores of everyday reality stem predominantly from the magical worldview

As Wilber expresses it elsewhere (Wilber, forthcoming, excerpt C/part3-2, p. 8) "When I am born into a society, I am landed in a labyrinth of already existing networks, individuals, cultural and social holons – a plethora of "I's" and "you's" and "we's" and "its" – and the tetra-selection pressures from those perspectives slam down on me from day one". As well, apart from secondary developments such as the growth of the body and sexual maturing, the biological *structure* of the human holon (Upper-Right) is pretty much concluded before the psychosocial (Left-Hand quadrants) development may begin. Accordingly, the interior development of each individual human holon is in fact the process of consciousness awakening, as its capacities are enacted by the said selection pressures. "Yet", as Wilber (*op. cit.*) emphasizes, "those are not merely selection pressures besieging me *from without*, but calls *from within* to awake my own indigenous perspectives and begin to inhabit them with consciousness, care, resonance, and radiance". Moreover, he adds, "That society can often cripple these native potentials and blind my own perspectives is no secret. What is less often realized or appreciated is the positive side of all those demands: the miracle of our togetherness circles as they call forth from us those extraordinary potentials that neither you nor I would ever find in ourselves without each other: the beauty and radiance that each of us is because of the other". And yet, as we saw, from the conventional perspectives on ours commons dilemmas, conscience

is supposed to be “self-eliminating”, whereas communication at the best inconsequential, but typically deceiving and “double binging.

In this regard, we shall see in a moment how vital Habermas’s theory of collective action is for a more sensible approach to communication. At this moment, what is worth noticing is that Habermas believes that communicative action represents the tail end of a developmental process in which complex forms of social action have given rise to competencies that have in turn permitted the emergence of more sophisticated forms of action (HEATH, 2001, p. 35). As clear, this is an empirical hypothesis, which can be tested, in Habermas’s view, by a historical reconstruction of cultural worldviews, just like the one portrayed in the Lower-Left quadrant of Figure 5.1. Habermas developed this argument in the fifth chapter of his *The Theory of Communicative Action* (HABERMAS, 1984), where he provides a phylogenetic account of communicative action in the form of an interpretation of work by Emile Durkhein and George Herbert Mead (HEATH, 2001). On the other hand, Habermas put forward, furthermore, that such phylogenetic path is recapitulated in ontogenesis, so that the emergence of each new form of sociocognitive competence can be observed in the growing child (HEATH, 2001). As noticed above, this onto/phylo parallel passes on just a rough approximation, for the individuals can be either above or below the collective level of development. Yet, just because the evolution of the self is in the main conditional on the existence of common significations and mutual understanding, the developmental level of the surroundings exercise important restraining effects or selection pressure upon the incidence of outliers. Hence, bearing in mind these qualifications, Habermas pursued support for the posited onto/phylo parallel in a paper entitled “Moral Consciousness and Communicative Action” (HABERMAS, 1990), where he tries to show that the stages in Lawrence Kohlberg’s model of the development of the sociocognitive and moral reasoning are isomorphic to the stages of that occur in his our reconstruction of the development of communicative action. As we saw, despite Kohlberg’s model deals with the earlier phases of development, Graves (2005) put forward the

matching of Kohlberg's construct with his own, alongside a number of other developmental models (see Appendix 1).



Fonte: Wilber (2001).

Figure 5.1 – Some aspect of the four quadrants as they appear in humans

5.3. The pattern with connects and the sociocultural evolution

As evolution in the sociocultural domain is a sensitive topic, which potential for abuse when not handled with care, we do well reminding how the above construction (Figure 5.1) ensues right from the foundations of the Integral Theory. The first point to notice is that, despite the convenience of expression leads us to talk about holons in such and such quadrant; the human holon (as any other) is actually positioned right onto the crux of the Integral matrix, so that the

four quadrants are four aspects or facets of a single, say, Human Holon, whose essential nature runs into the realm of noosphere – not in the realms of either biosphere or physiosphere, which are *parts* of the Human Holon (i.e. noosphere).

As every holon, the Human Holon display the four fundamental capacities of agency or self-preservation, communion or self-adaptation, Eros or self-transcendence, and Thanatos or self-dissolution. Also, the Human Holon is not a given but an ever-emergent patterned flow, which unfolds holarchically, transcending-and-including its predecessors as well as being subsumed into higher levels of increasing complexity, increasing differentiation/integration, increasing organization/structuration, increasing relative autonomy, and increasing *telos*. Therefore, the evolution has directionality. Also, at all levels of its depth, the individual facet is in relational exchange with the collective facet of the holon, sustaining the permanent tension between its four capacities.

Accordingly, it is in fact this patterned flow throughout the sole Human Holon (as the historical tracks of it) that is being portrayed in Figure 5.1. Sill, in order to expand on the evolution of the social systems (Lower-Right) and cultures (Lower-Left), preserving a plainer expression, one ought to bear in mind some important qualifications, so that we can, on the one hand, avoid a reified conception of the agency capacity of the communal holon (i.e. the "holons" in either one of the lower quadrants), and, on the other hand, get to an adequate conception of its self-transcendence capacity.

Then, as Wilber (2000b) explains, to say that a collective holon has agency does imply that it has a "life of its own", so that the evolutionary process in the collective domain cannot be reduced to combinations and permutations of that which occurs in the individual domain (and this is another way as the onto/phylo parallel breaks down) However, this development does not reach escape velocity from the individuals and become itself a development in a higher compound individuality or superorganism like a Leviathan. As, rigorously, the Human Holon is not in this or that quadrant but expresses itself in the entire AQAL matrix; the individual organisms rigorously are *not parts but members* of

the collective domain. The true parts or components of the systems are individuals' *interactions*.

As Wilber explains, an individual subjective holon (Upper-Left) is a member of a cultural network, intersubjective holon (Lower-Left) when its subjective intersections with other subjective holons follows the *regnant nexus* of the cultural network of with they are members. As well, an individual objective holon (Upper-Right) is a member of a social network, social system, or interobjective system (Lower-Right) when its behavioral (exterior) intersections with other organisms follow the regnant nexus of the objective network of with they are members (forthcoming, except D, note 51). It is the regnant nexus of either the social or the cultural holon that configures the agency capacity of the holon in the collective domain. It is this *nexus-agency* that has "a life of its own". Yet, one may doubt whether a nexus have agency. Does a system or collective holon have agency?

To expand on this, Wilber (except C/part3-2, p. 1) reminds us that the important differences between the quadrants need also to be factored into any discussion of agency. Accordingly, the agency capacity refers to the wholeness aspect of every holon. But since a holon is a whole that is simultaneously a part of a larger whole, its agency is always *agency-in-communion*. Also, every holon has four quadrants (i.e. it is placed in the center of the AQAL map). Thus, "Right-Hand holons" (i.e. how these holons appear when described from outside) have agency only in the exterior sense of mass-energy impacts and registrations, where they follow physical laws, habits, rules, and regulations. Consciousness and intentionality appear only in "Left-Hand holons", that is, they are the interior aspect of every holon (remember that in the AQAL metatheory holons are sentient beings, from quarks to ants to apes to humans)[47]. In other words, agency

[47] Of course, Wilber is not suggesting that quarks present reflexive intentionality as human beings (the anthropomorphic fallacy that Lovejoy called "retrotension"), but simply that all holons have their own forms of interior perception. Spinoza, Leibniz, Whitehead, Schopenhauer, Aurobindo, Schelling, and Radhakrishnan are just a few of the major theorist who have explicitly recognized that the interiority of individual holons is in essence the same as *consciousness*. This is likewise the ground of Henri Bergson's *Creative Evolution* (BERGSON, 1998). See also Arthur Young's (1976) critic of Ernst Cassires's analysis of whether the indeterminism established by quantum physics has a bearing on the question of free will and related questions of ethics.

as intentionality originates in the first-person spaces of free will but can be viewed from a third-person (exterior) stance of determinism. Incidentally, this is precisely why an epistemological attitude which intends to explain interiority (e.g. disposition to cooperate) only in terms of exterior surfaces and empirical forms (e.g. neurophysiology, evolutionary psychology, reciprocity) sticks with determinism and cannot come across meaning and values whatsoever. Thus, when Wilber refers to agency in the exterior or Right-Hand quadrants, it is the exterior *correlates* of interior intentionality that is meant.

Hence, when it comes to the evolutionary path social networks (Lower-Right), it includes both: (1) the behavioral intersections of the members of the network and (2) the exterior artifacts that are material components of the network (from techno-economic modes of production, architectural buildings, modes of transportation, types of media and modes of communication, foodstuffs, types of weapons, types of money, forms of business exchange, types of medical tool, and so on…). Whereas the behavior of the individuals has some sort of intentionality, artifacts themselves do not (artifacts, *as artifacts*, have no interiors, although their subholons do). So, the "agency" or defining pattern of an artifact is imposed on the artifact by the intelligence of the sentient being that produced it. Likewise, the behavior of an artifact is supplied by the network of sentient beings that are utilizing the artifact and it follows the various collective patterns or regnant nexus of the system. Thus, those artifacts, as artifacts, do not show stages of development. However, the consciousness that produced those artifacts does show growth, development and evolution, following the waves-like stages or patterns as discussed in the previous sections. Therefore, in a special sense, we can speak of growth and development in technology, agriculture, medicine, transportation, and so on, based on the degree of growth in the intentionality, cognition, or consciousness producing those artifacts. *These* are the associations between the Left-Hand and the Lower-Right quadrant implied in Figure 5.1.[48]

[48] Part of the entries in the Lower-Right quadrant is due to Lenski (1970), who has laid out the forms of social evolution in a way that is now uncontested by most scholars: foraging, horticultural, maritime, agrarian, industrial and informational. Wilber mentions that systems theorists (and structural-functionalists such as Parsons, Merton, Luhmann, Alexander, and Bellah) have shed an enormous light on social action systems, their maintenance and self-reproduction (WILBER, 2000b).

Still, the nexus-agency of the social systems does not only ensue from the interior space of culture and intentionality but also feedbacks an active influence upon the development of the individuals' mindset and collective worldviews. The identification of these profound influences is of course the main merit of the Marxist School. Accordingly, while in obvious disagreement with the Marxist attempt to reduce everything to the Lower-Right, Wilber (excerpt D, p. 129) acknowledges that *the mode of techno-economic production is the single strongest determinant for the average level of consciousness in a society*. Note that this feedback mechanism not only helps to explain the relative convergence of individuals' mindsets but also supports both the associations between the two lower quadrants and the onto/phylo parallel. Therefore, if the mode of production is foraging, the average level tends toward magical (purple); if the mode of production is agrarian, the average level tends toward mythical (blue); if industrial, rational (orange); if informational, pluralistic (green) (see Figure 3.12). [49]

Again, precisely because the social mode does not determine the consciousness, but only feedbacks on it, individuals in those societies can be higher or lower than the average. Hence, as Wilber explains further, a particular cognitive level produced the corresponding material artifact (e.g. preoperational invented foraging modes, concrete operational invented horticultural modes, formal operational created industrial, postformal created pluralistic). Once an individual consciousness (Upper-Left) has *conceived* a particular invention (e.g. the steam engine), and *behaviorally* communicates it (Upper-Right) to others such that they *understand* it (Lower-Left), then that group of individuals might eventually produce that artifact on a large scale and make it part of the social institutions of that society (Lower-Right). At that point, Wilber continues, the

[49] We are in this paragraph implicitly referring to *isomorphic* developmental lines or streams as they appear in the interior individual and collective domains. Accordingly, this isomorphism reappears as parallels between the cultural and the social domains. Still, Wilber call attention to the fact that various subsystems in a society are composed not only of the correlative or isomorphic individual streams (e.g. cognitive and education, psychosexual and marriage, defense and military); but also in part of occasions that have *no parallel* in the individuals. These occasions are thus *paramorphic*. Paramorphic streams arise because many aspects of the various the subsystems in a society are not merely reworking of individual intersections but are themselves novel emergents.

material artifact themselves, which were created and produced by a particular level of consciousness, in turn tend to inculcate the same level of consciousness that produced the artifact. For example, a person using a computer for much of his communication with others has his consciousness subtly molded by that mode of communication, whether or not he himself could invent and produce a computer. Likewise, on a bigger scale, when a entire society meets its subsistence needs using horticulture, every single member in that society has his or her consciousness molded, from birth, by the relentless ever-present realities imposed by that pervasive horticultural mode (whether they like it or not, created it or not, want it or not).

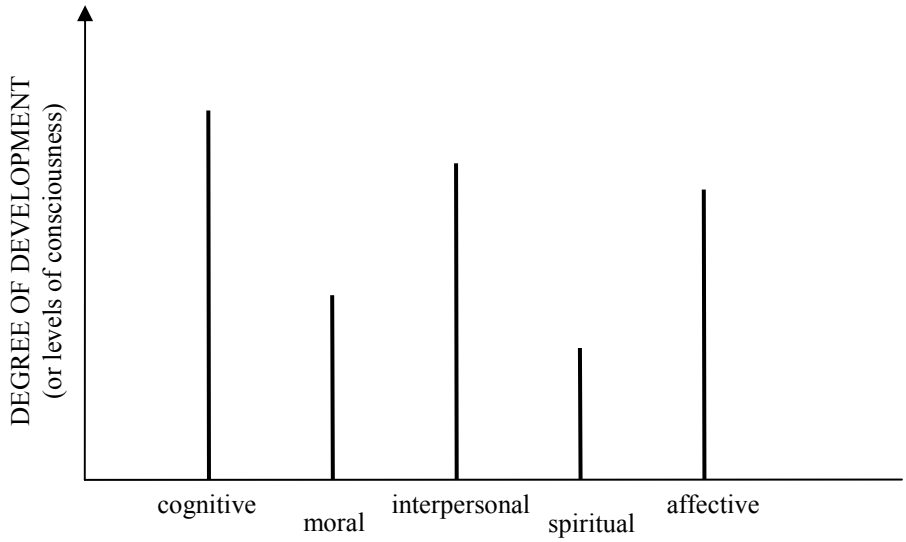
Accordingly, there is indeed something of a nexus-agency (regnant nexus or governing pattern) in the collective domain, but this is not because it subsumes its members but because it subsumes it own past (for evolution has directionality) in the present intersections of its members. There is no dominant monad (no super-I or social superagency) required to do this (in either cultural networks or social systems); but neither can this nexus be reduced to nothing but the interactions of isolated individuals. It is the idea of such irreducibility and integrality that is meant by the conception of a single Human Holon, which its manifold perspectives manifesting simultaneously all through the levels of its depth, in the four quadrants. Besides, such multiplicity is entirely subsumed, as potentials waiting for awake, into each and every one of us – the fractals of the Human Holon.

To resume, we must extract from this almost Integral vision of the Human Holon the implications on people ability to straighten out their collective action dilemmas using communication. Before we can do so, we must go a bit farther to consider the existence of multiple *lines* or *streams* of interior development, so that we take a better idea of limitations involved into our attempt to take the ECLET conception as *proxy* for the overall development of the Human Holon.

5.4. The many streams of a same river: limitations and possibilities of a generalization based on the ECLET conception

In indicated, the interior development of the individual fractals (Upper-Left) of the Human Holon presents multiple developmental lines or streams. Moreover, these about two dozen facets of subjective file present their own laws of evolution, following their own pace in relatively independent dynamics. Among the more preeminent developmental lines for which there are empirical evidences, Wilber mentions: morals, affect, psychosexuality, cognition, ideas of good, role taking, socio-emotional capacity, creativity, altruism, several lines that can be called "spiritual" (care, openness, concern, religious faith, meditative stages), joy, communicative competence, modes of space and time, death-seizure, needs, worldviews, logico-mathematical competence, kinesthetic skills, gender identity, and empathy. Just from this range one can grasp the senselessness involved in the attempt of getting to some all-purpose assessment of human interiority. Notwithstanding, as just mentioned, these lines or streams present their own tempo and logic of transformation, being influenced by diverse factors, so that a single individual can be very advanced in some lines, medium in others, and low in still others – all at the same time. This idea is illustrated in Figure 5.2. Note that, in this example, spirituality is presented as a particular developmental line. In this sense, as we saw, it would involve expressions of care, openness, concern for others, religious faith, and meditative stages. This view comprises possibly the most common understanding of spirituality, and is compatible with that that Wilber calls *shadow spirituality*, at least as far as this *stream* is just middling developed. *Deep spirituality*, on the other hand, pertains to the higher stages of development of the several developmental lines, and it is moreover constrained by the development of cognition, as well as the engagement into required behavioral injunctions designed just to foster the spiritual awakening (for a compressive discussion on the variety of conceptions of spirituality and its relation with cognitive development see Wilber 2000b).

Note that degree of development (or the level of consciousness), shown in the vertical axis of Figure 5.2, is an abstract measures of concrete realities that are unfolding in the lines themselves. There are all sorts of ways to measure and indicate those levels, and there is no "correct" number of levels or stages in any of the lines. In other words, the phenomena measured in the various lines are not the same thing, as they are disclosed and conceptualized on the basis of different, whereas appropriate, in opposition to disparate methodologies. The important point is that the very *structure*s of the stages in the various streams were proposed using actual phenomena of the stream, so that the structures of the levels of one line simply do not fit with the structures of the levels of other lines. Accordingly, the very terms of the stages themselves (and their proposed deep structures) adequately fit only the particular line or stream of with they are measured. This is why the common terms that were developed for the morals stream (e.g. Kohlberg's "punishment-obedience", "good boy-nice girl", "social contract") do not exactly match those of the cognitive stream (e.g. Piaget's "preoperational", "concrete operational", "formal operational") which do not match those of the self stream (Loevinger's "symbiotic", "protective", "conscientious", "individualistic") which do not match those of the Gravesian/SD psychosocial value systems ("egocentric", "absolutistic", "sociocentric", etc.).

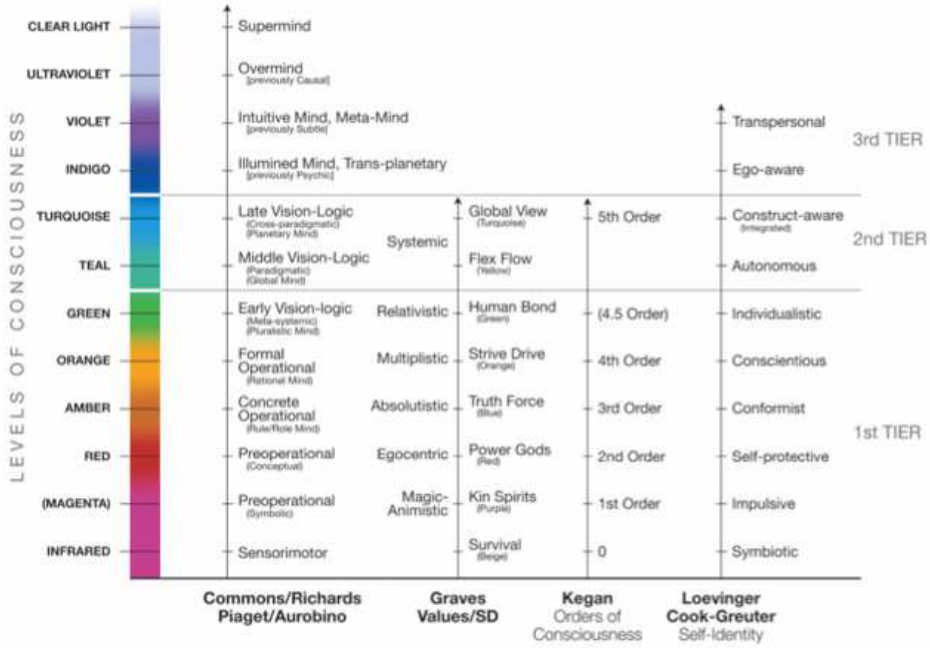


Source: Adapted from Wilber (2000, p. 30).

Figure 5.2 – The integral psychograph.

Therefore, as suggested, there is nothing of disparate in such dissimilar ways of approaching human interiority, whereas it certainly reveals the incredibly complexity and subtleness of the phenomenon itself. Furthermore, since these lines are "relatively independent" the *overall development*, that is, the sum total of all these different lines comprising the sensation of a "ghost in the machine" shows no linear or sequential development whatsoever. Yet, as Wilber explains, the bulk of research has continued to find that each developmental line itself tends to unfold in a stage-like, sequential, and holarchical fashion. In this regard, as we saw, Graves's model does not impart any exceptional conception of development. Wilber adds that there is a further general consensus that no matter how different the developmental lines might be, not only do most of them unfold holarchically, they do so through *the same set of general waves*, comprising, in this sequence, a physical/sensorimotor/preconventional stage, a concrete actions/conventional rules stage, and a more abstract, formal, postconventional stage.

In order to consubstantiate the empirical ground of the above statements Wilber reports to the widely regarded text *Higher Stages of Human Development* (edited by Charles Alexander and Ellen Langer)[50], where the works of thirteen top developmental psychologists – including Piaget, Kohlberg, Carol Gillian, Kurt Fisher, Howard Gardner, Karl Pribram, and Robert Kegan – are presented. Wilber reports that, of those thirteen models, all of them except one or two present conceptualizations that are at least in part hierarchical, including Gilligan's for female development. Also, in the same publication, after surveying the developmental research and data from Fisher, Sternberg, Kohlberg, Armon, Pascual-Leone, Powell, Labouvie-Vief, Arlin, Sinnot, Basseches, Koplowitz, and Alexander, Richard and Commons concluded that "The stage sequence [of all of these theorists] can be aligned across a common *developmental space*. The harmony of alignment shown suggests a possible reconciliation of [these] theories…" (in Alexander and Langer, 1990, p. 160, emphasis in the original). The suggestion of such possible harmonization is portrayed in Figure 5.3.



Source: Integral Institute (didactic material offered at the Integral Sustainability Seminar, Denver, September 2006).

Figure 5.3 – Corresponding levels in four developmental lines.

[50] Oxford University Press, USA (June 1, 1990).

Again, in order to adequately read the suggested associations in Figure 5.3 it is important to bear in mind the abstract nature of the "levels of consciousness scale" (color coded) in the vertical axis. Therefore, the levels of consciousness are always levels of something that is appearing or manifesting: a level of moral responses, a level of needs, a level of values. As Wilber puts it, "Consciousness is not a phenomena but the *space in which the phenomena appear*, and therefore 'levels of consciousness' simply means *levels of the phenomena appearing*. And when the *specific* stages or levels of those streams are specified, they are *always* in the terms of that actual stream". This is why "levels of consciousness", as used in the Integral Theory, can avoid that sort of *stream absolutism* that got Piaget in trouble and that Wilber refers to as a kind of occupational hazard of pioneering structuralism. He explains that many of the great pioneering got caught in this unfortunate confusion when they built an entire psychological system around a plausible stages sequence taken from a particular methodology or paradigm they have used to spot a particular type of human response to environmental challenges (needs, values, identities, morals, etc.). The problem of doing so, as Wilber puts it, is that the total system turns to be based on two shaky foundations: one, since no stage-conception can capture everything, the postulated system is necessarily based on partial capacities of the overall the human bodymind, and, two, when extrapolating from those particularities to the entire psyche, other capacities that might be captured by different stage-perspectives are excluded, so that a second partiality is created on top of the first.

More specifically, Wilber (forthcoming, excerpt D) stresses out that the important thing we particularly cannot do is use the way the "levels" are formulated in one line to refer to the "levels" in the other lines. For example, one of the *levels* in the (Piagetian) cognitive *line* is formal operational cognition or formal rationality (see Figures 5.3). This formal rationality can be used to adopt or support different levels in the values or moral lines. Therefore, an adult who is at the egocentric/power level in the values line, for instance, can be (and normally is) at the formal operational level in the cognitive line, and if one ask

him why his egocentric values are important, one will often get a very thoughtful and rational explanation of why the world is a very dangerous and raw place. Likewise, people at the same cognitive level of formal rationality can in fact have a preconventional moral sense (e.g. Nazi doctor), a conventional moral sense (e.g. fundamentalist preacher), or a postconventional moral sense (e.g. classical liberal). He then summarizes the general situation with a metaphor, saying that "structuralists are like individuals sitting the bank of a great river. They attempt to look for any patterns in those waves and streams – any recurrent or stable eddies in the river – and they formulate those patterns. But all that really means is that if you stand exactly where that structuralist stood and look at the river from that angle, you will see those same waves and streams. Yet, if you move ten meters down the river and look again you will see a different pattern emerging".

To remind, the angle from which Graves pictured the rivers flow, by drawing out the mindsets of his subjects using the vehicle of the conceptions of mature/health personalities, is regarded as providing an image of the Human Holon's interiority from the angle of the psychosocial value systems. As such it takes its rightful place sideways with a number of other perspectives, as in Figure 5.3. Therefore, my intent of using the ECLET conception as a *proxy* for consciousness development in no way should be taken as a suggestion that Graves's theory is truly general, as if encompassing everything worth of noticing in the domain of consciousness research. Much on the contrary, as already put, Graves's model was chosen owing to the convenience imparted by its very particular perspective, connecting agency and communion capacities of the self, to the inquiry on collective action dilemmas. Was I concerned only with the implications of the cyclical unfurling of those capacities, I would be done. However, the price to pay for this key fitness is that the same cyclical turn challenges my further aim of examining the implication of the more general property of the psychological development, i.e., the *lessening of egocentrism* and *increasing moral embrace* underneath the central thesis of this study.

Therefore, I am supposed to expand a bit more on the general character of the interior development in order to back up the aforesaid property with a

rigorously logic foundation, as it ensues right from the most basic tenets of Integral Theory, i.e., the basic unite of reality is better gasped through the concept of holon and how holons evolve holarchically.

5.5. Decentering of the self, moral skepticism, and the farther reaches of human potential

To begin, I want first remind the reader that the property of *lessening of egocentrism* and *increasing moral embrace* is quite present in the ECLET conception. To be sure, it is this property that explains the remarkable change of the ethical perspective as the evolving individual gets to the second spiral of existence. Then, at the GT level of existence, as we saw, the expression of what self desires is for the first time conceived so that it should not turn up at the expense of others. Of course that, as the overall psychology involves multiple and relatively independent developmental lines – and I am *not* putting forward a reduction of human interiority to the ECLET conception – one should not expect that this general trait, as measured in the ECLET scale, ought to show up totally apart from of the circumstances. We shall discuss this issue in connection with our experimental settings in the next chapter. At this time, as indicated, the point is to underpin the assertion relating to the broad-spectrum character of the property of lessening egocentrism and increasing moral embrace, which is what explains why it is present not only in the ECLET conception but is in fact a general trait, somehow underling the structure of virtually all developmental models (for a wide-range table of correspondence comprising over one hundred models of interior development showing the said property, see Wilber 2000b).

Accordingly, the fundamental reason why lessening of egocentrism and increasing moral embrace go along with the development of the self is just that the said development is the interior aspect (Left-Hand) of the tetra-enacted evolution of every fractal of the Human Holon (i.e. each one of us). As such, as we saw, it is a process of increasing envelopment, were earlier holons are transcended-and-included by the emergent holons, submitting their aims and

means to new internality codes. Note that, because such tenet of the Integral Theory is so fundamental is that Graves was able to find a way to conceptualize his otherwise baffling data only when he has acknowledged the principle of envelopment (see section the nested aspect of the ECLET conception, in the previous chapter). However, this is not yet an explanation as why the transcend-and-inclusion principle leads to lessening of egocentrism and increasing moral embrace – especially not if one is committed to the moral skepticism under the exclusionary instrumental view of rationality. From this perspective, the issue turns out as awfully puzzling, as the analyst disconnects the interior from the exterior aspect of reality and presumes that any valid explanation ought to be grounded on relations observed at the exterior (sensory-empirical) side. As we saw, this is the defining quality of that that Wilber calls subtle reductionism.

As Wilber explains, "this subtle reductionism can still champion all the twenty tenets [of Integral Theory], the notion of development, the emergent characteristic of evolution, the mutually interrelated nature of all holons, the holarchical nature of all holons, and so on. *All of that can still be heartily embraced by the subtle reductionist*; it's just that all holarchies of quality have been lost entirely in holarchies of quantity, and all gradations of interior depth have been replaced by gradations of meaningless exteriors". Therefore, as he concludes, "compassion is better than murder, but quarks are not better than photons, and thus to explain the entire Kosmos in terms of exterior surfaces and empirical forms is to find a cosmos with no value whatsoever: guaranteed" (WILBER, 2000a, p. 146). Perhaps even subtler is that fact to do so is ordinarily regarded as the most valued quality of the scientific endeavor, namely: to be "value-free". Again, this is the root of the moral skepticism, which as supposed to challenge directly even for the only matter of sustaining that we are indeed before a common dilemma whatsoever. Incidentally, as this recognition is not universal, it is just the developmental psychology which provides the answer as why this is so.

Hence, the first step toward increasing moral embrace is the cognitive capacity to put oneself on the other's shoes. This capacity ensues from a most

fundamental process that Piaget called *decentering* of the self, which begins when the child is for the first time able to stand aside from the egocentrism of early stages and instead take the role of other. As Wilber explains (2000a, p. 239), this process comes to a fruition with a *further decentering*, a further lessening of egocentrism, in the Piagetian formal operational stage (see Fig. 4.3) where one can take the perspective, not just of others in one's group, but of others in other groups, that is, not just an ethnocentric perspective but a worldcentric one. We already indicated how the up-to-date research of interior development, despite of important differences among researchers standpoints, are gradually coming to a consensus as to the potential harmonization of the various perspectives. As for the cognitive line itself, most researchers have found that cognitive development mover through three of four major stages (with numerous substages), namely: sensorimotor, concrete, formal, and postformal (WILBER, 2000b, p. 24). According to Susanne Cook-Greuter's research, reported by Wilber (2000b, p. 26), preoperational thinking has only first-person perspective (egocentric); concrete operational adds second-person perspectives (sociocentric); and formal operational goes further and adds third-person perspectives (which allow not only scientific precision but also impartial, postconventional, worldcentric judgments of fairness and care). Indeed, as Wilber explains, rationality (or reason in the broad sense) involves first and foremost such capacity to take perspectives. According to him this is why Jean Gebser calls it "perspectival-reason".

Therefore, as visible, it is quite significant that the "value-free" value of the positivist perspective is itself an expression of the decentering cascade. Furthermore, our epistemological puzzle ensues as well from further decentering at the postformal stages of cognition, while being a less the fortunate expression of it, largely under the onslaught of anti-Western cultural studies with a strong relativist prejudice, which, amalgamated with subtle reductionism, buttress the moral skepticism that ultimately level all qualitative distinctions.

However, while cognition is necessary it is not sufficient to bring about compassion instead of murder, as the above remark on the Nazi doctor illustrates

quite convincingly. But we have already discussed the issue as to the relative independence with which the many streams of subjective river flow. The point of note now it that such many streams are nevertheless streams of a same river, flowing in the space of consciousness, where the self is at once the very river and the navigator (plus the relation among them). Note that this self is not a "ghost in the machine", for it is just the interior face of the tetra-enacted "ghost-machine". Much on the confusion, as I see it, comes from the fact that the "machine" is seemingly quite concluded before the "ghost" come into being. This is an illusion owing to an ontological point of departure, though, for otherwise we are clearly before an "egg-chicken" kind of problem – at least as far as we attribute some degree of consciousness to our parents. But even if we cling to this point of departure, the *part* of the "machine" (i.e. the biological body) which is preset is only its most superficial casing, for the development of the psychological being is tied to the gradual activation of specialized neurological coping systems and involves, indeed, an integral transformation of the biopsychosocial system, as we in the previous chapter. Or, simpler put, the development is tetra-enacted.

Hence, in order to in order to better comprehend the developmental path of the "whole" self, as a process of increasing envelopment, comprising many different self-related streams, Wilber (2000b, p. 33) offers a useful distinction between the *proximate self* and the *distal self*. As he explains, the proximate self is some sort of observing self, an inner subject or watcher. It is said *proximate* since it is closer to that you call "you", so that it is *experienced* as an "I", or the "ghost", if you will. The *distal* self is some sort of observed self, some objective things that you can see or know about yourself, that is, how you appear in the Right-Hand quadrants – I weight so many pounds, have blond hair, etc. (Upper-Right); I am a father, mother, doctor, clerk (Lower-Right). Accordingly, the distal self is experienced as "me" (or even "mine"). The importance of such distinction ensues from the fact that during psychological development, *the "I" of one stage becomes a "me" at the next*". That is, what you are identified with (or embedded in) at one stage of development (and that you therefore experience very intimately as an "I") tends to become transcended, or disidentified with, or

de-embedded at the next. So, you can see it more objectively, with some distance and detachment. Wilber says that such understanding is shared by many researchers, East and West, from Sri Ramana Maharshi to Robert Kegan (op. cited, p. 34).

As example, Wilber reminds us that a young infant is identified solely with its body. The body is the infants' proximate self and it cannot really stand back and objectively observe its body. As Wilber puts it, "It simply *is* a bodyself, and as a body it looks at the world". But when the infant's verbal and conceptual mind begins to emerge, the infant will start to identify with the mind. Now, the mind becomes the proximate self or subject and the infant can then, for the first time, start to see its body as a distal object (or "me"). In other words, the subject of one stage becomes an object of the next. In this path, the proximate self is the *navigator* of the great and turbulent "river" of identification-disidentification, transcendence and inclusion. It is the central source of identity, and that identity *expands and deepens* as the self navigates from egocentric to sociocentric to worldcentric to the farther reaches of human potential.

We already remarked that such *overall development* do not exhibit linear or sequential development at all, owing to the relatively independent dynamics of the many different developmental lines. Notwithstanding, as far as the Human Holon configure an integral entity, the various developmental streams, which their line-specific substances or contents, nonetheless traverse the same general levels of awareness in the space of consciences (see vertical axis in Figure 5.3). The point is that no suchlike thing is able to open out at whatever level – or else those would not be levels of awareness. Accordingly, as also already indicated, the moral content or value judgments infused in such structure must at least cohere with the cognitive verge, so that one cannot be considerate before be cognitively able to put oneself the other's shoes. I regard that this would suffice to establish consideration for the collective good as a higher achievement compared with self-absorption, were such structuration of moral development not yet sufficiently established by empirical evidences (e.g. KOHLBERG, 1963, 1966, 1969, 1976a, 1976b). However, the important point about cognition as a

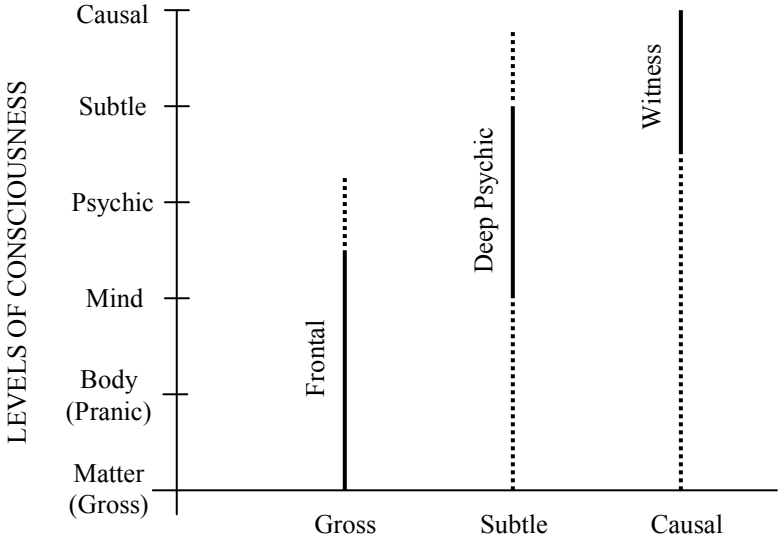
support for the other self-related streams is that, in order to fathom the possible content of the higher levels of the cognitive capacity (see Figure 5.3), cognition itself must the further factored according to the its overall object. In this vein, Wilber proceeds by distinguishing three *lines* of cognition based on the three natural *states* of consciousness – gross/walking, subtle/dreaming and deep sleep/causal states. By doing so, instead of having only one uniform line of cognitive development, with each stage stacked on top of its predecessors like many bricks (as in Figure 5.3), we would have these three relatively independent lines of cognition, each developing alongside the other. This view is illustrated in Figure 5.4.

Wilber explains that the main characteristic of the gross cognition is that *it takes as its objects the sensorimotor realm*. This line of cognition would begin with sensorimotor development itself, move into concrete operational, and then both peak and begin to trail off at formal operational cognition. Wilber suggests that the gross cognition tends to start trailing off at formal, and especially postformal, operations, because both of those increasingly take the world of thought as an object, and thus increasingly move into subtle cognition. Also, this type of cognitive development doesn't easily continue beyond postformal because in those higher stages the sensorimotor world, although not in any way abandoned, ceases to be the dominant object of awareness.

Accordingly, the main characteristic of subtle cognition is that *it takes as its objects the world of thought*, or the mental and subtle realms altogether. Wilber underlines that this line of subtle cognition involves precisely all those perceptions whose study has been downplayed by Western cognitive psychologists: first and foremost states of imagination, reverie, daydreams, creative visions, hypnogogic states, etheric states, visionary revelations, hypnotic states, transcendental illuminations and dozens of savikalpa samadhi (i.e. meditation with form). He adds that we would generally expect the subtle-cognitive stream to have available to it the same basic waves as most other streams: preconventional, conventional, postconventional and post-

postconventional, but the point is that it is a developmental line reaching all the way back to infancy, and not simply jumping out at the higher, adult stage.

Finally, the main characteristic of causal cognition is that *it is the root of attention* (and the capacity for Witnessing). As Wilber explains, this line, too, can be traced to early childhood, although it comes increasingly to the fore in the postformal stages. But the point is that by subdividing the cognitive line according to the three natural states of awareness we can preserve the characteristics of the cognitive phenomena that appears when the focus of attention is directed either to the sensorimotor realm, the realm of thought or the root of attention itself. This also helps to see that the postformal stages identified by Commons and Richards, although indeed entering the realm of thought in terms of mind's capacities to conceive systems within systems etc., still maintains the focus of attention upon the sensorimotor realm and, as such, they are more genuine higher stages in the Piagetian cognitive stream than those of Aurobindo, whose complete spectrum of consciousness mix the three suggested forms of cognition.



Source: Adapted from Wilber (2000b, p. 125).

Figure 5.4 – The development of the frontal (or ego), the deeper psychic (or soul), and the witness (or self).

The same subdivision can be applied to the self-sense stream, suggesting that this three great realms – gross, subtle, and causal – are home of three different lines of self, which Wilber generally call ego, soul and Self (or frontal, deeper psychic, and Witness), as in Figure 5.4. Accordingly, the ego (or frontal) is the self that adapts to the gross realm; the soul (or the deeper psychic) is the self that adapts to the subtle realm; and the Self (or Witness) is the self that adapts to the causal realm. Just as discussed *a propos* of cognition, these three modes of self can be viewed as relatively independent developmental lines. As such, once those selves emerge, they overlap; but great deal of research strongly confirms that they emerge in a generally hierarchical fashion, because the major stages of the self stream depends on the competences developed by the previous stages in other developmental lines.

Therefore, the frontal is the self that depends on the line of gross cognition (sensorimotor to preoperational to concrete operational to formal operational), and is for that reason responsible for orienting and integrating consciousness in the gross realm. The soul is the self that depends on the subtle line of cognition, and thus the soul is the self-stream that orients and integrate consciousness in the subtle domain. The Witness is the self that depends upon the causal line of cognition (the capacity for attention, detached witnessing, equanimity in face of gross and subtle fluctuations) and thus it is the self that orients and integrates consciousness in the causal domain.

As pointed out, these three selves develop relatively independently and all three of them usually are simultaneously present in various proportions throughout the overall development. Still, according to Wilber's explanation, the Witness or the Causal Self is responsible for the overall integration of all the other selves, waves and streams. And it is the Self supreme that prevents the three realms – gross, subtle and causal – from flying apart in the first place. Thus, as Wilber puts is, "even though the three domains can show relatively independent development, they are still held together, and drawn together, by the radiant Self, the purest Emptiness that can impartially reflect, and therefore embrace, the entire manifest domain". Accordingly, the *center of gravity* of

consciousness increasingly shifts from ego to soul to Self, so that the self developments more closely related with the subtle and causal realm can be adequately depicted as higher stage-waves in a simpler representation as in Figure 5.3.

The same reasoning is applied to suggest the three higher waves in the moral line. In *Integral Psychology* Wilber explains that his scheme includes not only principles of moral judgment (Kohlberg) and care (Gillian) – or how one reaches a moral decision – but also moral span, that is, the span of those deemed worth of being included in the decision the first place. As with most streams, the sequence runs from egocentric to ethnocentric to worldcentric to theocentric (or, more accurately, "pneumocentric", or spirit-centric, so as not to confuse the transpersonal realm with mythic theism). Wilber regards that nowhere is the expansion of consciousness more apparent than in the self's identity and morals. Indeed, the expanding identity is directly reflected in moral awareness, for, as Wilber says, "you will treat as yourself those *with whom you identify*".

As suggested, the proximate self navigates trough basic waves of existence by using its capacity to *identify* with each wave and ride it to some sort of completion. Note that *it is this capacity that gives to the self its sense of wholeness and the character of its agency*. Each time the self's center of gravity orbits around a new level of consciousness it has a new and different outlook on life. As Wilber puts it, "the self at each level sees a different world: it faces new fears, has different goals, suffers new problems. It has a new set of needs, new class of morals, a new sense of self" (WILBER, 2000, p. 38).

To put things simple, if you identify only with yourself, you will treat others narcissistically (preconventional stage in Figure 5.3). If you identify with your friends and family, you will treat them with care. If you identify with your nation, you will treat your countrymen as compatriots. If you identify with all human beings, you will strive to treat all people fairly and compassionately, regardless of race, sex, color, or creed. If your identity expands to embrace the Kosmos, you will treat all sentient beings with respect and kindness, for they are all perfect manifestations of the same radiant Self, which is your very own Self

as well. Wilber clarifies that this comes to you in "a direct realization of the Supreme Identity, precisely because identity can span the entire spectrum of consciousness, matter to body to mind to soul to spirit, with each expansion bringing a greater moral embrace, until the All itself is embraced with passionate equanimity" (WILBER, 2000b, p. 116).

As for the needs line, the great pioneer Abraham Maslow followed in Baldwin's footsteps, as Wilber says, and attempted to elucidate a full spectrum of motivations and needs. Thus, according to Wilber, Maslow definitely included the needs of the higher and further reaches of human nature (e.g. the needs of *self-actualization* and *self-transcendence*, which are, along with the other needs, what Maslow called "prepotent", which means they show a wave-like emergence). Wilber observes however that Maslow's research was not very specific in those higher waves, since fewer individuals reach them and thus studding them is extraordinarily difficult, although Maslow found abundant evidence of for their general existence. He also comments that Maslow's work fell into temporary disrepute during the eighties, when an extreme postmodernism, dominating both academia and the counterculture, made all forms of holarchy subservient to radical relativism (which Wilber refers to as a form of flatland dogmatism). Yet, Wilber regards that as the world awakens from that reductionism, Maslow's pioneering work are there to greet all who genuinely embrace a more integral and holarchical view. These comments are important because, as we saw. Graves's conception is closely related with that of Maslow.

5.6. On the occasion of omniscience and other arcane subjects

While our research is by no means reliant on the authenticity of the reports associated with the transpersonal realms or higher (see Figure 5.3), I would like to reemphasize the logical possibility of cognition unrestrained by our ordinary perception of time and space.

The starting point is the acknowledgement of proper position of the human sphere (noosphere) in the holarchy of being and knowing, as a holon

which transcends and includes the biosphere and the physiosphere (Figure 1.1). As Wilber (2000a) discusses at length, this is a transcendence-and-inclusion in terms of depth – not of span, of course. But such envelopment has as well a quite concrete implication, as both the biosphere and the physiosphere are *parts* of the Human Holon, and not otherwise. The first implication is that the existence of Human Holon presupposes the sustainability of its constituent parts, while the biosphere can keep on well enough without us – perhaps better than with us, judging by the damages caused by our unsettling deeds. The second implication is that the Human Holon subsumes the physiosphere and the biosphere into his or her internality codes. When one ingests a fruit, for instance, disassembles it altogether, absorbs just what is useful, and obligates the physical elements to obey one's internal codes, one is submitting the physiosphere to regnant nexus of one's biological vehicle. In turn, when Human Holon domesticates animals and sends them in millions to the frigorific, one is submitting the biosphere to the regnant nexus of human mind (noosphere).

What that means is, first, that all the evolutionary records of the emergence of biosphere are in many forms imprinted into our biological constitution, mostly notably in the genetic code. And, second, that we can in principle have conscious access to such records, since they are subsumed, as parts, that is, transcended-and-included in Human Holon's mind. Still this is absolutely contrary to ordinary experiences. The reason is shown in Figure 5.4, being that ordinary level of awareness is limited to gross cognition, which takes as its objects the sensorimotor realm, as we saw. In other words, gross cognition is focused on the exteriors. As we saw in the chapter two, this gross cognition corresponds to Bergson's conception of intellect, as relative as constituted according to the needs of action.

Notice that this is why Graves's subsistence levels of existence have no access the "other worlds" that he mention in his characterization of the HU system (global view or turquoise in Figure 5.3). Likewise, the subtle and causal streams of cognition correlated with Bergson's concept of *intuition*. Both are forms of direct apprehension of the internal order ingrained in human body,

mind, and causal spirit. In other words, these internal realities cannot be disclosed by the intellectual mind, which means that they cannot be disclosed by mapping their external correlates, even if such correlates do exist and may be mapped. As Bergson says, "if one gives in to [this sort of reduction], one will obtain purely and simply a physics of the mind traced upon that of bodies". Indeed, he says, "The intellect, so skilful in dealing with the inert, it is awkward the moment it touches the living. Whether it wants to treat the life of the body or the life of the mind, it proceeds with the rigor, the stiffness and the brutality of an instrument not designed for such use". In fact, Bergson concludes, "*The intellect is characterized by a natural inability to comprehend life*" (BERGSON, 1998, p. 165, italics in the original).

But we are currently still deeply immersed into an epistemological atmosphere with simply deny validity to whatever claim that be not grounded onto exterior observables. This is would not be a problem was science restricted to physics and chemistry. But the belief that empiricism ought to the method of human sciences is just what Wilber denounces as subtle reductionism, from which the most notable consequence is the moral skepticism. Please note that this epistemological atmosphere of subtle reductionism is as well the source on much criticism on structuralisms – after all, where is the matter of fact which establishes the alleged stages or levels of development? Of course there is not any primary matter of fact. The stages or levels are constructs within researcher's mind. They make sense in the measure that they reflect adequate intuitions of the interior reality.

Therefore, if we are able to put subtle reductionism off for a while, we can now take advantage of Habermas's hand to rescue the analysis of communication from the behaviorism of instrumental rationality.

5.7. Communication in the sacred mirror: integrating rationality

Nowhere the drawbacks of the commitment with the instrumental conception of rationality express itself clearer than when it comes to address

communication. On the one hand, the instrumental model of rationality leads standard noncooperative game theory to completely dismiss the role of communication just when it intuitively gives us the impression of being most crucial, namely, as a mechanism for coordinate actions in social dilemmas. On the other hand, since consistently reproduced experimental evidences confirm the aforementioned intuition on the service of communication in such settings, the commitment with the instrumental model of rationality drives the analyst interested in elucidating that usefulness right toward the scrutiny of the situational variables, including the structure of the social relations, that may explain the said effects as resulting either from strategic interactions or some presumed nonrational source, such as "culture" or "moral issues", for example.

In the first case, communication would ultimately finish up reducing the interdependent choice problem to a set of individual decision problems, by providing a mechanism for sharing information so as that the agents would supposedly be able to parameterize both the natural and social environments, and then handle individual decisions using the standard techniques of decision-making under risk. For this reduction to be carried out, as Heath explains, it must be the case that nothing fundamentally new arises in the context of social interaction that might call for the introduction of more sophisticated decision criteria. "However", he adds, "it is obvious that something fundamentally new *does* occur in social interaction. When there is another person in the environment, it becomes possible to select actions that are aimed at influencing that person's beliefs" (HEATH, 2001, p. 79, emphasis in the original).

Of course, there is nothing in the instrumental conception of rationality that prevents an agent from performing an action with the goal of "setting up" others to draw particular conclusions. However, as indicated, as far as one remains within the confines of instrumental model of rationality, such sort of attempt to influence other's beliefs is conceived, at best, as nonrational psychological or social mechanisms to exerting pressures of others. Indeed, we saw how stringently Hardin (1993) has outlawed any communicative appeal to deontic constraints as being essentially mischievous, and cause of "pathological

effects". The important point is that both arguments *may* provide adequate explanations *if* the individual is in fact acting instrumentally. As Heath (2001, p. 13) clarifies, "The first point to notice about Habermas's theory of communicative action is that it is a *typological* theory". Accordingly, Habermas does not reject instrumental conception of rationality and replace it with an alternative, "communicative" conception. Instead, he takes the point of departure the assumption that agents always have available to them a set of different, often incommensurable standards of choice. In this vein, communicative action will turn out to be action governed by a particular standard, while instrumental action will be action governed by another. Therefore Habermas is not committed to the claim that agents do not sometimes deliberate instrumentally; his claim is simply that instrumental models do not provide a sufficient basis for a *general* theory of rational action.

As suggested since the opening of this study, the different perspectives on rationality and the corresponding contrasting theoretical developments ensues from different points of departure. On the one hand, the approaches beached on the instrumental model *presupposes* the agents come to the action situation equipped with all sorts of intentional states, including beliefs, desires, values, and so on, whereas Habermas, on the other hand, puts forward that a general theory of social action should explain how this states come about – or at least how they acquire their content. I remind once more, in this regard, that to provide a theoretical foundation for the origin and structure of such intentional states, as a ground for his utilitarianism, was Mill's central concern in his *Logic* and *Principles of Political Economy*. However, as Heath (2001) observes, any action theory that starts out by presupposing agents' intentional states will be structurally incapable of providing such an account.

Habermas overcomes this deficit by incorporating *speech act theory* into his theory of communicative action, so that his theory of action includes a full-scale theory of linguistic action and the cognitive recourses that are simply presupposed by most action theories from an integral part of Habermas's own. As indicated, the key idea in Habermas's theory of communicative action is that

speech acts cannot be planed or executed with entirely strategic intent. To establish this claim, Habermas draws upon some rather controversial work in theory of language – in particular, J.L. Austin's distinction between illocutionary and perlocutionary components of speech acts. Heath then take upon himself the task of showing that Habermas central criticism of the instrumental view can be established without appealing to any of these problematic claims, but through an entirely game-theoretic analysis of the effects of communication systems on strategic interaction. Accordingly, rather than analyzing meaning, as did Habermas, in order to show that it cannot contain strategic elements, Heath analyses strategic reasoning (both in standard noncooperative and cooperative settings) in order to show that it cannot include meaningful elements, so that language could not arise out of strategic interaction (HEATH, 2001).

I previously indicated that the sort of problem Habermas was concerned with is just the very "puzzle" of collective action, that is, an action situation where the individual and the collective interests *may* collide. The conditional character of the puzzle, I suppose as clear at this point, is that it will depends on how narrow or wide is the range of beings that one's self is identified with. In other words it will depend on the level of consciousness development. In this regard, we also commented that Habermas backed his theory of collective action onto a reconstructive account both cultural evolution and individual development. In all these the key point is that, because Habermas does not approach the theory of action with a prior commitment to moral noncognitivism, as Heath (2001, p. 6) puts it, he is able to dispel much of the mystery that has surrounded the sociological conception of "norm-governed" action. Heath notices, however, that Habermas's perspective will only be persuasive if he is able to give some account of the moral reasoning that redeems the cognitivist intuitions. According to Heath's report, the attraction of Habermas's position, in this respect, is that instead of trying to provide an autonomous, "foundational" conception of moral reasoning that would ground the action theory, he imports elements of action theory into his philosophical ethics, in order to explicate the *practice* of moral argumentation. Therefore, at the same time that suspending the

noncognitivism postulate enables him to develop a more powerful theory social action, this enhanced action theory is itself deployed in order to relieve some of the skeptical pressures that have traditionally favored moral noncognitivism.

Heath locates the significance of Habermas's work on moral philosophy in the way he situates the problem, so as to expose the connections between the problem of moral skepticism (often put as the question "why be moral?") and the technical problem in theoretical sociology referred to as the "problem of order". The basic insight establishing such connection, which Habermas took from Durkheim, comes from the recognition that there is no *specific* socialization process through which agents acquire moral dispositions. In fact, *all* socialization is moral socialization, because all social interaction is governed by norms that function as deontic constraints. As Heath explains, "this means that acquiring the competences required to manage routine social interactions amounts to acquiring the *dispositions and personality structures* that we understand to be essential elements of moral agency" (HEATH, 2001, p. 8). Clearly, this provides the foundation for the mentioned onto/phylo parallel put forward by Habermas and embodied in the Figure 5.1.

As already indicated, Habermas's basic strategy in defending moral cognitivism (i.e. the rationality of moral claims) is to reject the correspondence theory of truth and the foundationalist model of justification, "both of which have served as major sources of moral skepticism", as Heath puts it. Instead, Habermas attempts to explain truth and justification in terms of practice of argumentation, thus location the inquiry on speech acts to its natural domain of hermeneutics, in the Lower-Left quadrant. This is what generates his "discourse" theory of truth and "dialogical" model of justification, which underlie the broader philosophical program that he refers to as "discourse ethics".

Heath points out that the theories of truth and justification that Habermas develops are both, in the broad sense, antirealist. That is, Habermas does not attempt to provide any foundation for moral judgment that is outside the logical space of reason. Note that this antirealist stance is equivalent to the enactment principle of Wilber's Integral Methodological Pluralism. As Heath suggests,

"The danger with such a position is that it can easily open the door to relativism", for "It becomes difficult to explain why we should ever expect to secure agreement on moral questions, it there is nothing to which they are "ultimately" anchored" (op. cited, p. 8-9). At this point, Habermas introduces a universalization principle governing moral discourse, plus some criteria for distinguishing between moral and ethical problems in order to demonstrate that moral argumentation will exhibit convergence. Although Heath endorses Habermas's general strategy in rejecting moral noncognitivism, and withdraws it from relativism by showing that Habermas's claim to "universality" follows fairly immediately from the pragmatic conception of language, he nonetheless argues that Habermas attempt to demonstrate that moral argumentation will exhibit convergence saddle him with an *unnecessary* burden of proof.

The "problem of convergence" (or the lack thereof) is important for us because, as posited by the Integral Theory, whether or not convergence will obtain is a matter of a quite concrete encounter of individuals' mindsets in the intersubjective we-space of mutual understanding, or, what amounts to the same thing, an encounter into the event-horizon of consciousness (Lower-Left, zone # 1). To understand this proposition one must first getaway from the commitment with an exclusionary instrumental conception of rationality and the corresponding *monological* model of justification, in order to acknowledge that justification *to others* is the primary phenomenon, so that justification is primarily *dialogical*. According to Habermas's view, speech acts, by drawing upon a set of shared linguistic resources, necessarily involve the agent in a set of public commitments, and therefore cannot be modeled instrumentally. In Habermas's terms, each speaker raises a set of "validity claims" through his or her utterances, claims that can only be redeemed through appeal to public justificatory resources. Thus, as Heath stresses out, public accountability is an intrinsic feature of linguistic communication, whereas under the instrumental conception of rationality, the agents' reasons for acting in a particular way are always, in at least an important sense, private.

Indeed, according to the instrumental view, actions are chosen as means for the attainment of agents' preferred outcomes. This implies that the reason for an action is always given through reference to some ranking that the agent has conferred upon the set of attainable outcomes. Therefore, under the standard interpretation of decision theory in terms of belief-desire psychology, as Heath explains, this ranking is understood to spring from a set of subjective motivational states. This way, the type of reason that such a ranking provides is necessarily agent-relative and nongeneralizable. In other words, it is a reason for some *specific* agent to perform a *particular* action. Also, it is private in the sense that the agent's is assumed to have privileged introspective access to information about her preference order, and, for those who lack this information about the agent, it will be impossible to identify any reason that she might have for doing anything.

At this point, it is worth noticing that the holarchical structure of the psychological development, as well as the universal aspect of the Human Holon, suggest that such impossibility which Heath points to is in fact only relative. Although there is no doubt regarding the agent's privileged access to her own reasons for acting, everyone can attest how much easier it is to infer the reasons under a child choices than under the choices of an adult – and this is more so if the child is our own, while the hypostatized adult is an stranger. This remark imports as an attempt to elucidate the concrete nature of the aforementioned encounter of consciousness in the intersubjective space.

Notwithstanding, Habermas/Heath's point is that the private nature attached to of instrumental rationality places limits on the extend to which an agent can be held accountable for her conduct by other agents, for the only constraint imposed by practical rationality upon an agent's conduct is that her actions accurately reflect her preferences over outcome. Heath observes that Habermas's emphasis on public accountability make his conception of "communicative action" closely related to the family of noninstrumental action conceptions owing to Kant, Weber, and Parsons, while his motivation for introducing a noninstrumental action type differs significantly. As he explains,

the primary concern of the "classical" sociological theorists was to solve what Parsons referred to as "the problem of order", as the instrumental conception of rationality, on this view, is incapable of explaining the empirical regularities exhibited in social interactions, because instrumental action does not provide the sort of "glue" needed to hold together stable human association. The problem with the classical sociological theory of action, according to Heath, is that its proponents have never be able to specify with any precision how values and norms interact with other intentional states – primarily beliefs and desires – in order to produce the observed interaction patterns.

As indicated, rather than introducing a type of public accountable action as a resource for solving the problem of order, Habermas uses it as a resource for explaining the *development of language*. Accordingly, the reason that public accountability is necessary for the development of language, in Habermas's view, is that the meaning of sentences is determined by the type of commitment that speakers undertake through their utterances. This way, the conceptual need for a noninstrumental action type is established without appeal to the problem of order. Yet, when language is used to coordinate social interactions, these linguistically incurred commitments (i.e. the "validity claims") reduce the range of action alternatives that are available to us, and so increases the "orderliness" of our interactions. Therefore, as Heath explains, the basic structure of the Habermas's theory is determined by the special constraints that an adequate theory of meaning and a theory of action must impose upon one another.[51]

Accordingly, the constraints imposed by public accountability vis-à-vis the uncertainty ensuing from private character of instrumental action serves, for Habermas, as the basis for the division of action in to two primitive analytic types: instrumental action and speech acts. As indicated, one type is not substitute for the other. Instead, Habermas argues that these two types of action

[51] Heath explains that it is widely accepted that a general theory of communication will have to parcel out the analysis of utterance meaning into semantic and pragmatic components. This is the origin of the mutual constrains between the theory of action and the theory of meaning. On the one hand, a general theory of meaning requires a theory of action to "handle" the pragmatic aspects of communication. The semantic theory, on the other hand, by inducing a particular pragmatics, thereby places restrictions on what will serve as acceptable theory of action. According to Habermas, it is considerations of this type that expose the inadequacy of the instrumental conception of rationality (HEATH, 2001, p. 19).

correspond to two different "stances" or attitudes that the agent can take toward others. On the one hand, instrumental action involves simple concern with success of action, whereas communication, on the other hand, requires that the agent adopt what Habermas calls *performative stances*, which commits the agent in series of idealizing presuppositions. So while the success of an instrumental action rests in the attainment of its objectives, the success of speech act rests in the achievement of *mutual understanding*, i.e., the acceptance of the commitments undertaken (HEATH, 2001, p. 19).

In order to grasp the insufficiency of the instrumental conception of rationality, so as to acknowledge the necessity of another *telos* for human interactions but success in maximizing no matter what generation of individual preference functions, it is required to take the origin of such preferences formation seriously. In this vein, one is supposed to recognize that the agent's beliefs and desires are *propositional attitudes*, which means that they take as their objects interpreted sentences of a natural language. Yet, under the standard "psychological" interpretation of instrumental rationality, inherited from Hobbes and Hume, how agent's come to acquire these attitudes – where the grasp of language come from – is usually ignored or posited as primitive.[52] In other words, the standard instrumental models of rationality start with a conception of agents in which the basic structures of theoretical rationality are already in place. The attempt is then made to generalize this apparatus from the theoretical to the practical realm. Heath comments that this leads to philosophical commitments with as deeply antipragmatist, since the basic explanation strategy precludes the use of practical (social) considerations to elucidate the structure of theoretical reason.

Given this, a more general explanatory strategy, put forward by Davidson (DAVIDSON, 1984, 1990, 1996) and followed by Habermas, should begins by treating both theoretical and practical rationality as equiprimordial, and

[52] Heath (2001, p. 20) comments that, according to Davidson – whose work Habermas takes as his point of departure in later formulations of his theory – the problem mentioned in the text does not go away if one opts for a "reveled preferences" interpretation of instrumental rationality, for the agent's capacity to understand to complex wagers of possible choices requires significant linguistic and computational abilities which still must simply be presupposed by the model.

take as psychological basis not only beliefs and desires and also meaning. Accordingly, *all these three propositional attitudes must be ascribed to agents in order to explain observed patterns of behavior*. As Heath explains, on this view, the utility-maximizing conception of rationality is a structure that we employ to resolve a set of observed preferences over actions into a set of ascribed beliefs and desires. Similarly, a *theory of truth* is a structure we can use to resolve a set of observations about which sentence an agent hold true into a set of ascribed beliefs and meaning. This analysis provides the foundation for the two primitive types of action (instrumental and communicative). Yet, Davidson acknowledged that all our propositional attitudes as interdependent, so that a single attitude is never directly responsible for a particular behavioral display. Therefore, when we observe an action the desires we ascribe to an agent as an explanation for that action will depend upon what beliefs we also ascribe to her. Therefore, Davidson is unsatisfied with this basic typological construction and suggests that it can be further unified, by merging the ascribed patterns of *preferring* and *holding true* into a single formulation of the kind *preferences over which sentences are true*. Still, Habermas do not follow Davidson's suggestion in this regard, and continue to view speech and instrumental action as fundamentally distinct types.

The important point of the distinction, as indicated, is that they impart two different *telos* to the agents' intentional states in their social interactions. This trait ensues from Davidson's view that, in order come up with an adequate interpretation of speech, it is not only necessary to focus on beliefs and meanings, but also at least temporarily to bracket the question of which desires the agent is acting upon. That is, in order to ascribe meaning to the agents' words on the basis of which sentences she held as true, he saw as necessary to grant that agents' pattern of affirmative and negative responses be accurate reflection of her underlying belief states. As Heath resumes it, this amounted to proposing something like a "desire to tell the truth" governing all speech. Similarly, in order to determine an agent's desires and beliefs, by observing preferences over actions, it would be necessary to presuppose as given the meaning that the agent attributed to various senses.

"Because a desire to tell the truth must be ascribed to agents in order to interpret speeches", Heath explains, "it follows immediately from Davidson's view that truth-telling must be constitutive of linguistic meaning". Accordingly,

> The action of speakers is 'intrinsically veridical', in the sense that it must be, on the whole, motivated by a desire to tell the truth. And since language, must be presupposed in order to ascribe desires to agents, the veridicality of speech is more basic than any particular desires that agents may possess. This makes it incorrect to say that agents tell the truth when speaking *in order to satisfy their desires*. The fact that they tell the truth when speaking is necessary condition for they have desires in the first place. Thus it is a priori that speech cannot be a form of instrumental action, that is, action oriented toward satisfaction of desire-based preferences over outcomes (HEATH, 2001, p. 22).

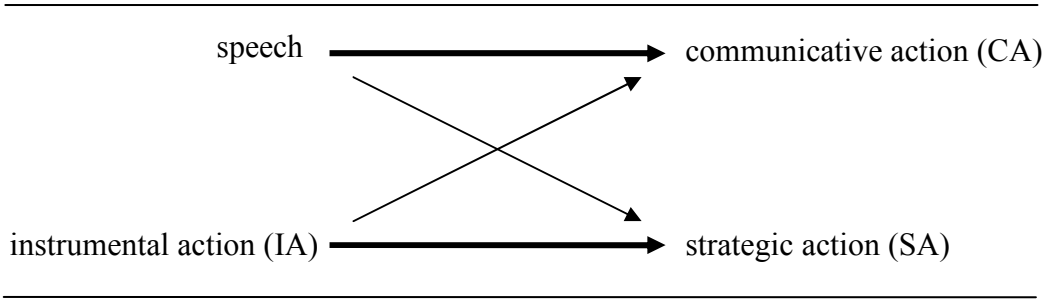
Heath obviously acknowledges that agents are capable of lying, just as they are capable of having false beliefs. The point is that lying and error are intelligible only against a background of veracity and veridicality. This is what motivates Habermas's claim that *mutual understanding* is the "inherent telos" of communication. Together with the outcome oriented instrumental action it comprises the two "elementary forms of action". Social action, in this view, is a complex phenomenon constructed out of the interactions of these two elementary forms of action (Figure 5.5). Therefore, according to this view every interaction involves the puzzle that game theory identifies as the problem of "the regress of anticipations", and that Parson referred to as "double contingency".

In order to see how the integration occurs, in Figure 5.5 resumes Habermas's typology of social action. The illustration makes clear that communicative action is not the same as speech. Accordingly, communicative action is still a form of teleological action, so that agents continue to pursue goals to bring about states of affairs in the world. As Heath (2001) explains, the distinctive characteristic is that agents *use language* is order to solve the problem of "double contingency", and therefore use language to determine a set of shared goals. This use of language as an explicit coordination mechanism (rather than simply consistently aligning beliefs through strategic reasoning) imposes constraints on the *type of goes* that agents can pursue, and the means that they can employ. Also, strategic action is not a simple generalization of the instrumental action, for it relies on linguistic resources as well. In this case,

however, the intralinguistic objectives of speech (i.e. to get to mutual understanding) are subordinated to each agent's individual projects. Language is still the background, but it does not exercise any constraint on the range of action alternative available to agents. This is, of course, the nonbinding character of communication, as set forward by the exclusionary instrumental conception of rational acts. As well, this is why, once caught under this partial view, the impact of communication on the outcomes of collective action dilemmas seems harder to explain.

Finally, it is important to notice that the reinforce that idea according to which it is the use of language that generates the structure of public accountability, and that Kant, Weber and Parsons took to be basic features of norm-governed action. The fundamental idea is just that when agents use language to coordinate expectations, they incur certain commitments. *This* is what their speech intelligible in the first place. However, as Habermas's puts it, these commitments are also relevant to the subsequent action sequence, so that they are *not* merely intralinguistic but impose direct constraints on agents' future conduct. These constraints are what generate the social order.

As indicated, the sort of case that Habermas has in mind goes right into the core or the commons dilemma. According to Heath report, the sort of case considered by Habermas's is when agents adopt a rule to govern their interactions – for example, when everyone agrees that a certain allocation problem will be handled on a "first-come, first-serve" basis. In the future, agents who were party to this agreement will be held accountable to the accepted norm. However, according to this view, it is not an intrinsic feature of the norm that agents are held accountable to it. Instead, it is because the norm us established using the mechanism of communication oriented toward mutual understanding that all agents are held accountable to it. Thus, as Heath puts it, the accountability of social action is not a primitive feature, but inherited from the accountability of speech acts. Communicative action is precisely the segment of social action that inherits this characteristic (HEATH, 2001, p. 25).



Source: Adapted from Heath (2001, p. 25).

Figure 5.5 – Elementary action types combine to produce social action types.

Accordingly, as the desire to tell the true must be assumed as governing all speech (or else language could not have emerged), the orientation toward mutual understanding is inherent to communicative action (from where something new may still emerge). And as strategic use of language are parasitic upon communicative uses so is breaking promises parasitic upon the practice of keeping promises. Heath notices that parasitic here has a precise meaning, so that "an action type is parasitic if it can only succeed in the event that ones' interaction partner *makes an error* in the determination of one's attitude or intend. If everyone always broke their promises, then the entire practice would collapse. It is precisely because people, for the most part, keep their word, that it is possible for some people to defect profitably. Similarly, if everyone always communicated strategically, then the institution would collapse […]. However, strategic uses of communication are possible if one player adopts and instrumental orientation while the second maintains an understanding-orientation". As indicated, using entirely game-theoretic analysis, Heath showed the impossibility of having successful communication emerging from strategic interaction alone.

5.8. The twofold dimension of mutual understanding: toward our proximate framework and empirical strategy

It was already indicated that Habermas offers a reconstructive account of the historical origins of the agent's capability for engaging in communicative action, grounded in the contributions of Herbert Mead and Emile Durkheim, and suggested that this historical process is recapitulated along development of the sociocognitive competences and moral reasoning of individual, is a specific report to Kohlberg's model. I then suggested that this onto/phylo parallel is basically what is presented in the Figure 5.1.

The evolutionary/developmental aspect of Habermas's theory is central to avoid the unwarranted inference that all social practices are grounded in rational consensus, as if the internalization of sanctions had played no role at all on the development of social institutions. As he explains (Habermas 1984, p. 39), "this conflicts with the repressive character evidenced in the fact that norms, demanding obedience, take effect in the form of social control". Habermas therefore argues that social institutions were *originally* stabilized through simple (i.e. nonrational) internalization of sanctions, but have slowly been transformed through the use of communicative action as a mechanism for social integration (HEATH, 2001, p. 34).

The importance of this developmental aspect is that it adds a "vertical" component to the problem of attaining mutual understanding, that is, to the "horizontal" component of similar signification most commonly known as *solidarity* (after Gadamer) and which configures the central notion in hermeneutics validity claim (Wilber, forthcoming, excerpt C). At this point, it is important to bear in mind the "antirealist" (i.e. anti subtle reductionist) character of such validity claims, as discussed above. In Wilber's terminology, the validity claims raised in the hermeneutic circle (see Figure 2.8), is not one of objective truth (Upper-Right), subjective truthfulness (Upper-Left), or functional fit (Lower-Right), but intersubjective justness, rightness, or appropriateness. Therefore, under the said qualification, horizontal solidarity refers to varieties of

possible mutual understandings as they unfold around similar depth or similar waves of consciousness. As Wilber (excerpt C/part4-1, p. 5) explains, this is related with the issue of *cultural legitimation*, or how a particular solidarity maintains the allegiance of its members. Vertical solidarity, on the other hand, means that the people involved (or whatever holon) *share a similar depth* or level of consciousness. Therefore, bearing in mind the multidirectional space defined by the AQAL matrix, genuine mutual understanding entails a quite concrete intersection of mindsets and worldviews within the phenomenological space of hermeneutics, at the Lower-Left quadrant. As Wilber puts it, "In order for intersubjectivities to be actual intersections, they must collide in a real phenomenological space: they cannot be near misses but must be direct hits or prehensions. This means that the intersections or shared signifieds that are internal to the "we" must be of the same general depth or else they will slide over and under each other and fail to mutually resonate" (Wilber, excerpt C/part4-2, p. 3).

Therefore, genuine mutual understanding demands the presence of both vertical solidarity (shared depth) and horizontal solidarity (shared horizons). We shall the discuss the implication of this on the research of collective action in the next section, in connection with the submission of our suggest framework to add the vertical depth to the otherwise flatland analysis of institutional change and mental model revision, as put forward by the IAD framework and the so-called "cognitive institutionalism".

5.9. Rescuing depth from flatland: toward and integral approach of the commons dilemma

Wilber often uses the term "flatland", borrowed from Abbott's 1884 novella (*Flatland: A Romance of Many Dimensions*), to refer to the leveling of all qualitative distinctions in the subjective domain (Left-Hand quadrants), ensuing from the current epistemological trend of acknowledging only quantitative distinctions and evidence of functional fit (Right-Hand quadrants) as

valid explanations for observed phenomena in whatever quadrant – in the root of moral skepticism. We have already discussed this issue in the chapter two, in connection with the evolutionary approach to the origins of human array of intentional states and psychological dispositions. There we also saw how such methodological discipline naturally drives the IAD framework to scrutinize the situational variables of complex socio-environmental systems, and to put forward a kind of behaviorism of cooperation, as its' valuable contribution for our understanding concerning how to foster the likelihood of self-organized solution in collective action dilemmas. Further, we saw that the significance of both individual intrinsic values and cultural traits are both plainly acknowledged as important factor affecting the final outcome of collective action. The limitation in this regard is that, by its very methodological decision, the IAD framework is not primarily focused on addressing such dimensions, so that they are treated as exogenous variables and handled using changeable assumptions as they appear consistent with each action arena (see Figure 2.6). Despite this, I argued that the methodological bent toward the game theoretical approach traps the IAD framework into the exclusionary instrumental view of rationality, which in turn makes impossible for it to hit the road beyond of the confines of strategic reasoning, in order to grasp the full significance of face-to-face communication (see Figures 2.8 and 2.9) by incorporating the concepts of communicative rationality (see Figure 5.5). I shall argue in this section that the same limitation applies to the so-called "cognitive institutional" and will suggest how it can be remediate by combining these approaches with the findings of stage developmental psychology, in general, and the ECLET conception, in particular, as they bring to the scene the significance of the vertical aspect of shared signification to support genuine mutual understanding.

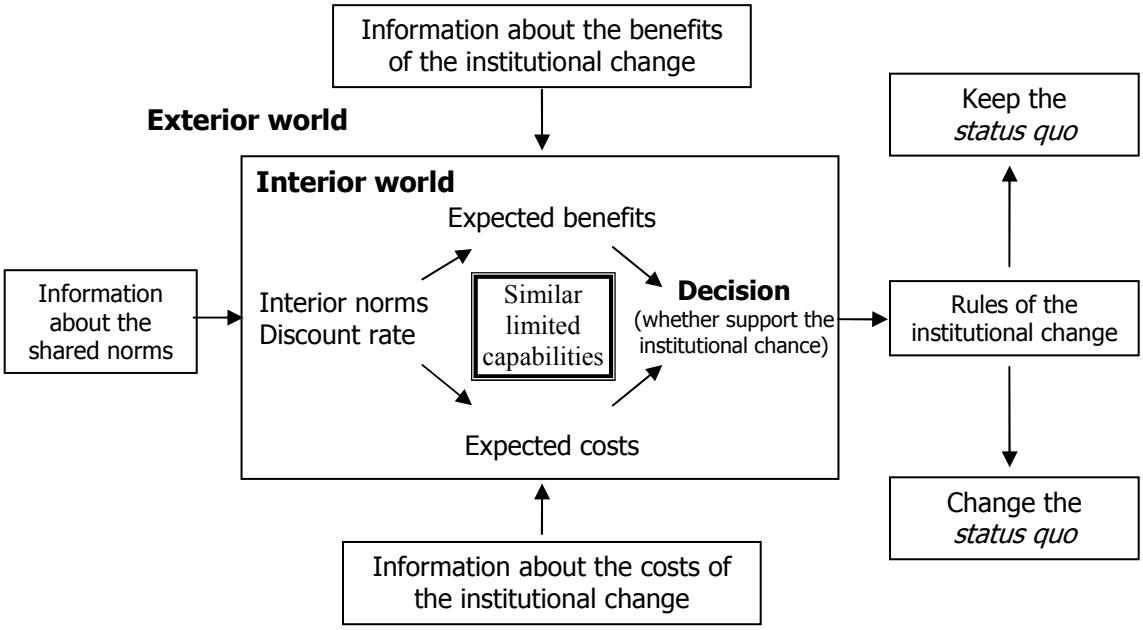
5.10. Buddhas, demons and ordinary people in the action arena

We saw that the methodological discipline at the bottom of the IAD framework substitutes the presumption that "some individuals are incompetent,

evil, or irrational, and others are omniscient", for the assumption that "individuals have very similar limited capabilities to reason and figure out the structure of complex environments" (OSTROM, 1999, p. 25-26). Since the IAD framework does not ignores the variety of intentional stated that may motivate individual behavior, as incorporated by the so-called "second generation" models of the human actor, the above methodological decision must not be tied to any particular representation of the preference function – such as in the traditional model of "rational egoists". Such methodological approach entails, nonetheless, a compromise with the instrumental conception of rationality, together with the methodological individualism which goes naturally and often unstated all through the institutional-economic approach. When applied to the choice situation involving institutional change, the IAD framework, consistently with the broad game-theoretical research program, reduces the interdependent choice dilemma to an individual choice problem, as in Figure 5.6.

However, we saw that what matters in order to figure out the structure of complex environments, especially with the aim of devising solutions for collective action dilemmas, is not only capability to reason, and particularly not to reason instrumentally (see Figure 5.5). Instead, it involves a number of different lines of interior development (e.g. as shown in Figures 5.2 and 5.3). Moreover, these multiple lines present their own and relatively independent internal codes of transformation.

Therefore, the laws of mind and of the formation of character that Stuart Mill was looking for, in order to ground his utilitarianism, should not be taken as mere adaptive responses to either the proximate circumstances of action or the evolutionary path of human species. As we saw, once we have acknowledge the tetra-enactment principle of the Integral Theory, we should no longer take the structure of those complex socio-ecological environment for granted, as a preset reality waiting to be correctly reflected into human cognition (as in the "reflection paradigm").



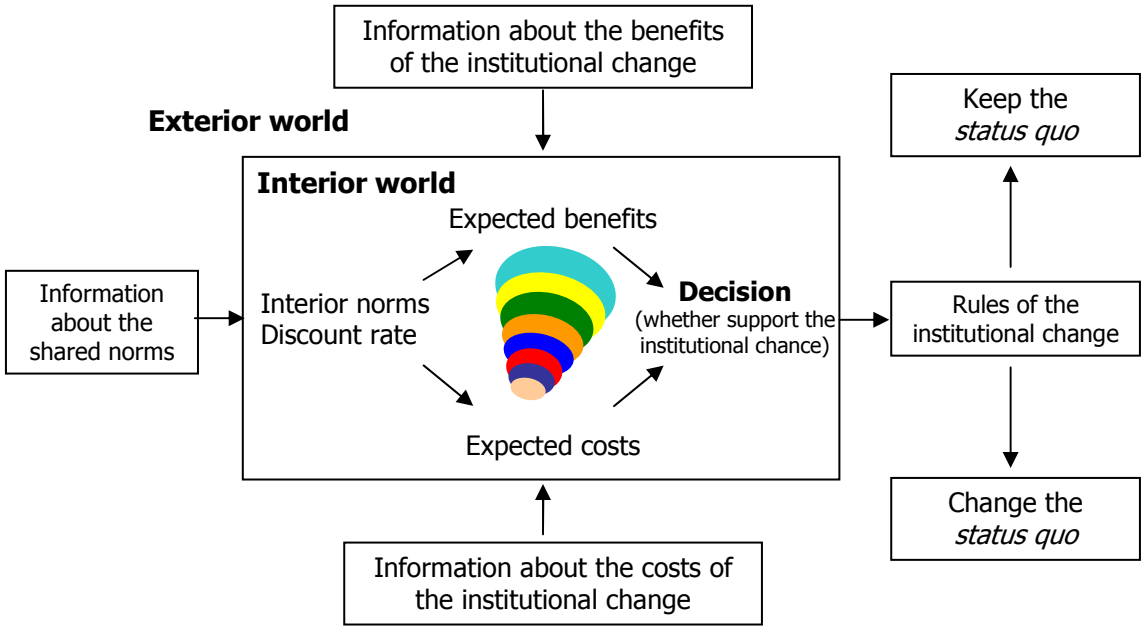
Source: Adapted and modified from Ostrom (1997, p. 193).

Figure 5.6 – Variables affecting institutional choice.

Further, once we have corrected the sight in relation to the proper location of the Human Holon (noosphere) in the holarchy of beings and knowing, as a tenet rigorously deduced from the fourfold aspect of every holon – one which avoids the individual-collective confusion observed in most systems theory and reproduced in the IAD framework – the Integral Theory opens the theoretical foundation to explain the character of enlightenment associated with the farther reaches of human potential.[53] Finally, since all through our discussion I have struggled to show the consistence between the holarchical unfolding of human awareness and the structuration of value systems and normativity which traditionally places Buddhas or Gods above and demons below, with all sort of ordinary levels of consciousness in between, it is time to relax the methodological assumption of the IAD framework and invite all those patterned levels of consciousness to join the action arena. Hence, in the measure that the

[53] We have attempted to show that this is connected with the envelopment of both the biosphere and physiosphere by the noosphere, on the one hand, and with the capacity to retrieve the internal codes of those domains in human consciousness through contemplative meditation, on the other hand.

ECLET conception can stand as a *symbol* of these awakening of human consciousness, our basic suggestion to explore its implication on individual choice regarding institutional change can be represented as in Figure 5.7.



Source: Author's configuration.

Figure 5.7 – The ECLET conception as weights and restrictions to expectations, norms, time horizon and choice concerning institutional change.

The significance of this apparently simple change in the original scheme of the IAD framework is that the ECLET conception adds considerable information and imposes restrictions on how the subjective benefit-cost analysis ought to be weighted according to the level of psychosocial development. Accordingly, "interior norms and discount rates" are longer just a matter of "information about shared norms" in the exterior world, but truly imputed and reflected – as far as the ECLET conception is here just a symbol of the overall process of consciousness awakening – in the overall psychograph (see Figures 5.2 and 5.3) of the human being considering the institutional chance. As well, the

entire psychograph, not only the instrumental type of rationality in the cognitive line of interiority, impacts the subjective evaluation on expected costs and benefits of the institutional change. Most important, however, is that the way how such array of information about the situational variables affect the individual choice cannot be predicted by the structure of action situation itself, as if the human actor could freely chose the best adaptive response to each situation, as presumed by the instrumental model of rationality. Instead, as we saw, the holarchical structuration the psychosocial development, as put forward by the ECLET scheme, establishes restrictions upon the dynamic of change. The implications of this aspect can be better appreciate together with the scheme suggest by the "cognitive institutionalism" (MANTZAVINOS et al., 2003).

According to the cognitive institutionalism, a deeper understanding of institutions' emergence, their working properties, and their effect on economic and political outcomes should *begin* from an analysis of cognitive processes. The proponents tie the processes of learning to institutional analysis, and explore the nature of individual and collective learning, stressing, as they say, "that the issue is not whether agents are perfectly or boundedly rational, but rather how human beings actually reason and choose, individually and in collective settings" (MANTZAVINOS et al., 2003).

The cognitive institutionalism (CI) draws in a view of mind as a complex structure that *actively* interprets and at the same time classifies the varied signals received by the senses. Such active interpretations form what CI generically defines as "mental models". The CI then acknowledges that "mental models gradually evolve during our cognitive development to organize our perceptions and keep track of our memories", thus consisting in flexible knowledge structures (MANTZAVINOS et al., 2003, p. 3). As visible, such understanding comes across the ECLET conception, while the CI remains within the boundaries of the instrumental conception of rationality and is not in itself a cognitive theory. Accordingly, the CI draws in a *pragmatic notion of mental models* so that a mental model is understood as "the *final prediction* that the mind makes or expectation that it has regarding the environment before getting feedback from

it" (ibid, p. 4, my italics). The proponents then put forward a behaviorist-like reasoning to posit the mechanism according to which the mental models are likely to be revised, refined, or rejected altogether. In this vein, "Learning is the complex modification of the mental models according to the feedback received from the environment" (ibid, p. 4). The general schema is illustrated in Figure 5.8.

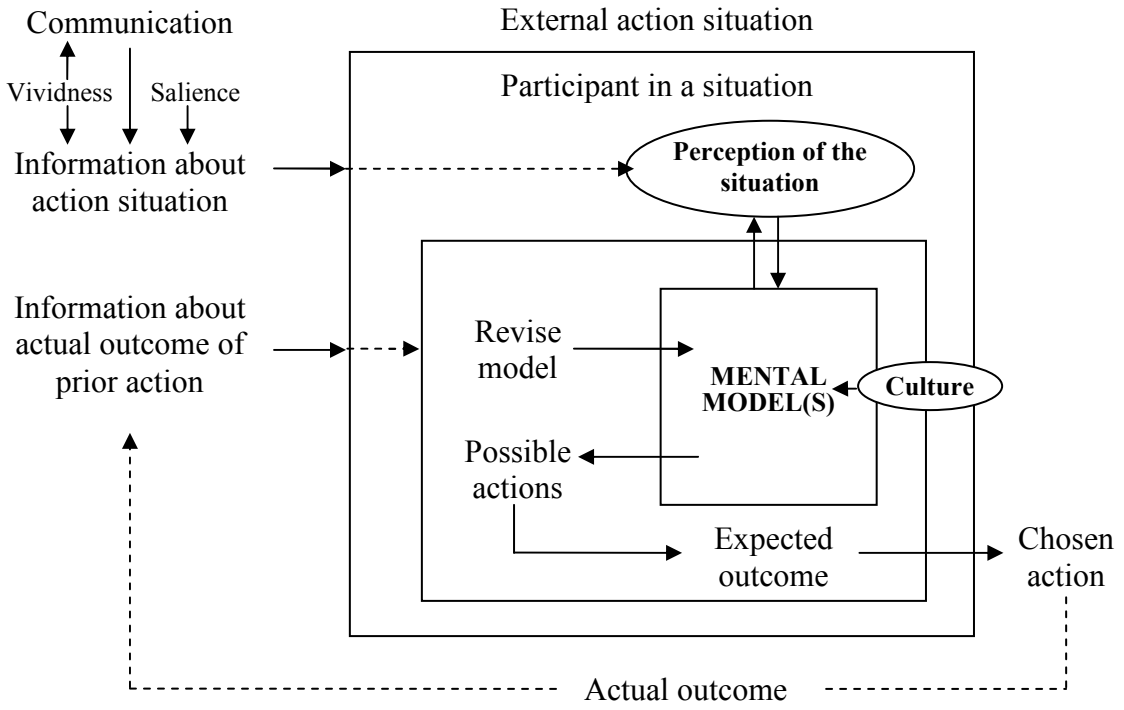
Since authors' perspective is committed with the instrumental model of rationality they basically passes over the enactment principle and assume a pregiven world which ought to be reflected into human mind as accurately as possible. Because of this, the authors seem puzzled by the persistence throughout history of dogmas, myths, superstitions, and ideologies based on "flawed belief systems" and suggest that research should pay as much attention to learning that produces such beliefs as we do to learning that appears to interpret correctly the problems confronting humans. "Because the mind *actively* interprets all sensory input", as they put it, "the message regarding the success or failure of the solution attempted will often be misinterpreted" (ibid, p. 5, italic added). As the authors explain:

> When environmental feedback confirms the same mental model many times, it becomes stabilized, in a way. We call this relatively crystallized mental model a "belief;" and we call the interconnection of beliefs (which can be either consistent or inconsistent) a "belief system". Having enabled an individual to survive in his or her environment in the past, the belief system becomes connected with the motivational system. To put it another way: the belief system is progressively shaped by the involvement of a parallel emotional adaptation, and it therefore takes on the character of a general filter for all new stimulus processing, so one may safely hypothesize that it is relatively resistant to abrupt changes structures (MANTZAVINOS et al., 2003, p. 4).

Note that this understanding is quite in line with Clare Graves's ECLET conception, particularly with his explanation of system closeness or arrestment, as well as with the challenges associated with the normal dynamic of development (see Figure 5.4). Moreover, we saw that the said emotional adaptation is not *parallel* but constitutive of the very cognition itself (e.g. Damasio 1994), together with other self-related lines which mix together in the sense of self-identity. Accordingly, the difficulty to revise mental models is tied

with process of identification and disidentification involving the proximate and distal selves, according to Wilber's explanation. "In fact", as Wilber (2001, p. 38) puts it, "the self-related stages are generated, in part, precisely from the self's *identifying with a particular level of consciousness*" (emphasis in the original).

The complexity of the problem, regarding the likelihood of self-organized solutions for collective action dilemmas, is that individuals in the action arena may come with divergent mental models of the action situation. Denzau and North (2000) handle this puzzle by pointing out the importance of communication as a way to enable convergence of the different views of the action situation through repeated exchange of information about likely action-outcome linkages and the feedback they obtain from a situation (Figure 5.8). This convergence would be particularly facilitated if the participants already share some broad cultural views of the world. Still, the commitment with the instrumental model of rationality reduces communication to a strategic interaction mediated by shared information regarding the action situation and possibly the expectations of other "players".



Source: Denzau and North (2000) modified by Ostrom (2005).

Figure 5.8 – The relationship between action-outcome linkages, information feedback, communication, and mental model revision.

In this regard, we saw how Ostrom (2003) efforts to translate the effects of face-to-face communication into a behaviorist triangle of indirect reciprocity connecting trust and reputation. Respecting the CI schema, she observes that individuals may not always be able to engage in open repeated communication, so that convergence of mental models might be difficult to occur in most common situations (OSTROM, 2005). In particular, she draws attentions to studies showing that unless there is open and active communication among participants over some time, they may simply use different mental models when interacting in the same external situation. Ostrom mentions, in this regard, John Schiemann (2002) research, who addressed this problem in an analysis of the mental models used by different age groups coming from the same culture but who have had vividly different experiences in life. Schiemann found that older Hungarians, who had directly experienced the bloodshed of World War II,

interpreted the situations following the ending of the Soviet Union differently than younger Hungarians who only read about the earlier, dramatic events. A generational difference also exists on how individuals view the Chinese leadership's decision to use force in the Tiannemen incident in 1989 (OSTROM, 2005).

Still within the frames of instrumental rationality, Frohlich and Oppenheimer (2001) also address the puzzle of multiple internal models by asking what factors in an environment are likely to affect the way a participant perceives a situation and the way the individual optimizes in that situation given that "paying attention" is costly. As Ostrom (2005) explains, few humans gain full mastery of all of the potentially relevant details in a complex, ongoing situation. Accordingly, Frohlich and Oppenheimer single out two properties of a situation as being the most important in affecting the way a situation is perceived, namely: *salience* and *vividness*. The authors explain that the salience of an element in a choice situation related with "the degree to which [the] element is linked to possible changes in the welfare of the decision maker" (FROHLICH; OPPENHEIMER, 2001, p. 8). We already saw how the IAD framework links this notion of salience with the likelihood of self-organized solutions for collective action by suggesting that the users must be dependent on the resource for a major portion of their livelihood or other variables of importance to them. As for the vividness of the situation, Frohlich and Oppenheimer refer to the "amount and quality of the sensory details of the objects encountered" (idem.). The authors explain the importance of these attributes in driving individuals' attention within the variety of signals they receive.

> In order for something to grab one's attention it must displace something else to which one is attending. To accomplish this, a new focus of attention must have a higher claim. Attention shifts from one object of attention to another as if there were a threshold of value attached to the former which has to be surpassed for the competitor to displace it (FROHLICH; OPPENHEIMER, 2001, p. 8).

Ostrom (2005) then adds these elements in Denzau and North (2005) general scheme for the relationship between mental models, action-outcome linkage, information feedback and communication, as illustrated in Figure 5.8.

This construction clearly adds something significant but certainly do not solve the puzzle because what defines vividness and salience, i.e., the value attached to alternative objects of attention is the mental model itself. To be sure, the examination of relationship between perceptual readiness and the hypothesized levels of human existence was one the first lines of investigation pursued by Graves in order to check the corresponding dominant behavioral characteristics associated with them.

At that time (it was the 1960s), operational definitions of the levels of existence were obtained by subject's score on tests for dogmatism and rigidity within the personality. It was then predicted that subjects whose behavior indicated the presence of a particular level of psychosocial development would perceive words representing the dominant ethical and motivational components of the level quicker than the words representing other levels. Upon tachistoscopic presentation of the words, subjects hypothesized to be at a particular level perceived the words of that level quicker than those of the other levels. The results of statistical analysis (chi-square test of independence) indicated that the theoretically sound associations were not random, at high level of confidence (GRAVES, 1965). This was, nonetheless, only the first steps of extensive investigation involving up to 36 psychometric dimensions from which Graves determined quantitative, system specific, cyclical and cyclical trend characteristics associated to his suggested developmental sequence of development (see Appendix).

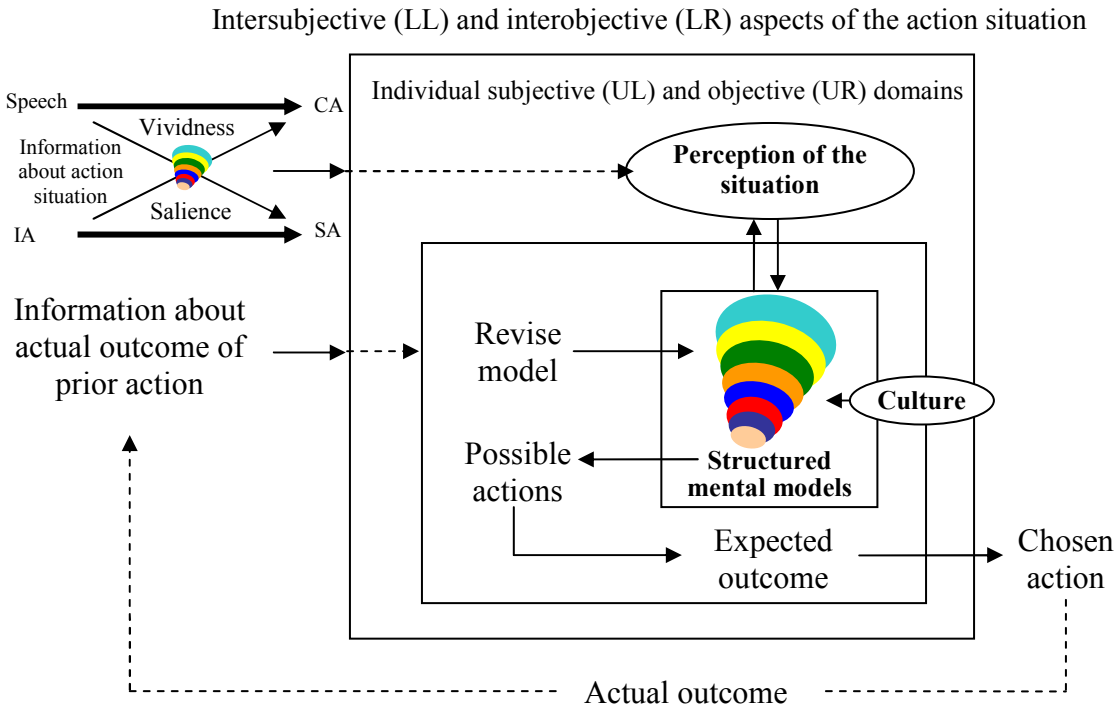
Again, it is here that the ECLET conception can certainly add something to our understanding, including a better account as to the learning that produces the said flawed belief systems, since Graves has indeed connected different learning processes attuned to each of the psychosocial systems he has identified (GRAVES, 1971a).

My overall suggestion is that, lacking a theory to bestow developmental structure and specific content for the mental models, the conceptual schema provided by the cognitive institutionalism, as in Figure 5.8, has no much use for predictive purposes, for we can predict neither the possible actions associated with different mental models nor how distinct mental models will perceive the eventual discrepancy between the actual and the expected outcome. Of course that the most general conjecture to introduce specific content into the mental model box in Figure 5.8 is but the model of "rational egoists", but this would lead us back to the starting point. Fortunately, the ECLET conception is available, providing us which a quite sophisticated and empirically ground theory of psychosocial development which fits pretty well our needs to inquiry into collective action problems. Accordingly, I place it in the core of Denzau and North (2000) schema of mental model revision, as illustrated in the Figure 5.9.

Again, the ECLET conceptions provide the information and the restrictions to predict in which extent, and possibly is which direction, mental model revision shall most likely occur. Note that I attempted to emphasize this aspect by redefining both the participant's and the exterior environments originally suggested by Denzau and North. Accordingly, the participant's domains is given by both her subjective level of existence (Upper-Left quadrant) as well as the corresponding neurological coping systems (Upper-Right) – or in fact the entirety of her biological constitution which supports the development of consciousness awakening. Likewise, the prior exterior (objective) environment, comprising the socio-ecological system (Lower-Right) is now extended to include the intersubjective domain (Lower-Right). This conceptual change allows us the analysis of communication to disengage from the exclusionary instrumental conception of rationality in order to incorporate the extremely noteworthy contribution due to Habermas's communicative action theory. Our suggestion is indicated in the Figure 5.9 through the substitution of the previous exclusively instrumental view of the role of communication for Habermas's typology of social action, as discussed in connection with Figure 5.5. The pictographic representation of the ECLET conception in the center of

Habermas's schema reports to the fact that communication takes place in hermeneutic circle of similar significations, where "information about the action situation", "salience", and "vividness" are all reliant upon both horizontal and vertical solidarity. Further, it reports also to the fact that the validity claims raised in the hermeneutic circle are not restricted to objective truths about the action situation but include also valid claims *a propos* of justness, rightness, or appropriateness and alternative actions – all of them are perceived quite differently, depending on the level of awareness or, more limitedly, level of psychosocial development.

This completes our conceptual framework to address collective action, and that orients the interpretation of our results. We shall now turn to the empirical strategy designed in order to check out whether or in which extent the ECLET conception has some predictive power regarding the observable responses of our participants in the experimental settings.



Source: Author's configuration.

Figure 5.9 – The ECLET conception as the "spiral dynamics" in action situation.

As just stated, the relevant aspect of the problem has less to do with objective validity claims – like which should be the proper coordinating strategy so that the collective outcome would be maximized – but with the intersubjective claims of justness, rightness, and appropriateness of setting up whatever rule to be followed by the entire group. Accordingly, the general hypothesis of this study, in this regard, is that the effect of communication opportunities will be a function of the different stages, levels, of waves of consciousness the participants in a group are navigating along the path of their ontogenetic revival (both Upper quadrants) of the phylogenetic history which have established the relevant sociocultural background (both Lower quadrants) (see Figure 5.1).

This is of course quite relevant to the study of collective action problems, particularly when the definition of *boundary rules* for limiting membership cannot be set up upon either objective (e.g. local of residence) or formal institutional (e.g. nationality) criteria (first design principle put forward by the IAD framework). Notwithstanding, however important, the horizontal aspect of solidarity does not suffice to foster mutual understanding, as the validity claims involving different perspectives on the justness, rightness, and appropriateness relating to how the social dilemma should be handled will be strongly affected also by the level of psychosocial development (Greaves) or more broadly the level of consciousness (Wilber, see Figure 5.3). This is the vertical dimension of solidarity of understanding, meaning that the people involved *share a similar depth* or level of consciousness, and, according to this level of consciousness, can form part of the fabric of cultural or horizontal solidarity that is the prerequisite for mutual understanding.

Note that this is why, however important it might be (as it is), the publication of design principles grounded in empirical evidences attesting their relevance for successful self-organized solutions for the commons dilemma, it does not suffice to foster the emergence of new solutions, since the very design principles shall be filtered or evaluated by the most diverse intersubjective perspectives as to their justness, rightness and appropriateness.

Apart from the relativist resistance in acknowledging the evolutionary/developmental views applied to culture, value judgments, and moral issues, which is the essence of vertical aspect of solidarity, the importance of the horizontal aspect of it is quite recognized, and has been actually fostering the current emphasis on self-organized solutions and collaborative inquiry to address local social dilemmas. In this extent, the horizontal aspect of solidarity is incorporated in the IAD framework as the "attributes of community" (see Figure 2.6), while it's vertical dimension starts to come up in highly controversial studies which try to find the link between cultural values and material progress (HARRISON; HUNTINGTON, 2000). In this regard, Shweder's contribution, in the just cited reference, provides spirituous insights as to how much carefully the discussion about how culture matters ought to be conducted (SHWEDER, 2000). Granted this understanding, I will not expand along with either Habermas's or Wilber's reconstructive explanations supporting the correspondences conveyed in the Figure 5.1. The fundamental logic of it, as already discussed, ensues right from the fundamental tenets of Integral Theory, grounded on the concept of holon and it's evolution by envelopment (transcend-and-inclusion principle).

As indicated, the purpose in this section is to make clear both the limitations and the possibilities of our intended generalization, which connects the ECLET conception with the most general process of consciousness awakening, toward the farther reaches of human potential, and ultimately getting to the validity claim of mysticism. In this regard, I remind that my purpose is not to defend this claim in isolation, but only in the measure that it naturally connects with our subject matter, once we have granted that holding neither *secularism*, as a privileged epistemological stance, nor *subtle moral skepticism*, as preferred attitude before individuals' preferences, apply for the matter of addressing our global social dilemmas.

6. EMPIRICAL STRATEGY AND THEORETICAL EXPECTATIONS

In order to examine the significance of consciousness evolution on the issue of collective action, we set up an empirical strategy which combines the assessment of Clare Graves's constructs with individual and collective behavior observed under experimental conditions.

We first discuss the procedures for measuring participants' behavioral functioning according to Graves's psychosocial systems. In the sequence we present the experimental designs and derive the behavioral hypothesis associated to both the game theoretical approach and each of the selected psychosocial systems as set forth by Graves's model. We then comment on the process of sampling and finally describe the statistical procedure for hypotheses testing.

6.1. Accessing the Gravesian psychosocial levels of existence

The assessment tool to identify the Gravesian psychosocial systems applied in this study consists of a Portuguese translation of the instrument developed by Marylin Hurlbut, in her PhD dissertation at the North Texas University, in August, 1979. For the development of the test instrument, Hurlbut worked in conjunction with Don Beck and Christopher Cowan at the National Values Center. The National Values Center (NVC) is an independent corporation

in Denton, Texas, established for the purpose of furthering research in values systems through the use of Graves's ECLET conception of adult development.

Hurlbut (1979) comments that Graves was personally acquainted with the work of the National Values Center, having visited the center himself in 1978 and having conferred with Beck in New York on numerous occasions. She adds that Graves was thus aware of the workshop activities of Beck and Cowan and has given his approval to their interpretation of his theory. Given, besides, Beck and Cowan's experience presenting and applying the theoretical concepts of Psychosocial Levels of Existence all over the United States as well as into Alaska and Spain, Hurlbut regarded them as the most qualified people in her geographic area to examine the test she was developing.

According to Lee's (1982) literature review, Hurlbut's test was the first truly compressive instrument to measure an individual's behavioral functioning according to their value systems as set forth by Graves's classification system. Differently from earlier attempts, Hurlbut's test was designed to reveal a person's psychosocial levels of existence with reference to his/her overall lifestyle and not any compartmentalized area of life such as his/her professional occupation, family life, religious beliefs, etc. Thus, while it is theoretically possible that a person may have mixtures of different systems and functioning on bases of different developmental lines or streams, Hurlbut's test was designed to reveal a person's overall dominant value system, secondary value system, and so on.

Hurlbut first translated Graves's theoretical levels of existence into discrete components of attitude and behavior which could then be assessed with a written test instrument. She began by collecting as many beliefs, one-word descriptors, typical behaviors, etc., as she could think of for each of the psychosocial systems from C-P through G-T. She explains that, some of the work of compiling characteristics for each level had already been done by Graves himself and that his descriptions were, of course, used. However, many additional characteristics were compiled in an attempt to translate the generalized

features of each system into specifics indicating typical ways that each wave would manifest in contemporary American culture (HURLBUT, 1979, p. 134).

These lists of characteristics were then examined by Beck and Cowan to evaluate their content validity (validity by definition). They examined each item on the test carefully and any question which they deemed non-representative or inaccurate was omitted from the instrument or revised according to their recommendations.

Items were chosen on the criteria of what statements were most representative of each system and at same time were atypical of all other systems. The final test consisted of forty multiple choice questions in the Most Like Me/Least Like Me format and it was administered to a total n of 4,029. The instrument was then subjected to several types of validity testing, including test for internal consistency, concurrent validity and construct validity.

Various internal consistency tests were made during the construction of the instrument itself. It was accomplished by means of field testing the effect of systematic inclusion of additional questions containing new "most representative" sentences per value system. The goal was to verify whether such increasing number of representative sentences per value system were consistently identified and selected by the people taking the test. Instances of sentence inconsistently selected in relation to the respondent's total score were object of further examination for either exclusion or reformulation. At the end of such a process, a hundred eighty sentences were distributed along forty multiple choices questions (thirty containing four statements and ten with six statements) so that there were equal representation among the levels of existence being tested including both negative and positive aspects of the various values systems.

The reliability indices (Spearman Rank Order correlation coefficients) were obtained using the split-half technique and correlating the two halves of the each individual test. In order to estimate the reliability of the full test the Spearman-Brown prophecy formula was applied to the median. The indices of 0.91 (n 105) and 0.88 (n 2,220) were thus obtained. The same split-half technique was applied to obtain reliability indices for the various levels of existence

separately. These indices (Pearson Product Moment correlations coefficients) ranged from 0.51 to 0.85 but Hurlbut commented that examination of the data revealed that low coefficients were primarily a function of low numbers. She thus suggested that a test-retest procedure would be more appropriate to assess the reliability of the test for separate systems (HURLBUT, 1979, p. 162).

Concurrent validity (discriminant validity) was established by comparing the predicted scores for thirty respondents with their actual scores on the test instrument. The judges – Beck, Cowan or Hurlbut – predicted the participant overall test score by ranking the person within the Levels of Existence classifications, identifying the person's dominant wave, secondary wave and so on, assigning a rank to each wave (CP through GT). The test was then administered to all participants. Test scores were compared with the judge's evaluation using the Spearman Rank correlation coefficient. The median of the correlation coefficients for Beck was 0.8928, for Cowan 0.8286, and for Hurlbut 0.8858. The median for all thirty Spearman Rank correlation coefficients was 0.8858.

Finally, construct validity was established by means of testing three hypotheses, namely: (1) the results of the Levels of Existence test for persons under thirty years of age would differ significantly for persons thirty years of age and over; (2) the Levels of Existence test would discriminate between different occupational groups of people (comparison were made between counselors, managers and clerical employees on their scores on Absolutism, Achievism and Sociocentrism); and (3) the primary and secondary waves of existence would be within three levels of each other.

The first hypothesis was rejected at the 0.05 (n 179) level of significance, showing no statistical difference in the test results between the groups of age. The second and third hypotheses were accepted. The distinction among occupational groups was established on the basis of both t-Test for two independent samples (n 179) and hierarchical grouping (n 99). The limit range of three levels between the primary and secondary psychological systems were found to be true for 98.9 of those tested (n 216).

Thus, Hurlbut's general conclusion was that the Levels of Existence test meets the standards of reliability and validity accepted within psychometrics sufficiently to recommend that it be revised with respect to certain details as specified in her study, and that further research be undertaken. Particularly, it was suggested that future research considered the revision of some few statements (distributed among five questions) which received twenty or more anomalous answers (within an n of 100 for each level of existence). These statements were individually analyzed and Hurlbut's recommendations for revision were incorporated in the online version of assessment tool currently in use at the National Values Center.

It is such revised version that we translated into Portuguese for this present application in Brazil. Clearly, Hurlbut's test was expected to be valid only for speakers of Standard American English, as well as for persons with the equivalent of a high school education. In spite of this, we undertook the translation with the consent and in collaboration with the National Values Center, since Graves regarded his ECLET conception to be fundamentally cross-cultural valid. Other applications of the Spiral Dynamics around the world, as well as our own, have continuously supported this claim, as we shall see in the next Chapter.

The process included three independent translations which were confronted each other. One translation was made by a professional translator, the other by the author of this study and the third by an author's colleague at the Federal University of Viçosa. The three translations were compared and when differences were found decision was made according to this author understanding of which translation reflected better the spirit of the particular statement according to Graves's theory and Hurlbut's test design.

The author's qualification for doing so was conferred by the National Values Center Consulting through the completion of two Spiral Dynamics workshops in Gravesian theory and further study of Graves's classification system. This included the study of Hurlbut's (1979) and Lee's (1982) PhD dissertations on test design, various publications concerning Graves's theory (BECK; LINSCOTT, 1991; BECK; COWAN, 1996; BECK, 2000a, 2000b;

GRAVES, 1959, 1961, 1966, 1970, 1971a, 1971b, 1974, 1981) and the recent Cowan and Todorovic's edition of Clare Graves's complete work (GRAVES, 2005). In addiction, it was consulted also Ken Wilber's interpretation of Grave's theory (WILBER, 2000a) as well as the correspondence between Graves's classification systems and Wilber's own (WILBER, 2000b, 2000c) and various other psychological developmental theorists reviewed by Hurlbut (1979), Lee (1982), Wilber (2000c) and Graves himself (2005).

As observed by Hurlbut (1979), in full recognition of the truth of Loevinger's statement a propos of the non-existence of absolutely certain signs of any level of ego development but only probabilistic ones (LOEVINGER, 1977), the estimation to the psychosocial systems is limited by the fact that attitudes, beliefs, and thoughts patterns are not accessible for direct measurement and therefore can only be measured indirectly trough self-report. As such, the test accuracy is proportional both to the test taker's awareness and understanding of his own personality and to his truthful disclosure of that personality when taking the test. On top of that, the necessity of using a translated instrument adds some concern regarding cultural differences. In special, it was required to take into account that each of the representative statements in Hurlbut's test was designed not only to reflect the content of each system but also to be in familiar wording, so that it could "sound right" from the perspective of a person centered in that particular level of existence.

Though this delicate aspect could appear at first sight as a source of additional difficulties, it in fact worked as reference to ascertain that the translation was reflecting properly the intended meaning and vernacular expression of each statement. As a last steep in the translation process, the result was submitted to a Native American speaker colleague, proficient in Portuguese, for a final revision and concurrence. This submission resulted in few minor changes in some subtler statements representative of the C-P level.

The National Values Center then implemented the translated instrument online and we ran a brief pre-test with only twelve respondents. The respondents' individual answers and overall profiles were analyzed for internal consistency.

The results leaded us to suspect that one C-P and one G-T statements were not being correctly identified. Thus, after submitting our conclusion to the National Values Center people, we reviewed and changed these statements accordingly. The changes were implemented in the online system and the instrument was regarded ready for our intended application.

6.1.1. Participants and constitution of the sample

We deem that the developmental character of Graves's psychosocial levels of human existence asked for an intentional sampling technique, in order to increase the likelihood of having instances of people centered in the different systems. This should be attempted by means of looking for voluntaries for the experiments in particular settings and organizations.

According to Wilber's (2000a) suggestion, for the US culture, we could look for B-O-like (tribalistic) people in athletic teams and corporate "tribes". The C-Ps (egocentric) would be more likely found amid the rebellious youth and soldier of fortune, while the D-Qs (absolutists) amid religious fundamentalists, Boys and Girls Scouts and patriotic organizations. The E-Rs (achievist) could be found among the emerging middle class, trophy hunters and amid people exploring the cosmetic and fashion industry. Along with the sympathizers of animal rights, deep ecology, ecofeminism, humanistic psychology, diversity and human rights movements we could find instances of people centered at the F-S (sociocentric) level. It would be more difficult to find instances of people centered at the G-T (individualistic/integrative) level, since according to Wilber (2000a) they would represent about one percent of the world population and would be very unevenly distributed.

In view of Wilber's suggestions, we believe that cultural differences between the United States and Brazil, though certainly important in many specific matters, do not play a major role for the purpose of identifying the Gravesian psychosocial systems. In fact, his indications come together with the settings were we could expect to gather instances of these systems in Brazil as

well. Moreover, Graves himself suggested that his classification system for adult behavior and psychosocial development would be applicable to any adult human being, regardless his culture (GRAVES, 2005, p. 4).

However, the practical circumstances of our research lead us to follow the sampling practice generally found in the literature on experimental economics. This means that we were restricted, for the most part, to the undergraduate and graduate students in our geographic area.

For the recruitment of the participants we used the service of the student office that sent out our message announcing the experiment to all students who had email in the University Federal of Viçosa (virtually all students). We also had the support of the Associations of Graduate Students, University's Employees and Professor Staff to email our announcement to their respective lists of contacts. Finally, we fixed a printed version of the announcement into virtually all buildings in the University Federal of Viçosa and in various spots distributed all around the downtown.

The announcement asked people to make contact with the author or with people at the *Campic* in order to access the questionnaire which was required for those interested in taking part in the experiments. We offered a fix R$ 5.00 (currently about US$ 2.25) payment for the completion of the questionnaire, conditional to the effective participation in the experimental phase of the research. The announcement briefly explained that the experiment was part of a research in "Applied Economics" where individual should make decisions concerning alternative investments. It was also informed that the experimental incentive consisted of variable financial rewards depending upon the participant's decisions. We added that the variable incentive plus the R$ 5.00 would be paid immediately after each experimental section.

As a result, we received more than 500 contacts of people interested in taking part in our research. In total we have 322 people who filled out the Spiral Dynamics questionnaire, being 142 (44%) females and 180 (56%) males. As for the ages, since the majority of them were undergraduate and graduate students, 43% were between 18 and 21 years old and 52% between 22 and 29 years old.

We have only 10 people in their thirties, 3 in their forties and just one man in his fifties. Concerning birthplace, 92% were from the Minas Gerais state, 3% from São Paulo, and 2% from Rio de Janeiro. The few remaining were from Bahia, Espirito Santo, Brasília (D.F), and Pará.

Unfortunately, the Spiral Dynamics on-line system was not set to register the information concerning the course or main occupation of the respondents, and we lost the opportunity of doing so during the experimental sections as well. Still, from the information supplied in the recruitment responses we know they come from the most diversity of graduate and undergraduate programs offered at the Federal University of Viçosa, including Veterinary Medicine, Agronomic Engineering, Environmental Engineering, Civil Engineering, Production Engineering, Physics, Math, Economics, Geography, Business Administration, Foodstuff Technology, Law, Biology, Botanic, Biochemistry, Administration of Cooperatives, among others. We have also a few staff professors, other university employees and Viçosa's citizens.

Had we systematically collected such data we could test the associations between the levels of existence and the chosen undergraduate course by the subjects as did Hurlbut (1979) as to different occupational groups of people. Specifically, Hurlbut was able to show that her Levels of Existence test discriminated between counselors, managers, and clerical employees on their scores in Absolutism, Achievism, and Sociocentrism. This was accomplished on basis of the results of two statistical tests, t-Tests ($\underline{n} = 179$) as well as hierarchical grouping ($\underline{n} = 99$).

6.2. Producing behavioral observations

The behavioral observations to be correlated with the individuals' psychosocial profiles were produced through three laboratorial experiments.

One strength of the experimental approach to address the issue of collective action is that it allows the creation and thus the control of the seven components of the action situation, namely: (1) the number of *participants*, (2)

the *positions* they may hold, (3) the specific *actions* they can take, (4) the *outcomes* they can affect, (5) the *transformation functions* that link actions and outcomes, (6) the *information* they obtain, and (7) the potential *payoffs*.

Also, experiments can be tailor-made for controlling the particular attributes of the environment, such as the physical, cultural or deontological attributes, required for investigating a particular model. Experiments can also induce and measure values, costs, and information flows that may be impossible to measure in natural settings (OSTROM; WALKER, 2002).

Owing to such qualities the experimental approach is regarded particularly adapted to test the foundations of game theory as well as for delving deeper into findings obtained in the field (ROTH, 1991). Roth particularly stresses the importance of field and experimental studies when economics are to be used as foundation for recommending social polices. And Gints (2003) suggests that the progress in modeling human behavior has been hampered by the underutilization of controlled experiments, which are common only in behavioral psychology.

Yet, as Ostrom et al. (2002, p. 101) point out, "The strength of control in experimental analysis is a weakness from the standpoint of the richness that one gains from studying field data". Thus the authors emphasize the complementary nature of the laboratory investigations with other approaches. For instance, Jager (2000) stresses some important limitations of the experimental approach that can be surmounted by means of computational simulation techniques. With regard to the factor affecting behavior, Jager suggests that whereas experimental games are useful for studying the effect of group factors, such as the attributes of the community, it is less practical for studying the individual factors. He explains that individual factors such as personality, social orientation, social values and ideas of morality, cannot be experimentally varied. Thus, he says, it is virtually impossible to develop an experimental design with exactly the same distribution of person characteristics in each of the experimental conditions (JAGER, 2000).

Yet, though Jager's last observation is certainly true, the study of the effects of the individual factors affecting behavior does not need to be

accomplished by means of reproducing the same distribution of individual characteristics in each of the experimental conditions. Indeed, in the present study we put forward precisely the opposite strategy by fixing the experimental condition and allowing the distribution of individual characteristics to vary. Since we have access to those individual traits, as possible through the levels of existence test, we can correlate them with the individual behavior under the controlled conditions. Thus, at least for the cases where significant statistical association between the variable measuring individual behavior and that measuring the individual factors exists, we have some empirical evidence of how that individual trait affect behavior under the specified experimental conditions.

A more compelling issue discussed by Smith (1982) is the creation of adequate *inducting value* so that the laboratory situation leads participants to perceive and act on payoff that have the same mathematical properties as the payoffs in the theoretical model with which it is linked. Smith then suggests that the following conditions constitute a set of sufficient conditions for inducing value:

1. **Nonsatiation**: Subjects' utility must be monotone-increasing in payoffs, so that subjects will always choose the alternative with the greater reward.
2. **Saliency**: The level of rewards must be motivational relevant, so that rewards are directly tied to subjects' decisions.
3. **Dominance**: The cash reward structure must dominate the other subjective benefits and costs of an action present in the subjects' utility function.
4. **Privacy**: Subjects' payoffs should be made privately.

We will see in a moment how the laboratory experiments were designed in order to secure inducing value. First, though, we regard that a brief word concerning the condition of dominance of the cash reward is opportune, since we are primary interested in exposing the effects of other subjective arguments in the subject's utility function.

In fact, the dominance of the cash rewards over the other arguments of the subjects' utility function is necessary to guarantee a *common knowledge* concerning the reward structure involved in the experimental design. Together

with the conditions of nonsatiation and saliency, the condition of dominance assures us that all subjects act upon the payoff structure with the intent of obtaining the maximum cash reward. In other words, the dominance of the cash reward must work as a common factor affecting behavior. Yet, *how* (specifically) the subjects act in order to maximize their cash payoff depends on other subjective arguments in their utility functions, such as other regarding preferences, ideas of morality, and many other individual values. Alternatively, from the analytical viewpoint, the dominance condition could be represented by seen these other subjective arguments as *restrictions* imposed upon a utility function limited to the cash reward. Either way, the fulfillment of the conditions for inducing value is critical for distinguishing between the common and the individual factors underneath the behavioral variability observed in the experimental settings.

We now turn to the description of the experimental designs and settings used to generate the behavioral observations that were correlated with the individual factors, as measured by the levels of existence test.

6.2.1. The common-pool resource experiment

There are a number of experimental protocols to investigate the general situation involving the appropriation from a common-pool resource. Since we are interested in examining the interlay among consciousness development, alternative institutional settings and economic performance, we followed Ostrom et al. (2002) experimental designs in order to test the contribution of the ECLET conception in explaining the capacity to improve the collective performance under three different institutional settings: unregulated or "open access" (baseline), costless communication and costly sanctioning.

6.2.1.1. Experiment summary

The experiment consisted of thirty rounds in which eight participants must decide how to allocate an endowment of fifteen tokens, $x_i \in [0, 15]$, between two alternative investment opportunities.

The alternative opportunities were generically labeled as Market 1 and Market 2. Market 1 is a safe, outside activity in which each token yields a fixed (constant) rate of output and each unit of output yields a fix (constant) return. Market 2 (the CPR) is a market that yields a rate of output per token dependent upon the total number of tokens invested by the entire group. The individuals' share of the total output produced in the Market 2 is equivalent to the percentage of total group tokens invested in that alternative. Thus, the return from Market 2 is variable and uncertain while each unit of output yields a fix (constant) return.

Each round of the experiment is a *repetition* in the sense that the endowments, the production functions, and the payoff functions are not dependent upon decisions made in any round. Also, the endowment is provided by the experimenters so that the experimental situation separates appropriation activity from provision.

The experimental design involved three different conditions, run in the following sequence: (1) ten rounds where decision were made without communication; (2) ten rounds following ten minutes of imposed and costless communication, and (3) then rounds with an sanctioning mechanism supplied. Additionally, a new and not obligatory communication opportunity was offered between the 15^{th} and the 16^{th} rounds. In short, the structure of the experimental design schematizes as in Figure 6.1.

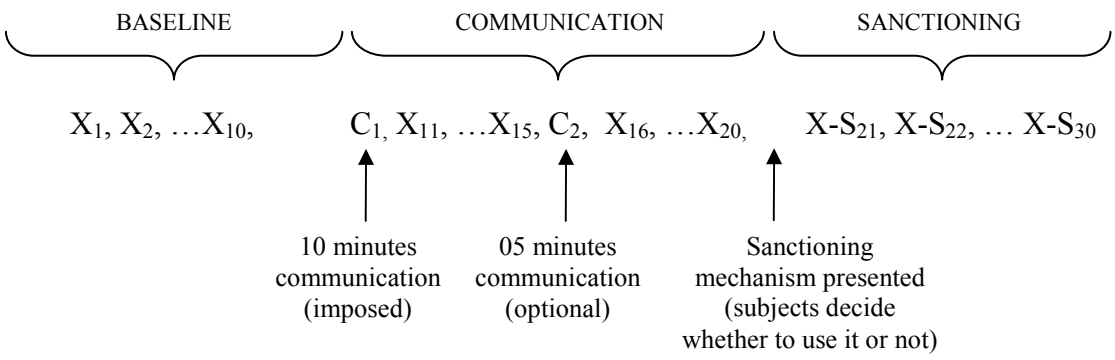


Figure 6.1 – CPR experimental design.

Subjects knew with certainty the total number of rounds in the experiment, the total number of decision makers in the group, total number of tokens, and that the endowments were identical. Moreover, subjects were informed that the experiments consisted of the three different situations specified above. We discuss the necessary adjustments in our procedure due to this information in the following sections.

Following Ostrom et al. (2002), Market 1 was described as an investment opportunity in which each token yielded a fixed (constant) rate of output and each unity of output yielded a fixed (constant) return. Market 2 (representing the CPR) was described as a market that yielded a rate of output per token dependent upon the total number of tokens invested by the entire group.

The subjects were informed that they would receive a level of output from Market 2 that was equivalent to the percentage of total group tokens they invested, times a fixed (constant) rate of return per unit of output. The values of the cash rewards at each level of investment in both alternatives were described in tabular form.

6.2.1.2. Participants, experimental setting and procedures

A total of 200 participants (94 female and 106 male) took part in the CPR experiment, in the total of 25 experimental sections with 8 players each. We have 4 missing cases regarding the Spiral Dynamics profile owing to compulsory recruitment of individuals to substitute more then expected absences. The subjects' ages ranged from 18 to 47 years, though only a few were above 30 years old and only one man was in his forties.

Since at this time we have no computational laboratory facilities for conducting this kind of experimental research at the University Federal of Viçosa, the experiments were run into a large classroom at the Rural Economy Department. Participants who have previously completed the Spiral Dynamics questionnaires were consulted mainly via email as to the likelihood of taking part in alternative experimental sections. As confirmations were arriving we settled lists with 10 names (8 titular and 2 substitutes) and confirmed time and local with them. We offered extra R$ 5.00 (about US$ 2.25) as incentive for substitutes showing up; in case they needed to be dismissed without actually participate in the experiment.

In conducting the classroom experiments, we followed Murphy and Cardenas (2004). Accordingly, at the experimental setting participants were accommodated as to guarantee the privacy of individual decisions and received an envelop containing: (1) the writing instructions, (2) the payoff tables (Market 1, Market 2 and Total Payoff), (3) a sheet with 30 experiment cards (one for each round), (4) one record sheet, (5) a sheet with an simulated round used as example, (6) a three test-rounds sheet, and (7) a calculating machine. Copy of the written instructions and other distributed material are presented in the Appendix.

Before beginning, identification numbers were randomly ("pick one") distributed, questions were answered and the test-rounds were played. At the test-rounds, assistants checked participants' computation in their test-rounds sheet, and this required the identification of individual decisions. When errors were found, doubts were elucidated and a second-round test was conducted. No third

test-round was necessary in any group. Once full understanding of the game and required procedures were assured, identification numbers were randomly assigned again.

In each round, participants filled out the experiment card informing: (1) his or she identification number, (2) the round number, and (3) his or her decision as to how to invest in the Market 2. The experimental cards were then collected by the research assistant and individual decisions were publicized and kept in the blackboard according the identification numbers. Note that players do not report their names on the experiment cards, only the player number. Names are, however, reported on the record sheet so that, in post-game procedures, we can match individual decisions with the corresponding individual psychosocial profiles. The privacy of the individual decision during the experiment was nevertheless assured by utilization of the experimental cards and collecting procedures.

Once decisions were published in the blackboard, experimenter tallied them in an Excel worksheet and then announces the total sum invested in the Market 2. No information on average or marginal returns was supplied. Participants consult the payoff matrix, fill out the record sheet, find out their earnings in that round and are ready to take a new investment decision. This make the constituent game X, which was repeated as indicated in Figure 6.1.

As pointed out, up from the outset subjects knew the complete experimental design, including the specification of the three experimental conditions. Such previous information was decided owing to our particular milieu at the Federal University of Viçosa. We were concerned with the occurrence of asymmetry of information due to the possibility of post-experimental communication among acquainted participants, as the experimental sections were succeeding (we conducted 25 repetitions of the CPR experiment limited to one experiment per day during the weekdays and two at the weekends). So, in order to avoid that some participants came to the experimental site knowing the sequence of treatments whereas others did not have this information, we make the experimental design common knowledge from the

begining. Yet, this high level of information called for some further adjustments in our procedures.

In fact, the prior knowledge that a sanctioning mechanism would follow the communication condition hazards the nonbinding assumption during the previous treatment and, moreover, insinuates some expectation regarding the issue of a collective strategy. We then attempted to neutralize these adverse effects by: (1) exposing clearly the general nature of the appropriation problem, not by suggest any naturally occurring parallel decision environment, but simply by making clear the distinctive characteristics of the two investment alternatives; (2) by suggesting, contrary to our intention, that at the 19th round participants could choose whether or not they want a sanctioning mechanism; and (3) by establishing that, if they do want it, no sanction would be allowed as to occasions on the communication condition.

As for the suggestion concerning the potential advantage of a collective strategy, it was alleged that different groups came out with diverse conclusions as to how to proceed. The words repeated to every group were:

> Some groups have set up a joint strategy. Others have chosen to continue to play as before. There is no necessary requisite about this issue. Still, since some may want to set a rule while others do not, we have created two situations to handle this matter. During the ten rounds after communication, even if some people in this group regard advantageous to set up a joint strategy, there is no way to oblige others to behave accordingly. Only you will be given an additional and optional five minutes communication opportunity after the round 15th. As we said, you are not required to use up this second opportunity. At the moment we will ask you whether you want to use it or not. At the end of the 19th round, we will ask you whether or not the group wants to be supplied a sanctioning mechanism, in which case at the time we will explain how it works. In any case, the mechanism shall operate from that point on. No sanction upon behaviors in prior rounds will be allowed.

6.2.1.3. Hypothetic-deductive model of "rational egoists"

Once we have disentangled ourselves from the *symmetry thesis* embedded in the hypothetical-deductive model of science and have recognized that prediction do not implies explanation, we can take advantage of the "rational choice" model in order to obtain theoretical outcomes that should result from the

assumed behaviors in our experiments. To reinforce, such "predictions" are in fact just analytical solutions, that is, mathematical deductions that become possibly precisely owing to the highly abstract character of the hypothetical behavioral responses, and that imparts to them the unambiguousness required for mathematical treatment. The more concreteness we infuse into our model of individual and its social interactions the more economics ought to capitulate before insurmountable uncertainty (see e.g. ARTHUR, 1994).

As such, what the analytical model we shall present actually supplies is just a referential point of hypothetical equilibrium (mathematical solution) against with the actual behaviors can be contrasted. It offers neither a description of actual human behavior nor a necessary explanation, which should arise from logical symmetry, for the outcomes that can eventually be predicted based on its assumptions. Accordingly, the working hypotheses under the following deduction are:

1. Individuals possess as much *information* about the structure of a situation as in contained in the situation itself.
2. Individuals assign a complete and consistent, internal *valuation* to outcomes that is a monotonic function of an individual's own net external payoffs.
3. After making a complete analysis of the situation, individuals choose an action in light of their resources to *maximize* expected material net benefit to themselves given what others are expected to do.

These are the three fundamental assumptions of the models developed by the contemporary "rational choice" theory and that define the model of individual behavior that Ostrom (2005) refers to as the "rational egoist" model. As she explains, each of these assumptions can be unpacked into multiple layers "thus making the holon of a participant a very complex entity" (Chap. 4, p. 6). Moreover, she continues, once one decides to explore alternative theories and models of human behavior, the number of choices that researches has to make about each of the three components and their combinations is immense. Ostrom regards that "Alternative assumptions are not likely to be as clean and elegantly simple as those of rational choice models, and agrees with Frohlich and Oppenheimer (2001) and Levi (1997) in that the explanation of actual human behavior require more theoretical superstructure. Still, she adds, "The grave

hesitation of some theorists to adopt 'more realistic' assumptions stems, to a large extent, from the messiness of the alternative superstructures".

As we saw, Graves's ECLET conception was conceived just with the explicit purpose of putting some order in such a mess. As to what extend he was successful, one can judge by its consistence with the many other models of consciousness evolution cohered by Ken Wilber (2000).

Concerning the analytical tractability, Ostrom reports that one strategy adopted by many authors, including herself, is not to change all of the assumptions at the same time when trying to model a particular type of situation. She explains that, in stable and repetitive situations where *intrinsic values* are important, one may retain the assumption regarding full information about the structure of the situation (i.e. assumption one, above) while focusing primarily on the types of values of importance to individuals and resulting strategies they select.

To begin, let us see how the *Homo economics* behave in the CPR experiment.

"Rational egoists" in the CPR experiment: baseline condition

Following Ostrom et al. (2002), assume a fixed number n of appropriators with access to the CPR. Each appropriator i has an endowment of resource e that can be invested in the CPR (Market 2) or invested in a safe, outside activity (Market 1). The marginal payoff of the outside activity is normalized equal to w, measure in reais (R$). The payoff of an individual appropriator from investing in the CPR depends on the aggregate group investment in the CPR, and on the appropriator's own investment as percentage of the aggregate.

Let x_i denote appropriator i's investment in the CPR, where $0 \leq x_i \leq e$. The group return to investment in the CPR is given by the production function $F(\sum x_i)$, where F is a concave function, with $F(0) = 0$, $F'(0) > w$, and $F'(ne) < 0$. Thus, for relatively low levels of investment in the CPR, the returns are greater than the opportunity cost of the foregone safe investment [$F'(0) > w$], but as the

investment in the CPR increases the marginal return decreases and eventually become negative. Assume the level of investment for which the marginal return from the CPR become negative is $x_n < ne$. Then, the yield from the CPR reaches a *maximum net level* when individuals invest some, but not all, of their endowments in the CPR.

Considering the general a model for n participants, let $\mathbf{x} = (x_1, x_2, \ldots x_n)$ be a vector of individual appropriators' investing in the CPR. The amount the individual i does not appropriate to the CPR, $e - x_i$, is automatically invested in the safe outside alternative. Thus, considering the two investment opportunities, the payoff to an appropriator, $u_i(\mathbf{x})$, is given by:

$$\begin{aligned} u_i(\mathbf{x}) &= we && \text{if } x_i = 0 \\ w(e - x_i) + (x_i / \textstyle\sum x_i)\, F(\textstyle\sum x_i) && \text{if } x_i > 0 \end{aligned} \qquad (6.1)$$

That is, if players put all their endowment (e) into the safe alternative that returns a fixed rate of (w) per unit of investment, they get the sure value of (e) times (w). If players put some of their endowment (e) into the safe alternative and some into the CPR, they get a return of $w(e - x_i)$ on the part of the endowment $(e - x_i)$ invested in the safe alternative, plus the return they get from the CPR, which is $(x_i / \sum x_i)\, F(\sum x_i)$. The return from the CPR is proportional to the investment of the individual, x_i, in relation to the total (aggregate) investment in the CPR, $\sum x_i$ times total CPR output $F(\sum x_i)$, measured in reais (R$).

The analytical model concerning the actor component assumes all three conventional working hypotheses of the "rational choice" model, that is, we are dealing here with "rational egoists". It is thus posited that each player maximizes payoff given the strategies chosen by the other players. The analytical solution is then the Nash equilibrium, which is not efficient; hence the dilemma or, eventually, the "tragedy".

Accordingly, let x_i be player i's strategy and $u_i(\mathbf{x})$ be player i's payoff function, in a general formulation of which (6.1) is a particular instance. Player i seeks to maximize her payoff by her choice of x_i, which is itself a constrained

variable [$0 \leq x_i \leq e$]. Assuming, as usual, that $u_i(\mathbf{x})$ is totally differentiable, player i is posited to be before the following calculus problem:

maximize $u_i(\mathbf{x})$
x_i
subject to $0 \leq x_i \leq e$

The Nash equilibrium requires that all players have solved their individual maximization problems simultaneously. But since the solution of each maximization problem depends upon the choices of the other players, for $\mathbf{x} = (x_1, x_2, \ldots x_n)$, the analytical approach requires also *common knowledge*. That is, it requires that each player assumes the others are before the same maximization problem and each one arrives to a solution x_i^*. Given these, at Nash equilibrium, the problem that player i faces is:

maximize $u_i(x_1^*, \ldots x_i, \ldots x_n^*)$
x_i
subject to $0 \leq x_i \leq e$
which is solved by x_i^*.

In our experimental design each player has the same endowment, the same set of strategies (either pure or mixed), and the same payoff function in reais (R$). Under these conditions, the game is symmetric and the equilibrium is unique. To find this equilibrium, it suffices to solve a single player's maximization problem, together with the restriction that each x_i^* will be equal at equilibrium.

Given the specifications of the CPR production function [$F(0) = 0$, and $F'(ne) < 0$], neither $x_i = 0$ nor $x_i = e$ solve player i's maximization problem. Therefore, there must exist an interior solution, where the first-order condition is satisfied. Considering the payoff function (6.1), and differentiating it in relation to x_i one has:

$$-w + (x_i / \sum x_i)\, F'(\sum x_i) + F(\sum x_i)\, ((\sum x_i - x_i) / (\sum x_i)^2) = 0 \qquad (6.2)$$

Since the solution for this problem is unique and symmetric, $\sum x_i = nx_i^*$. Substituting this shortcut into (6.2) one has:

$$-w + (1/n)\, F'(nx_i^*) + F(nx_i^*)\, ((n-1) / x_i^* n^2) = 0$$

In order to see that the aggregate investment in the CPR at the symmetric Nash equilibrium is not an efficient solution we must compare it with the optimum solution to the present CPR problem. Summing across individual payoffs $u_i(\mathbf{x})$ for all appropriators i, one has the group payoff function $u(\mathbf{x})$,

$$u(\mathbf{x}) = nwe - w \sum x_i + F(\sum x_i) \qquad (6.3)$$

which is to be maximized subject to the constraint $0 \leq \sum x_i \leq ne$.

Given the above productivity condition on F, the group maximization problem has a unique solution characterized by the condition:

$$-w + F'(\sum x_i) = 0 \qquad (6.4)$$

meaning that, in order to obtain the maximal net yield from the resource, the marginal return from the CPR should equal the opportunity cost of the outside alternative. Since this result is different from equation (6.2), the Nash equilibrium is not an optimum.

Both the Nash and the optimum equilibriums above were derived for a single period. Yet, if the baseline game has a unique symmetric equilibrium, as we showed, then the finitely repeated game has a unique symmetric subgame perfect equilibrium (SELTEN, 1971). Thus, equation (6.2) characterizes a finite sequence of equilibrium outcomes where symmetry holds within a decision period as well as throughout the entire set of repetitions.

Operationalization of the constituent game: baseline condition

Still, following Ostrom et al. (2002), we have operationalized the CPR situation with eight appropriators (n = 8) and quadratic production function.

The authors comment that, while many CPR problems involve more than eight participants, with eight participants one encounters most of the strategic complexity inherent with larger groups. Moreover, they add that there is sound theoretical reason to believe that eight is a large enough number to surmount small-group effects (OSTROM et al., 2002, p. 109).

The quadratic production function $F(\sum x_i)$ is specified as below:

$$F(\sum x_i) = a\sum x_i - b\,(\sum x_i)^2$$
$$\text{with } F'(0) = a > w \text{ and } F'(ne) = a - 2\,bne < 0 \qquad (6.5)$$

For this quadratic specification, on has from (6.4) that the group optimal investment satisfies:

$$\sum x_i = (a - w) / 2b \qquad (6.6)$$

And from (6.2) that the symmetric Nash equilibrium group investment is given by:

$$\sum x_i = (n\,(n + 1))\,(a - w) / 2b \qquad (6.7)$$

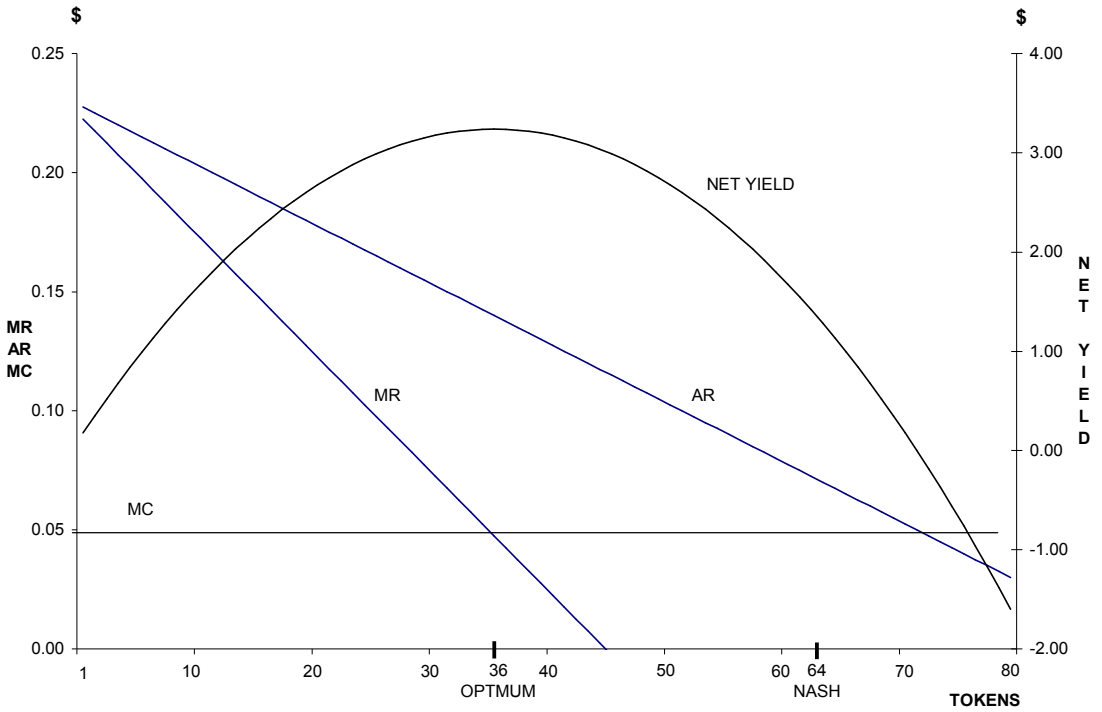
This level of investment is between the maximal net yield (6.6) and zero net yield, approaching the latter as n gets larger. The zero net yield occurs when investment is twice as large as optimum, $\sum x_i = (a - w) / b$.

One additional constraint that arises in the laboratory setting is that x_i be integer value, which is accomplished by choosing the parameters a, b, n and w in such a way that the predictions associated with $\sum x_i$ are all integer valued. In particular, the parameters utilized in the experiments reported here are the same used in Ostrom et al. (2002), except for the individual endowment. The

parameters' values are presented in the Table 6.1. With these parameters, the maximum net yield from the CPR (Market 2) is obtained when the aggregate group investment equals 36 tokens. With eight players and the restriction that x_i be integer, the optimal level of investment requires that four subjects invest five tokens and the other four subjects invest four tokens. This symmetric game has a unique symmetric equilibrium (Nash) with each subject investing 8 tokens (Figure 6.1).

Table 6.1 – Parameters for a given decision round (baseline)

Parameter	Value
Number of subjects	8
Individual token endowment	15
Production function (Market 2)	$23\ (\sum x_i) - 0.25\ (\sum x_i)^2$
Market 2 return / unit of output	$ 0.01
Market 1 return / unit of output	$ 0.05



(MR = marginal revenue, AR = average revenue, MC = marginal cost).

Figure 6.1 – Theoretical predictions.

<u>"Rational egoists" in the CPR experiment: communication condition</u>

From our brief summary on Habermas's theory of communicative action we saw how limited is the instrumental conception of rationality to address the effects of communication. Indeed, the ultimate conclusion drawn from standard noncooperative game theory is that communication, unless backed by enforceable agreements, is so unessential that can be entirely excluded from the analysis. As Harsanyi and Selten (1988, p. 3) put it:

> the decisive question is whether the players can make enforceable agreements, and it makes little difference whether they are allowed to talk to each other. Even if they are free to talk and to negotiate an agreement, this fact will be of no real help if the agreement has little chance of being kept. An ability to negotiate agreements is useful only if the rules of the game make such agreement binding and enforceable.

Ostrom et al. (2002, p. 146) observe that this implies that much of the contemporary noncooperative game theory in fact treats one of Nash's conditions

(Nash, 1950, 1951) separating the branches of cooperative and noncooperative games (i.e. no communication in noncooperative games) as superfluous. As such, as far as we keep the standard assumptions of noncooperative game theory, the possibility of a verbal commitment during the communication period C in our experiment does not change the theoretical equilibrium derived for the constituent (baseline) game X. Formally, when the symmetric constituent game X has a unique symmetric equilibrium x^*, neither finite repetition nor communication creates new symmetric equilibrium outcomes. Thus, if c denotes a communication strategy, and as long as saying one thing and doing another has no payoff consequences, then any strategy of the form (c, x^*) is an equilibrium of the one-shot game (C, X), and finitely repeated x^* is a subgame perfect equilibrium outcome of the one-shot communication $(C, X_1, X_2, \ldots, X_n)$ or repeated communication $(C_1, X_1, C_2, X_2, \ldots, C_n, X_n)$.

However, since repeated evidences have shown the significant effect of communication in increasing the selection of cooperative strategies in repeated interaction settings, researches have forwarded a series of hypotheses to explain such "anomalies". As summarized by Ostrom et al. (2002, p. 149) the most common hypotheses include: (1) offering and extracting promises; (2) changing the expectations of other's behavior, (3) changing the payoff structure, (4) the re-enforcement of prior normative orientations, and (5) the development of group identity. The authors observe that experimental examination of communication has demonstrated the independent effect all five of these processes, but also their interactions, where they appear to re-enforce one another. In any case, as far as the explanations remain committed with the exclusionary instrumental conception of rationality, they must either involve some nonrational source—whose typical example not by chance falls upon morality concern—or else be represented in the typical behaviorist arrangement, as in the core relationship involving reciprocity, trust and reputation.

As I have been empathizing, all these may provide adequate explanations. Yet, if rationality is not exclusively instrumental, then Habermas's typology of social interactions provides a better foundation to examine the effects

of communication. Accordingly, the requirement of both vertical and horizontal solidarity of signification as condition for rational mutual understanding regarding the justness or rightness of self-organized norms of appropriation from the CPR, together with the proposition that people may be navigating different waves of consciousness or levels of psychosocial development, can explain the uneven individual responses to similar communication opportunities and differences in groups' performance better than the mere assertion that people just differ on the grounds of incommensurable idiosyncratic aspects. Moreover, by controlling groups' composition respecting the psychosocial profile, one may be able to figure out and distinguish promising composition from those which approximates inefficient Nash solutions. Note that this very significant dimension of social interaction comes to add with the explanations based on strategic interactions alone (see Figure 5.5).

Procedures: communication condition

As already indicated in the experiment summary, following the 10th round participants were imposed 10 minutes communication opportunity. Participants were invited to move on a previously arranged site on the classroom and were offered to randomly pick one badge with an identification letter.

Since most of the instructions have already been introduced at the out set, we just add the following:

> You now will have the opportunity to discuss the decision problem you face. As we said before, you may discuss anything you wish during your 10-minutes period, with the following restrictions: (1) you are not allowed to discuss side payments (2) you may consult the payoff tables during discussion but are not allowed to see the private information of anyone's registration sheet, and (3) you are not allowed to make physical threats. Remember that you are not required to get to any agreement and no one is obligated to stick with an eventual agreement forcefully. Only you be given a second but no compulsory 5-minutes communication opportunity at the end of the 15th round.

Before leaving the room we start the video camera in order to record the participants' discussion. After this opportunity to communicate, the subjects returned to the constituent game as indicate in Figure 6.1.

"Rational egoists" in the CPR experiment: sanctioning condition

In order to represented formally the sanctioning condition in our CPR experiment, let s, the sanctioning matrix, be a matrix of 0's and 1's, where $s_{ij} = 1$ means that player i has sanctioned player j, and $s_{ij} = 0$ means that i has not sanctioned j. Accordingly, row i of the matrix s codes all of player's i's sanctioning behavior. As before, let $\boldsymbol{x}$ be a vector of individual investments in the CPR and $u_i(x)$ be i's payoff function in the game *without* sanctioning. Player i's payoff function in game *with* sanctioning, $u_i(x,s)$, is given by

$$u_i(x,s) = u_i(x) - f1\sum_j s_{ij} - f2\sum_j s_{ji} \tag{6.8}$$

The parameters $f1$ and $f2$ represent the cost of fining and the cost of being fined, respectively. Notice that the term *fine* is used not in the context of redistribution. The crucial point here in that real resources are used up, and not redistributed, by efforts to sanction. The sum $\sum x_{ij}$ is the total is the total number of fines j levied by player i, costing him or her $f1$ each; the sum $\sum x_{ji}$ is the total is the total number of times player i is fined i, costing him or her $f2$ each.

Adding this sanctioning mechanism to our constituent game X produces a game X-S with unique symmetric subgame consistent equilibrium. In a one-shot game with a unique symmetric equilibrium x^*, any sanctioning is costly and cannot lead to higher payoffs. Thus, the symmetric equilibrium of the one-shot game with sanctioning is the pair $(x^*,S^*) = (x^*,0)$; that is, the equilibrium sanctioning matrix is the 0-matrix (no sanctions at equilibrium). In our design the game X-S is repeated for $T = 10$ times or rounds. This finitely repeated game has a unique symmetric subgame consistent equilibrium given by (6.9):

In every round, play $(x^*,0)$.
In the event of any deviation from prescribed play,
resume playing $(x^*,0)$ after the deviation. (6.9)

Given the three fundamental assumptions, this equilibrium follows from *backward induction*. Accordingly, the "rational egoists" are presumed to reason thus: at the last round T, no deviation is profitable. At the next to last round $T-1$, given that no deviation will occur in the last round, for all players are "rational egoists" replicas and all know that all know this, then no deviation is profitable as well, and so on backwardly all through round one. So, repeating the game should not lead to sanction either.

Since in our design the sanctioning condition follows a communication condition, and the issue about communication is that players may agree to invest *less* than they would according to (6.9), there are many imperfect equilibria besides this unique symmetric subgame consistent equilibrium. Accordingly, let $z_i < x_i$ be the same for all i. One of these imperfect equilibria run as follows:

In every round except T, play (z, 0)
In the event of any deviation, play ($x_i = e$, $s = 1$) for one round,
then resume playing (z,0).
If no deviation took place in round $T-1$, play (x*,0) in round T. (6.10)

That is, given the agreement of investing $z_i < x_i$, if some player cheats, then every player dumps all their resources into the CPR ($x_i = e$), and every player issues one sanction for one round. Then play returns to normal (z, 0). The equilibrium (6.10) is imperfect because the trigger punishment threat ($x_i = e$, $s = 1$) is too harsh to be credible at the end of the game, so that in the final round everyone play the one-shot Nash equilibrium (x*,0). To show that (6.10) is indeed an equilibrium it suffices to prove that no deviation from prescribed play ($z_i < x_i$) pays. Following Ostrom et al. (2002, p. 173), let $F(ne)$ be a very large negative number. For $f1$ and $f2$ large enough, a player who deviated optimally for one round would lose [cic?] some positive amount, depending on the level of z_i, but the nest round would lose $(1/n)\ F(ne) + f1 + f2$ due to punishment from overinvestment and sanctions, as in (6.10). This threat the authors call *dire threat*, as it is the worst threat imaginable for one round in this design. So, given such a threat does not pay to deviate, even for one round. As mentioned, if a

punishment is not called in the last round, the endgame equilibrium is played at that round. This shows that (6.10) is an equilibrium.

Finally, there are a large set of equilibria along the lines of (6.10), involving variations of the length of punishment (1 or more rounds), the base level of investments z_i, and the direness of the one-period threat (dump not quite all tokens in the CPR, levy fines with some probability). In particular, by varying $f1$ and $f2$ (or offering options as we did?), we hope to allow the subjects to find equilibria of the family (6.10) that involve punishments of the form $(z_i,1)$ – that is to say, reduced investment in the CPR, but sanctions for everyone if a deviation occurs.

Operationalization and procedures: sanctioning condition

As we said, at the end of the 19th round we asked the participants whether or not they want to be supplied a sanctioning mechanism. At this moment, however, it was explained that the sanctioning mechanism was costly, and consisted into a set of fines that could be imposed at a corresponding fee. Following Ostrom et al. (2002, p. 174), subjects were given an announcement summarized below.

> Subjects were informed that in all remaining rounds each would be given the opportunity to place a fine. Each subject could levy one or more fines, directed to different player, at specific fees. The subject fined would pay a fine of a specific amount. It was possible for a single subject to be charged multiple fines. After each round, each subject filled out a fining form. These forms were collected and tallied by the experimenter, who then reported the results privately to each subject. Note that any subject who was fined did not know the identity of those who imposed a fine. At the end of the experiment, the experimenters subtracted from subjects' total profit the total of all fees and all fines.

Contrary to Ostrom et al. (2002), however, we added that no sanction was expected to be imposed either if the group has not settled a norm or if the norm has not been violated. Further, as already noticed, we established that no sanction were allowed as to defections occurred during the communication condition. The authors, otherwise, explain that, in their *imposed sanctioning-no communication* experimental setting, "No discussion was held on why the

subjects might want to use the procedure or its possible consequences", and that "This created an experimental setting as close as possible to the noncooperative assumptions of no communication and no capacity to engage in enforceable agreements".

Still, since our setting does have communication and, moreover, circumstances forced us to reveal the sanctioning treatment in advance (in order to avoid information asymmetry), our concern was focused on assuring the nonbinding assumption during the communication condition, not during the sanctioning condition. As already pointed out, we attempted to do that by suggesting that the group could choose whether or not they want a sanctioning mechanism and, at the same time, inducing the participants to accept it by exposing that, after all, even with the sanctioning mechanism at their disposal, they were neither forced to apply it nor expected to do so if the agreement is being respected.

We thus successfully induced all groups to choose for the implementation of sanctioning mechanism in order to keep the same design all through the 25 repetitions. Table 6.2 present the alternative fines and corresponding fees. The fee/fine ration ranged from .25 to .50. After subjects read the announcement, questions regarding the implementation of the procedure were answered and play resumed as indicated in Figure 6.1.

Table 6.2 – CPR experiment – sanctioning condition: alternative fines and corresponding fees

Fine (R$)	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80
Fee (R$)	0,05	0,08	0,10	0,12	0,14	0,16	0,18	0,20

6.2.2. The public good experiment

The public good experiment is based on Poppe and Zwikker (1996) design. In Poppe (2005) the approach was meant to show that not only the decision structure but also the content or context framing the social dilemma is important in determining behavior. In order to investigate this issue, Poppe correlated behavior in the experiment with factors scores describing contextually different social dilemmas. His results showed that the number of times the subjects made (monetary) contributions in the public good dilemma correlated significantly only with the scores on the factor with situations in which money was at stake (POPPE, 2005).

Poppe and Zwikker (1996) and Poppe (2005) design involves a strategy for revealing the subjects' motive underneath their decisions whether to contribute or not for the provision of the public good. We then followed the authors' approach aiming to investigate the association between both the frequency of subjects' contributions and the motivations beneath the alternative decisions with our data describing the psychosocial systems of our subjects.

6.2.2.1. Participants and experimental setting

A total 127 subjects (62 female and 65 male), mostly undergraduate students in different programs at the Federal University of Viçosa participated in this experiment. From these, only 4 had not participated in the previous CPR experiment. There is one missing case concerning the psychosocial profile.

Participants were told that payment depended on their own choices and those of the other participants in the experiment. However, because not all participants participated at the same time, it was not possible to pay them immediately after each session, as we did in the CPR experiment. We scheduled a reunion for paying personally those who were able to attend it and let envelops in the department attending desk containing the cash and a summary of the

experimental results for those who could not or not wanted to be present at the meeting. All participants agreed with this procedure.

6.2.2.2. Experiment summary

The experiment was a step-level give some dilemma (POPPE; ZWIKKER, 1996). Enough participants had to contribute in order to obtain outcomes for all. Discussion between the participants was not allowed. Participants were told (and this was actually the case) that 172 people had answered our call and were scheduled for taking part in this new phase of the study. However, it was said that, owing to normal absences, we could expect about 120 to 140 subjects effectively participating in the experimental sections and thus making decisions that would impact the provision of the "public good".

Following Poppe (2005), the experiment consisted of nine trials. In each of the trials every participant had 0.50 reais (as if US$ 0.25) at his or her disposal. The participants were asked to decide whether they would keep the 50 cents or would contribute them to a common pool. (''I do/do not contribute''). If at least X percent of the participants contributed the 50 cents to the pool, all participants – those who had not contributed included – would receive 1.00 real (R$). If a participant contributed him- or herself and not enough others contributed, s/he lost his or her contribution. So, in each trial the participant could get 0 cents (s/he contributed but not enough others contributed); 50 cents (s/he did not contribute and not enough others contributed); 1.00 real (s/he contributed and enough others contributed) or 1.50 real (s/he did not contribute but enough others contributed).

The percentage (X) of the total participants who had to contribute in order to receive 50 cents from the pool varied. In the sequence of the nine trials the percentages were 40%, 70%, 10%, 60%, 90%, 30%, 50%, 80% and 20%, respectively. In order to reveal the motivation underneath participants' decisions, before every trial the participant was asked to indicate which percentage of all participants s/he expected to contribute ("I expect that – percent of all

participants contribute"). To stimulate the participants to mention accurate expectations, we settled that each of the 9 participants with the most accurate expectations would receive a bonus of 5.00 reais (about US$ 2.25). So, in every trial it was established whether the participant did expect that enough others would contribute or did not expect that enough others would contribute and whether the participant did contribute or did not contribute him- or herself.

As Poppe (2005) puts it, when it is known whether a participant did or did not expect that enough participants contributed, and whether the participant did or did not contribute him- or herself, it is possible to determine what motive was underlying the decision. If participants expected enough others to contribute and did not contribute him – or herself, the motive was *greed*. If participants expected enough others to contribute and did contribute him – or herself, the motive was *solidarity*. If participants expected not enough others to contribute and did not contribute him – or herself, the motive was *fear* (of incurring a loss or being gypped). If participants expected not enough others to contribute and did contribute him – or herself, the motive was a sense of *duty* (Poppe and Zwikker, 1996). For each motive a dichotomous score reflected the presence or absence of that motive in a trial.

The nine trials were divided into three contribution levels: Low Contribution (the trials in which it was required that 10%, 20% or 30% of the participants contributed in order to obtain the 50 cents), Medium Contribution (40%, 50% or 60% required) and High Contribution (70%, 80% or 90% required). For each participant and for each Contribution level the percentage of the three trials was computed in which the participant expected that enough others would contribute, contributed him- or herself and showed *greed*, *fear*, *solidarity* and a sense of *duty*.

6.2.3. The ultimatum game experiment

Our last behavioral observations were produced in the context of the well-know Ultimatum game. The first Ultimatum game was developed is 1982 as

a stylized representation of negotiation, by Güth et al. (1982). It has since become the most popular of the standard experiments in economics, and is said to be "catching up with the Prisoner's dilemma as a prime show-piece of apparently irrational behavior" (NOWAK et al., 2000).

The experiment was conducted immediately following the Public Good game as an attempt to take advantage of having our participants already gathered in the experimental site in order to produce some additional data. We regarded that the decision problem settled in the Ultimatum game, in which two parties interact anonymously and only once so that reciprocation is not an issue, may possibly add some new insights or at least reinforce our understanding of the associations between the psychosocial levels of existence and the behavioral responses produced in the other two experimental situations.

6.2.3.1. Experiment summary and equilibrium analysis

In the Ultimatum Game, two players are offered a chance to win a certain sum of money, say, $100. All they must do is divide it. The first player (i.e. the *proposer*) suggests how to split the sum. The second player (the *responder*) can either accept or reject the deal. If the deal is accepted it is the money is divided as proposed by the first player. If the deal is rejected, neither player gets anything. Figure 6.2 present the Ultimatum game in the extensive form.

The "rational" solution (RUBENSTEIN, 1982), suggested by game theory, is for the *proposer* to offer the smallest possible share and for the responder to accept it. For illustration, we will suppose there is a smallest division of the good available (say 1 cent). Suppose that the total amount of money available is x.

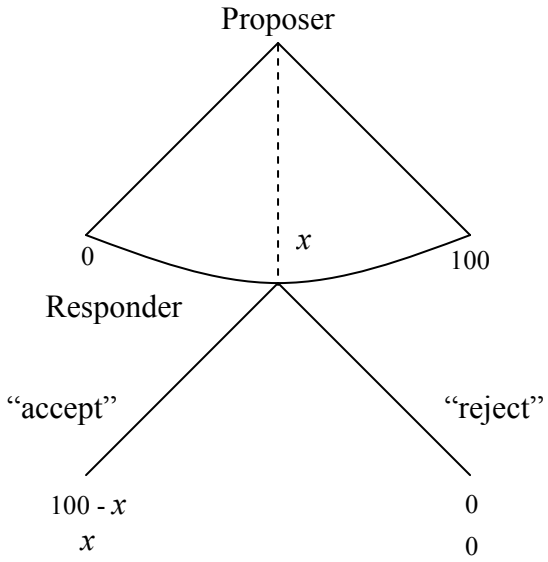


Figure 6.2 – Ultimatum game.

The first player chooses some amount in the interval [1,*x*]. The second player chooses some function *f*: [0, *x*] → {"accept", "reject"} (i.e. the second chooses which divisions to accept and which to reject). We will represent the strategy profile as (*p*, *f*), where *p* is the proposal and *f* is the function. If *f*(*p*) = “accept” the first receives *p* and the second *x*-*p*, otherwise both get zero. (*p*, *f*) is a Nash equilibrium of the Ultimatum game if *f*(*p*) = “accept” and there is no $y > p$ such that *f*(*y*) = “accept” (i.e. *p* is the largest amount the second will accept). The first player would not want to unilaterally increase her demand since the second will reject any higher demand. The second would not want to reject the demand, since he would then get nothing.

There is one other Nash equilibrium where $p = x$ and *f*(*y*) = “reject” for all $y>0$ (i.e. the second rejects all demands that gives the first any amount at all). Here both players get nothing, but neither could get more by unilaterally changing their strategy.

However, only one of these Nash equilibria satisfies the more restrictive equilibrium concept of subgame perfection. Suppose that the first demands a

large amount that gives the second some (small) amount of money. By rejecting the demand, the second is choosing nothing rather than something. So, it would be better for the second to choose to accept any demand that gives him any amount whatsoever. If the first knows this, she will give the second the smallest (non-zero) amount possible[54].

6.2.3.2. Results from prior experimental research

Results from experimental research generally show that if humans play the game the most frequent outcome is a fair share. Cross-cultural studies (e.g HENRICH, 2000; HENRICH et al., 2004) showed that, in many cultures, people offer "fair" (e.g., 50:50) splits, and offers of less than 20% are often rejected. For example, researchers have found that Mongolian proposers tend to offer even splits despite knowing that very unequal splits are almost always accepted. Similar results from other small-scale societies players have led researchers to conclude that "reputation" is seen as more important than any economic reward. Another way of integrating the conclusion with utility maximization is some form of "inequality aversion" model (preference for fairness).

An explanation which was originally quite popular was the "learning" model, in which it was hypothesized that proposers' offers would decay towards the sub game perfect Nash Equilibrium (almost zero) as they mastered the strategy of the game. (This decay tends to be seen in other iterated games). However, this explanation (bounded rationality) is less commonly offered now, in light of empirical evidence against it (see e.g. Brenner and Vriend forthcoming).

[54] Technically, making a zero offer to the responder, and accepting this offer is also a Nash Equilibrium, as the responder's threat to reject the offer is no longer credible since they now gain nothing (materially) by refusing the zero amount offered. Normally, when a player is indifferent between various strategies the principle in Game Theory is that the strategy with an outcome which is Pareto optimally better for the other players is chosen (as a sort of tie-breaker to create a unique NE). However, it is generally assumed that this principle should not apply to an Ultimatum game player offered nothing; he is instead assumed to reject the offer although accepting it would be an equally subgame perfect NE.

It has been hypothesised (e.g. by James Surowiecki) that very unequal allocations are rejected only because the absolute amount of the offer is low. The concept here is that if the amount to be split were ten million dollars a 90:10 split would probably be accepted rather than spurning a million dollar offer. Essentially, this explanation says that the absolute amount of the endowment is not significant enough to produce strategically optimal behavior. However, many experiments have been performed where the amount offered was substantial: studies by Cameron (1999) and Hoffman et al. (1996). Have found that the higher the stakes are the *closer* offers approach an even split, even in a 100 USD game played in Indonesia, where average 1995 per-capita income was 670 USD. Rejections are reportedly independent of the stakes as this level, with 30 USD offers being turned down in Indonesia, as in the United States, even though this equates to two week's wages in Indonesia (CAMERON, 1999).

These results (along with similar results in the Dictator game) are taken to be evidence against the "rational choice" model of individual decisions. Since an individual who rejects a positive offer is choosing to get nothing rather than something, individuals must not be acting solely to maximize their economic gain. Several attempts to explain this behavior are available. Some authors (see e.g. BOLTON, 1991 and OCHS; ROTH, 1989) suggest that individuals are maximizing their expected utility, but money does not translate directly into expected utility.

Another hypothesis is that individuals may get some psychological benefit from engaging in punishment or receive some psychological harm from accepting a low offer. Based on fMRI studies of the brain during decision-making, different brain regions activate dependent upon whether the participating subject "accepts" or "rejects" an offer (see e.g. SANFEY et al., 2003). Since to "rejects" means that neither receives any money, the responder is actually "punishing" the player who makes a low offer. Punishing activates the part of the brain that is associated with the dopamine pathway – i.e. it provides pleasure to punish. Hence, the subjects who refuse and punish in the process, possibly receive more pleasure from punishment than they would from accepting a low

offer. This is, therefore, an expected utility argument where the currency is in pleasures received rather than goods or their associated values in money.

Clearly, our argument here is that both individual and cultural dimension associated to decision-making in the Ultimatum game correlates in predictable ways with stages of adult psychosocial development proposed by Clare Graves. Thus, let us see how we intend to test this.

6.3. Statistical procedures for hypothesis testing

6.3.1. Factor analysis

The first step in the statistical procedures for testing our hypothesis concerning the behavioral responses of the participants in the experimental situations was the submission of the results from Hurlbut/Spiral Dynamics' Levels of Existence test to the procedure of factor analysis (method of principal components).

Factor analysis attempts to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. There are two fundamentally different approaches to factor analysis (STEVENS, 1996). The first is called *exploratory factor analysis*, where the researcher is attempting to determine if there are some small number of underlining constructs which might account for the main sources of variation in a complex set of correlations. Exploratory factor analysis is often used as a variable reduction scheme that will indicate how the variables cluster or hang together. This kind of application can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis (for example, to identify collinearity prior to performing a linear regression analysis).

The other approach, called *confirmatory factor analysis*, rests on a solid theoretical or empirical base. In this case, the researcher "knows" how many factors there are and whether the factors should be correlated or not. Also, the

research generally forces items to load only on a specific factor and whishes to "confirm" his hypothesized factor structure with data.

In the current application, the factor analysis has a quite *confirmatory character*, for Graves's theory imparts the restrictions as to how the psychosocial value systems or levels of existence must combine in order to produce meaningful factors. Indeed, as we saw, given the actual weight of multiple systems within the overall personality of any particular individual at a given point of his or her existence, meaningful linear combinations of systems, as produced through the factor analysis, configure possible be best approach to the theoretical concept of *center of gravity*, defining the principal levels of existence present the sample. Furthermore, since this is the first time that the Spiral Dynamic survey is being applied to a sample of Brazilian subjects, and cultural factors may influence respondents' perception of the representative statements, there is absolutely no promise that we shall obtain theoretically sound factors. In this sense, I claim that the extraction of meaningful constructs is by itself an indication in support of Graves's ECLET conception.

Yet, although linear factor analysis is often used with dichotomous, ordinal, and other types of discrete variables, it is meant for continuums observed indicators. As explained, our data set consists of discrete frequencies of responses associated with both the acceptation and the rejection of each of the six selected psychosocial value systems. Moreover, the cyclical turn of sacrifice-self and express-self systems clearly impart a non linear component into the correlation pattern. Magidson and Vermunt (2001) presented a nonlinear factor-analytic model based on latent class analysis that is especially suited for dealing with categorical indicators, such as dichotomous, ordinal, and nominal variables, and counts. The approach is called latent class factor analysis (LCFA) because it combines elements from latent class (LC) and traditional factor analysis (FA). A disadvantage of the LCFA model is, however, that its parameters may be somewhat more difficult to interpret than the typical factor-analytic coefficients.

In order to overcome this problem, Vermunt and Magidson (2004) propose using a linear approximation of the maximum likelihood estimates

obtained with a LCFA model. This makes it possible to provide the same type of output measures as in standard FA, while retaining the fact that the underlying factor structure is identified by the more reliable nonlinear factor-analytic model. Yet, the author's results from three example applications where the assumptions of standard factor analysis were violated suggested that it cannot be used at all only with nominal categorical indicators. Overall, the results indicated that improved interpretations from the LCFA approach ensue especially in cases where the nonlinear terms represent a significant source of variation.

As a reasonable rule of thumb, most statistical packages manuals indicate that data for which Pearson correlation coefficients can sensibly be calculated should be suitable for factor analysis (SPSS 1990). Further inspections of the suitability of the data set for submitting to factor analysis can be accomplished be means of Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity. The KMO measure of sampling adequacy tests whether the partial correlations among variables are small. Bartlett's test of sphericity tests whether the correlation matrix is an identity matrix, which would indicate that the factor model is inappropriate.

The assumptions underlining factor analysis are that data should have a bivariate normal distribution for each pair of variables, and observations should be independent. The factor analysis model specifies that variables are determined by common factors (the factors estimated by the model) and unique factors (which do not overlap between observed variables); the computed estimates are based on the assumption that all unique factors are uncorrelated with each other and with the common factors.

Formally, the relationship between a variable X_i and the factors is given by

$$X_i = A_{i1} F_1 + A_{i2} F_2 + ... + A_{ik} F_k + U_i \qquad (6.11)$$

where the F's are the *common factors*, provided the variable-component correlations; U_i is the unique factor, representing the non-explained variance; and

the A's are the constants in the linear combination or the *factor loadings*. The factor loadings squared sum produces the *communalities* or the variance explained by the common factors.

As we said, the common factors underlining the correlations among the Graves's constructs are precisely the levels or waves of consciousness, since the concept involves a patterned interplay among the various stages, not a discrete centralization into a unique stage. Accordingly, as we said, the extraction of meaningful factors is significant result in support of Graves's conception.

The basic steps in the factor analysis procedure are: (1) computation of the either the correlation or the covariance matrix; (2) extraction of the common factors; (3) factor rotation; and (4) computation of the individual (cases) factor scores.

In general, social sciences researchers would be used to extracting components from a correlation matrix. The reason for this standardization in that scales for tests used in educational, sociological, and psychological research are usually arbitrary. Yet, Morrison (1976, p. 222) explains that if the scales of the variables are reasonably commensurable, performing a component analysis on the covariance matrix is preferable for statistical reasons. Since the scale of the SD test is just the discrete frequency of "most like me" and "least like me" responses, bounded by the total number (40) of questions in the test, we used the covariance matrix. The method of factor extraction was the *principal component*. This is the most frequently used extraction method and a necessary one in our case since it can be applied when the correlation matrix is singular. The method is used to form uncorrelated linear combinations of the observed variables. The first component has maximum variance. Successive components explain progressively smaller portions of the variance and are all uncorrelated with each other. Principal components analysis is used to obtain the initial factor solution.

As for how many components to retain, probably the most widely used criterion is the Kaiser's eigenvalue-greater-than-one rule. This criterion is suitable when the factors are extracted from the correlation matrix, though. An alternative rule when the factors are obtained from the variance-covariance

matrix is to retain as many factors as the cumulated explained variance remains lower than the sum of the principal diagonal (i.e. the sum of the variances). Also, according to Hakstian *et al* (1982), a ratio < .30 between the number of factors and the number of variables (12 in this case) supply a much more credible estimate for the number of true factors. Nonetheless, these and other statistical criteria are only assistance tools that should not replace researchers critical evaluation regarding the number of components that would be appropriate if factors are to be considered as real influences rather than merely statistical derivatives. As pointed above, in a confirmatory application the key input criteria ought to ensue from the theoretical base.

Accordingly, the nested structure of Graves's scheme tells us that the six selected psychosocial systems (from BO to GT) are certainly not independent from each other. On this basis, we should not expect to find six distinct components, one for each system, clustering together only the contrasting variables denoting either the acceptance or the rejection of either system. Of course, we cannot predict exactly how the variables shall hang together. Yet, they must cluster in such a way that either the *sacrifices-self* systems are discriminated from the *express-self* systems or else the systems of each type appear negatively correlated into a same component. In the first case, we would have only two factors. Otherwise, we can obtain from one to three factors. Evidently that if all variables cluster together, it will be harder to come with an insightful interpretation. In any case, in theory, the sacrifice-self systems must correlate negatively with the express-self systems.

In order to facilitate interpretation of the factors and avoid multicollinearity in the regressions analysis we applied orthogonal rotation using the Varimax method with Kaiser Normalization. It should be mentioned that when the Varimax rotation is done the *maximum variance property* of the original components is destroyed. The rotation essentially reallocates the factor loadings and, thus, the first rotated factor will no longer *necessarily* account for the maximum amount of variance.

Finally, the factor loading are obtained by the multiplication of the normalized values of the variable *i* by the corresponding factor score coefficient. The general expression for the estimation of the factor *j* (F_j) is given by:

$$F_j = \sum_{i=1}^{n} W_{ji} X_i = W_{j1} X_1 + W_{j2} X_2 + ... + W_{jp} X_p \qquad (6.12)$$

where W_{ji} are the factor scores coefficients and *p* indicates the number of variables.

A final issue regarding the factor analysis concerns the stability of the solution. We tested this stability by repeating the factor analysis using only the cases of the subjects who actually took part in each experiment (n = 196 in the CPR game, n = 126 in the Public Good and Ultimatum games), as well as using the entire sample of SD respondents (n = 322). The results, in the Appendix, showed that while the relative importance of each component changes, the interpretation of the factors themselves remains the same, so that the solution was found to be stable. Therefore, since by using the entire sample we are able to produce *factor loadings* that are statistically more significant and *factor scores* that are more representative of the population or sub-populations from which our participants came from, we decided for using the entire ample in order to obtain the factor scores used in the hypothesis tests. This gives us more confidence to test the statistical relationship between the factors scores and the behaviors of the individuals who actually participated in the experiments.

6.3.2. Statistical tests

The test of the expected behavioral responses in the experimental situations relied basically on bivariate (Pearson correlation) and multivariate (linear regression and binary logistic regression) analyses.

The Pearson correlation analysis was used mainly to explore the pairwise relationship between the frequencies of each separate SD response and each

(quantitative) behavioral response observed in the experiments. As the linear regression models, Pearson's correlation coefficient is a measure of linear association. Two variables can be perfectly related, but if the relationship is not linear, Pearson's correlation coefficient is not an appropriate statistic for measuring their association. Also, Pearson's correlation coefficient assumes that each pair of variables is bivariate normal.

The statistical tests concerning the relationships among the SD variables themselves often rejected the normality distribution assumption for our variables. This is, nonetheless, a likely result in small samples in view of the complex nature of the phenomenon under analysis (i.e. the fluid and flowing pattern of consciousness development, with its spirals, swirls, streams and waves). One may expect that the levels of existence are asymptotically normally distributed. Notwithstanding, while the correlations among those variables run from small to moderate, the number of statistically significant correlations and the meaningful pattern produced in the factor analysis indicate that the found associations are substantive. As for the correlation of the SD variables with the behavioral responses, most of the psychometric data produced by Graves, which bear resemblance with behavior with the *agency* versus *communion* dilemma, present cyclical or cyclical trend patterns along the developmental sequence. This fact weakens the statistical significance that may be found by means of liner models. Still, results from nonlinear regression models are valid only if one has specified a function that *accurately* describes the relationship between dependent and independent variables. Additionally, the choice of good starting values is very important. Even if one has specified the correct functional form of the model, the use poor starting values may cause the model fail to converge, or one may get a locally optimal solution rather than one that is globally optimal. We thus have not explored nonlinear estimation but have simply relied on the linear multivariate regressions as a way to submit the associations to more stringent test of the null hypothesis.

Moreover, in the multivariate regressions we used the three factors (principal component) as predictors of the behavioral responses in the

experimental situations, thus preserving the suggested causal relationship, even if the models are not intended to supply reliable quantitative predictions. The same idea holds for the binary logistic regressions applied to examine the relations between the factor scores and the dichotomic responses such are if the individual was or not a defector. As to the relationship between the motives underlining decisions in the provision dilemma, multinomial logistic regressions did not work well enough. We thus have limited ourselves to the multivariate regression using the frequency of motives as predictors of the factor scores.

Finally, the effects of the institutional settings (communication and sanctioning) on the economic performance in the appropriation dilemma, and of the contribution levels (low, medium and high) on the contribution decision in the provision dilemma was tested using ANOVA.

6.3.3. Working hypotheses

A first note regarding the specification of our testable hypotheses is that, at this point, we can do so only with regard to the original variables describing each of the six selected psychosocial systems. The presentation of the hypotheses in connection with the principal components must wait until we have seen the results from the factor analysis.

A second note is that the intrinsic qualities of the ECLET scheme, as a nested structure of transcend-and-include, underlined by a cyclical turns of the agency and communion capacities of the self, challenged considerably our indent of submitting clear-cut behavioral predictions, capable of exposing the distinctive qualities and implications of the Gravesian systems, within quite limited alternative actions open by the experimental conditions.

Notwithstanding, we should remind that the key implications we are attempting to expose are:

1. That different attitude before the collective action dilemmas are not merely a matter of dissimilar individual “preferences” or idiosyncratic traits of character, but are instead connected with a quite regular and predictable

pattern of psychosocial development of our species, given the factors of history.

2. That the institutional means of dealing with collective action problems interact with such a pattern, in likewise predictable ways, so that the finding of the stage developmental theory must be taken into account when conceiving such institutional means of intervention.
3. That the interior development has its own inner impulse and configure a separate factor to be considered when devising institutional solution for social dilemmas, as it is marked by an inherent process of *decentering of the self* – out of the awakening of consciousness to increasing encompassing realities – which empowers the capacity to be trustworthy, exercise free of charge self-restraint, and willingly cooperate for the collective good.

As quick reference for checking the consistence of our hypotheses with the characteristics of Graves's constructs, Table 6.3 presents the major themes of each of the selected waves of existence.

Table 6.3 – Cyclical aspect, way of thinking and themes of the selected Gravesian stages or waves of existence

Stage/wave	Color code	Cyclical aspect	Thinking	Theme
G-T	**Yellow**	Express-self (agency)	Ecological	*'Express self for what self desires, but never at the expenses of others and in a manner that all life, not just my life, will profit'*
F-S	**Green**	Sacrifice-self (communion)	Consensus	*'Sacrifice now in order for all to get now'*
E-R	**Orange**	Express-self (agency)	Strategic	*'Express self for what self desires, but in a fashion calculated not to bring down the wrath of others'*
D-Q	**Blue**	Sacrifice-self (communion)	Authority	*'Sacrifice self now to receive reward later'*
C-P	**Red**	Express-self (agency)	Egocentric	*'Express self, to hell with others and the consequences, lest one suffer the torment of unbearable shame'*
B-O	**Purple**	Sacrifice-self (communion)	Animistic	*'Sacrifice self to the way of your elders'*

6.3.3.1. The levels of existence within the appropriation dilemma

We remind that the appropriation dilemma settled by the CPR experiment involves three distinct institutional settings or conditions: unregulated (or "open access"), costless communication and costly sanctioning.

As suggested since the Introduction, the assumptions underlining the "rational actor" model are suitable precisely in the extent that the scope of analysis is limited to market interactions, so that coordination failures other than those owing to "externalities" are not a concern. So, in the unregulated setting (baseline), the Nash equilibrium may be indeed a relevant prediction for aggregate behavior, at least in the measure that participants are well-established at the ER-level (strategic thinking) or above. Still, more relevant than the Nash equilibrium itself (i.e. precisely 64 tokens invested in the Market 2), is the weaker prediction of inefficient appropriation from the CPR due to overinvestment, without requiring strategic thinking but merely a go-getting trend. Such a behavior (go-getting) is an expression of the agency capacity of the self, and thus we expect to find it correlating positively with the express-self systems. We thus propose the following hypothesis (H1).

H1-CPR: Overinvestment in the unregulated conditions is higher for the express-self systems (CP, ER, and GT) than for the sacrifice-self systems (BO, DQ, and FS).

As for the nonbinding communication condition, we consistently expect that the cooperative behavior is favored by the sacrifice-self stance while hazarded by the express-self waves. Still, this general expectation requires a qualification concerning the GT stage or wave of consciousness, for at GT the individual "*Express self for what self desires, but never at the expenses of others and in a manner that all life, not just my life, will profit"*. So, we expect that GT also cooperates while being an express-self wave of existence. Note that this is a precise instance of what we mean by *lessening of egocentrism* or the *decentering property* of consciousness evolution. Accordingly, for the nonbinding communication condition we raise the following hypothesis (H2).

H2-CPR: The degree of cooperative behavior (limiting investment is the CPR) in the nonbinding communication setting is higher for the sacrifice-self systems (BO, DQ, and FS) and for the express-self systems GT, than for the express-self systems CP and ER.

During the nonbinding communication treatment, however, the cooperative behavior (or at least the economic performance of the collective appropriation strategy) may the challenged not only by cheating (primary defection) but also by secondary (reciprocal) defections as a form of punishment. We shall attempt distinguishing these forms of deviation one from the other by identifying first defections. Also, we shall sort out the "never defect" behavior as a stronger form of supporting the joint strategy. We acknowledge that the "never defect" expression sounds absolutistic (DQ). However, this behavior is expected to be associated not with DQ but with FS. The reason is that, in the experimental situation, the absolutism of DQ is not expected to show up as a forbearing stickiness with the norm but instead as an intolerant reaction to deviations from the norm. Conversely, a forgiven assimilation of cheating and altruistic restraint in observance and support to the collective agreement is a behavior underpinned by higher levels of consciousness development, precisely because it requires that the reproachable conduct be comprehended by transcendence-and-inclusion. Thus, while GT is, according to the ECLET conception, capable of doing so, it is not expected to "never defect" in nonbinding condition if it perceives that agreement has lost the general support. On the ground of such consideration we add the following hypothesis, limited to the nonbinding communication condition:

- **H2.1-CPR**: Defection is higher for the CP, DQ, and ER waves of consciousness, and lower for BO, FS and GT;
- **H2.2-CPR**: Cheating (primary defection) is higher for the CP and ER waves of consciousness, and lower for BO, DQ, FS and GT, and;
- **H2.3-CPR**: The "never defect" behavior is higher for FS wave of consciousness, and lower for the others.

Finally, in the costly sanctioning treatment we have behavioral responses relating to both the appropriation strategy and the use of the sanctioning mechanism. Concerning the first, the general expectation is that the sanctioning mechanism suffices to correct norm-deviant behavior initiated by either CP or ER. That is:

H3.1-CPR: The CP and ER systems are expected to reduce investment in the sanctioning condition.

Still, Graves's characterization of the CP wave allows the conjecture that CP may react negatively to the imposition of sanctioning mechanism, for one of the defining traits of CP is be defiant to authority. As we said, given that the cash rewards should dominate we expect CP can be put in line, even if it is also expected at least to give it a try. Likewise, since the number of rounds is common knowledge, last-round strategic defections are expected, particularly for ER. We have not separated last-round defection from defections occurring in other rounds of the sanctioning condition, though. In fact, the noncooperative behavior in the sanctioning condition was approached through the total expenditures with fines, not by the frequency of defections. In such a base, we posit the following conservative hypothesis.

H3.2-CPR: The CP and ER systems are expected to pay more fines, while BO, DQ, FS and GT pay less.

As for the use of the sanctioning mechanism, measured by the total expenses with fees, some distinctions are in order. First, as observed a propos of the punishing behavior in the Ultimatum game, the use of the costly sanctioning mechanism in the CPR experiment cannot be unambiguously interpreted as evidence of strongly reciprocity (i.e. the willing to incur into personal cost for the collective good). The use of the sanctioning mechanism *as evidence of strong reciprocity* is theoretically amenable to the characteristics of BO, DQ, FS and GT waves of consciousness. Still, Graves's theory indicates that there are subtler issues involved. For one side, BO and DQ may attribute some reward value to the act of punishing itself, so that the cost of punishing ought to be somewhat discounted. Conversely, while FS is expected to behave as strong reciprocators, it

is not expected to do so by means of punishment, that is, by imposing a lost on others, owing to FS high sensitivity and preference for reconciliation and consensus. Accordingly, the cost of punishing is possibly subjectively higher for FS. Finally, the idea of imposing lost on others is not a pleasant one at the GT level of consciousness as well. Still, at GT, "functionality have the highest priority" (WILBER, 2001, p. 12), so that we do expect to find GT paying more fees in order to support the collective efficiency associated with constraint imposed by the social norm.

As for CP and ER, conversely, we do not expect strong reciprocity. This is unambiguous in the case of ER, who would enjoy to "free-ride" in this second-order dilemma just like in the primary one. For CP, on the other hand, while spending money aiming to reinforce the social norm is definitely not a kind of behavior we might expect from them, the act of imposing lost on others is likely to present rewarding value at this level of psychological development. So, like in the case of BO and especially DQ, the cost of punishing may appear somewhat discounted from the CP point of view. In average, however, we do not expect that perverse punishment can lead CP to spend more with fees. So, bearing such consideration in mind, we raise the following and final hypothesis concerning the behavioral responses in the CPR experiment:

H4-CPR: The BO, DQ and GT waves are expected to expend more with fees, while CP, ER and FS are expected to expend less.

6.3.3.2. The levels of existence within the provision dilemma

The behavioral responses produced by means of the Public Good experiment includes subjects' expectations as to the contribution of others and the actual decision whether to contribute themselves for the common-pool. As we saw, by comparing expectations and actual choices, Poppe and Zwikker (1996) suggested four motives underling the participants' decisions: either "fear" or "greed", for the noncontributors, and either "solidarity" or sense of "duty", for the contributors. As a within-subject factor affecting decision the authors

defined three levels of required contribution for the provision of public good: low (10-30%), medium (40-60%) and high (70-90%).

Concerning the expectations, a theoretical observation is in order. The ECLET conception suggests that individuals centered at a given wave of consciousness cannot form expectations as to the behaviors that are still not open to them. Accordingly, for instance, BO, CP and DQ are presumed to do not form *strategic expectations* of the kind open at ER. Of course, individuals at the various waves of consciousness do form expectations as to the behaviors of other individuals that may be both at lower and higher waves of existence. Still, and typically, all thorough the six *subsistence levels* of existence levels (first-tier) these views are not accomplished *from* the perspective of the other waves. Instead, they are distorted by the prejudice that the worldviews and values of others are worse and incorrect. As Graves, Beck, and Cowan, quoted by Wilber (2001, p. 14) point out, "without second-tier thinking humanity is destined to remain victim of a global 'autoimmune disease', where the various vMemes turn on each other in an attempt to establish supremacy".

I elaborate on this topic to justify that the various waves of existence can form similar expectations as to the behavior of others, even if founded in diverse reasoning. For example, taking the sequence of the nine trial s it is fairy reasonable that all value systems expect that contributions will fall as the requisite contribution level increases. However, the way the expectations are formed will depend on how each level imagines how similar or diverse the other people are from themselves. If one regards that population is manly "most like me" he or she will form a kind of "self-mirroring" expectation, either cooperative (sacrifice-self systems plus GT) or aiming to free-ride (CP and ER). If one regards that population is manly "least like me" the expectation will be reversed. Even if this reasoning might impart no more than flimsy speculation, I just assume that BO, DQ, FS and GT think that the general population is not much like themselves, so that this led them the expect contributions will fall. Conversely, ER tend to see the world as a game and all others as rational players.

Since ER is assumed to be unwilling to contribute, it also expects contributions will fall.

Now, on the *average* if the nine trials, the following hypothesis is the best we could possible advance:

H1-PG: If some significant *positive* correlation between the expectations as to the contributions of others and the values systems exists, it might by with the FS value system.

Regarding the frequency of actual contribution on the average of the nine trials we expect:

H2-PG. CP and ER are expected to contribute less frequently, and the sacrifice-self systems (BO, DQ, and FS) plus GT are expected to contribute more frequently.

As for the interactions with the contribution levels, the only interesting conjectures are grounded on what we can advance as to the underlining motivations of the various value systems. So, beginning with the motivations, we posit that under the decision to cooperate we may possible find that:

H3.1-PG. BO, DQ, and FS cooperate out of *solidarity* and sense of *duty*; ER and GT cooperate out of *solidarity* but not sense of *duty*; CP does not cooperate.

As to the decision to do not contribute, we posit:

H3.2-PG. When BO, DQ, FS and GT do not cooperate, the underlining motive is *fear* bur not *greed*; the motives under CP and ER noncooperative decision are both *fear* and *greed*.

So, out of the above hypotheses we can add that:

H4-PG. BO, DQ and FS contribute more at the high contribution level (sense of *duty*) and CP and ER contribute less at the low contribution level (*greed*).

We don't hypothesize on GT in this regard because we don't think GT can be induced to expresses sense of duty in the experimental situation (hypothesis **H3.1-PG**). Also, the motives *fear* and *solidarity* present no particular

association with the extreme contribution levels, so that we don't regard them as helpful basis for discriminate behavioral interaction with them.

6.3.3.3. The levels of existence and the ultimatum game

Finally, the Ultimatum game produced two additional behavioral observations, the level of the offer and the decision whether to accept or reject the offer. As mentioned, reciprocation is not an issue in the Ultimatum game. Still, on the part of the proposer's decision, and beyond the "rational egoist" assumptions, when the proposer makes a high offer it is involved either a taste for fairness, fear of rejection (risk aversion), or both. On the part of the responder, on the other hand, the rejection of a low offer similarly involves a taste for fairness, but now it must also involve a taste for punishing, as we saw.

On this ground we set forth the following hypotheses:

H1-UG. Lower offerings are proposed by CP and ER, while higher offerings are proposed by BO, DQ, FS and GT.

H2-UG. The decision to rejections a low offering is associated to the CP and DQ waves or consciousness.

6.4. Concluding remarks

In closing, I would like to remark that, despite the difficulties and limitations inherent to the complex and flowing inner phenomenon under examination we were able to set forth a number of quite clear theoretical and testable expectations concerning the behavioral responses associated with Graves's constructs, in three well established experimental protocols.

Yet, as the very construction of the hypotheses indicates, the ultimate test of our central thesis should not be evaluated with reference to any single relationship between the psychological value systems and the behavioral responses in any experiment. The subtleness of the psychological dynamics indicates that significant statistical relationships might be rather difficult to

establish. In this vein, I claim that our results, to which we now turn, ought to be analyzed in conjunct, where the strength of the ECLET conception materialize as an overall consistence of the significant relationships scattered amid the sum total of our evidences produced along the three experiments taken together.

7. RESULTS AND DISCUSSION

7.1. Some patterns in the great river of existence

As underlined throughout the theoretical chapters, consciousness development is not a linear ladder but a fluid and flowing affair, approaching what appear to be an almost infinite number of multiple modalities. Accordingly, what Clare Graves and Spiral Dynamics offer us is a (third-person) description or reconstruction of a flowing psychological (first-person) reality, as it appears when scrutinized from a particular angle of observation-namely, that of the psychosocial value systems stream.

Therefore, owing to the fluid and flowing nature of the interior phenomenon, the contours of the psychosocial waves of existence is rather difficult to devise clearly in any particular individual. In any given moment, the majority of the individuals present patterns which expose multiple combinations of systems, and just a few of them exhibit clear indications of being centralized into a particular level of existence. As we saw in the section about adequate structuralism, it is mostly in the conjunct of large groups of individuals that such wavering patterns can be more clearly discerned.

The careful identification of the center of gravity of an individual involves a further consistence check of her profile with a set of open questions

presented at the beginning of the SD survey, plus the completion of the Change State Indicator test, in order to identify the change state phase (see Figure 5.6) the individual is passing through at the time of the test. Furthermore, a very important issue not yet systematically addressed concerns the distinction between closed and open systems. Plausibly, an otherwise similar profile may produce contrasting behavioral responses within a particular situation depending on its degree of openness to behavioral change.

We now need to face such difficulties and limitations in order to describe the characteristics of our sample with regard to the Spiral Dynamics profiles. As we saw, each individual SD profile contains 12 dimensions, since each of the six tested psychosocial systems (from BO to GT) figures as both "most like me" (acceptance/identification) and "least like me" (rejection/disidentification) responses. The simplest (and acknowledged simplistic) way of approximating the psychosocial center of gravity of our 322 respondents, without scrutinizing each profile and open questions individually, is by calculating their "net" results (i.e. the difference between the "most like me" and the "least like me" frequencies of responses for each psychosocial system) and then looking for the highest positive difference. We can thus produce a tentative image of the distribution of the psychosocial value systems present in our sample, at least as the "net accepted" representative statements.

By following this procedure we get to the results shown in Figure 7.1. From this view, our sample indicates a move toward the higher levels of existence, since the FS (green) and GT (yellow) values systems present the highest frequencies of "net acceptance". The wave-like move toward GT is indicated by the adjusted trend line (the negative range between BO and CP has no meaning, and no implication for our illustrative purpose).

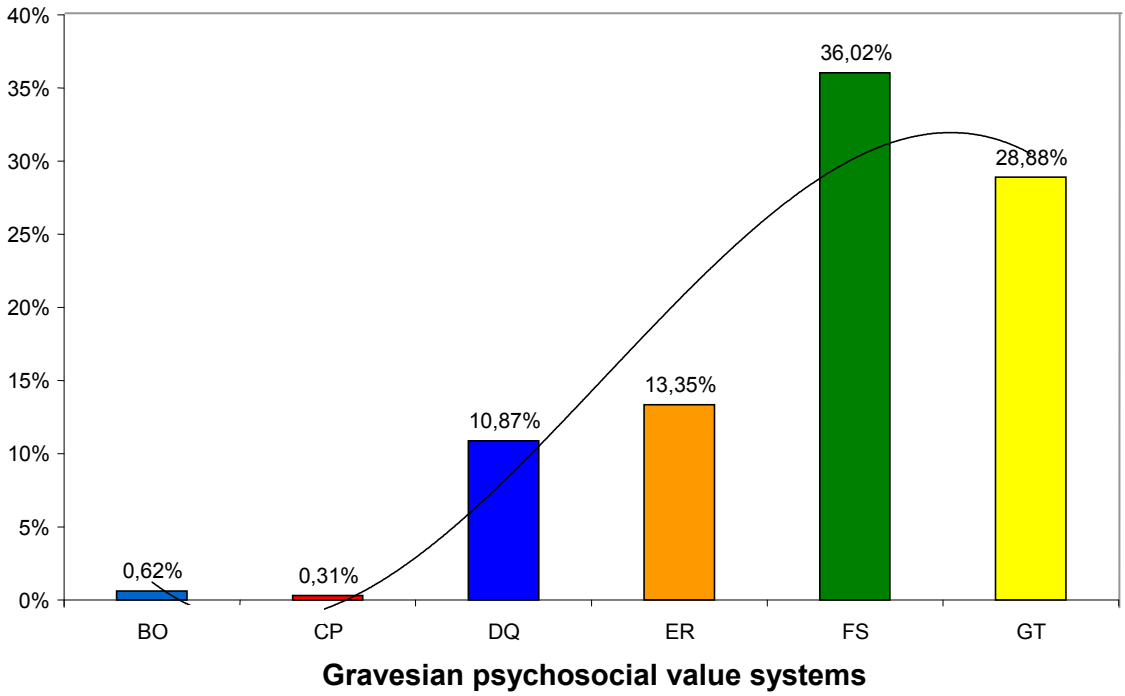


Figure 7.1 – Distribution of the individuals' centers of gravity, as the highest difference between the "most like me" and "least like me" frequencies of responses.

As far as our procedure in good enough, this result indeed approximates the expectations. As Hurlbut (1979, p. 213-214) pointed out, "In the early 1950s, Graves found 34 percent of his students operating at the fourth [DQ or blue] level, but felt that the modal level was the fifth [ER or orange]. In the 1960s, he talked about the sixth [FS or green] level being very visible in American society". Also, "It was in 1959 that Graves says he got the 'shock of his life' because he first began to encounter persons who had grown beyond the seventh [GT or yellow] level. Through the 1960s and into the 1970s, Graves still felt that hat not encountered enough of these people to be able to comprehensively describe this level". Hurlbut reports that, on the tests given in her own study, the seventh [GT] level was the modal primary level for nine of a total of eleven groups. According to her, "All of this indicates upward movement and leads one to conclude that there must be an eight [HU or turquoise] system becoming more and more visible". In fact, she concludes that, "Since such a high percent of

participants in this [her] study indicated their primary level of existence in the seventh [GT] level, some of them should, theoretically, be halfway into the eighth [HU] system now. This might have shown up had choices for an eighth system been included on the test instrument".

We should remind, however, that according to Cowan and Todorovic the SD survey still overestimate the representation of the GT value system[55]. As already mentioned when discussing about the assessment tool, this is related with a misclassification of ER, possible due to an aspiration effect that may appear in the last part of the test, when the representative statements of all six systems are presented together. Given this, we regard that Figure 7.1 could approximate better the profile of our sample if corrected in order to indicate at least 15% of ER first "net acceptance". This difference should be taken from the GT yellow bar.

But while the "net acceptance" construct may be passable for the purpose of assigning a roughly characterization of our sample, it is certainly insufficient for exploring the statistical associations between the values systems and the behavioral responses in the experimental settings, because the primary psychosocial system is *not* the only relevant system determining behavior – a proposition which we have repeatedly emphasized. In order to explore such relationships, we must take into account the whole spectrum of responses present in the individual profiles. As we saw, we pursued this by means of the Factor Analysis procedure, to whose results we now turn.

7.2. Three principal waves of existence

Before presenting the results of the Factor Analysis, an additional methodological note is in order, namely, that not all the individuals who filled out the SD survey (n = 322) and are included in the present analysis took part in the laboratory experiments[56]. Yet, by using the entire sample in the Factor

[55] Personal communication during the Spiral Dynamics Workshop in Santa Barbara, CA. (April, 2005).
[56] The participants who actually took part in the experiment were those who confirmed possibility of participation in the established experimental schedule.

Analysis we are able to strengthen the statistical significance of the *factor loadings* and obtain *factor scores* that are more representative of the population or sub-populations from which our participants came from. This bestows more confidence to test the statistical relationship between the factors scores and the behaviors of the individuals who actually participated in the experiments.

To check out the stability of the *factors* we repeated the Factor Analysis using only the data of the subjects who actually took part in each experiment ($\underline{n}$ = 196 in the CPR experiment and $\underline{n}$ = 126 in the PG and UG experiments) [57]. The results (Appendix 3) show that the factors are stable.

As discussed, we applied the Factor Analysis (method of principal components) on the *covariance* matrix. Unfortunately, the covariance matrix (as well as the correlation matrix) is not positive definite, so that it was not possible to accomplish the KMO and Bartlett's test of sphericity[58]. Nonetheless, by inspecting the psychosocial systems correlation matrix we observe that, while the value of the correlations range from weak to moderate, the correlations are in the most part statistically significant (Table 7.1). Also, the correlations run in the expected direction, suggesting that the assessment tool worked well enough and that the Factor Analysis can produce meaningful components, as we shall see in a moment. The communalities amongst the variables are presented in Table 7.2. The only moderate values reflect our commentary relatively to the correlation matrix, and indicate that the factor scores are somewhat scattered around the space defined the principal axis. Once again, this ensues from the intricacy involved in the Gravesian system, and already insinuates that we shall experience some difficulty in establishing strong statistical relationships between the factors scores and the experimental observations.

As for how many components to retain, we already discussed the main methodological issues. In this regard, we just need to reinforce our point relating to the confirmatory character of the present investigation, so that Graves's theory

[57] The actual number of participants in the CPR experiment is $\underline{n}$ = 200 (25 groups with 8 participants each), but we have four missing cases concerning the SD profiles (participants who did not completed the survey). In the PG and UG experiment there is one missing case.

[58] These statistics offer a synthetic indication of the suitability of the data set for submitting to the Factor Analysis procedure.

provides the key basis for deciding about both the number of components and the quality of the results. On this basis, after scrutinizing several outputs, we found that by extracting only three components, explaining 70.15 percent of total (raw) variance in the sample (Table 7.3), we are able to construct more meaningful waves of existence. When needed, we will disaggregate the factors and analyze the relationships of the behavioral responses with the separate psychosocial systems.

As explained, we applied orthogonal rotation using the Varimax method with Kaiser Normalization, in order to facilitate interpretation of the factors and avoid multicollinearity in the regressions analysis. To remind, when the Varimax rotation is done the *maximum variance property* of the original components is destroyed. The rotation essentially reallocates the factor loadings and, thus, the first rotated factor will no longer *necessarily* account for the maximum amount of variance. This can be observed in Table 7.3 where, after rotation, the first principal component explains 23.44 percent of the total variance; the second explains 21.15 percent and the third explains 24.56 percent. If the variances are rescaled (lower panel in the Table 7.3) the property is restored. In any case, the three principal components or factors explain quite comparable amounts of the total variance in this sample.

In order to interpret the factors we remind that with a sample of $\underline{n} = 322$ the factor loadings greater than | .30| are statistically significant (CLIFF; HAMBURGER, 1967). These are loadings detached in bold in the Table 7.4. Among them, the strongest variable-component correlations are underlined.

Table 7.1 – Values systems correlation matrix

		BO_most	BO_least	CP_most	CP_least	DQ_most	DQ_least	ER_most	ER_least	FS_most	FS_least	GT_most	GT_least
BO_most	Pearson Correlation	1	-.473**	.165**	-.103*	.175***	.040	-.359**	.298**	-.095*	-.215**	-.329**	.351**
	Sig. (1-tailed)		.000	.001	.032	.001	.236	.000	.000	.044	.000	.000	.000
	N	322	322	322	322	322	322	322	322	322	322	322	322
BO_least	Pearson Correlation	-.473**	1	.061	-.364**	-.249**	.054	.391**	-.532**	-.301**	.045	-.403**	-.394**
	Sig. (1-tailed)	.000		-139	.000	.000	.168	.000	.000	.000	-.208	.000	.000
	N	322	322	322	322	322	322	322	322	322	322	322	322
CP_most	Pearson Correlation	.165**	.061	1	-.575**	-.280**	.465**	.121*	-.137**	-.410**	.391*	-.042	.136**
	Sig. (1-tailed)	.001	.139		.000	.000	.000	.015	.007	.000	.000	.227	.007
	N	322	322	322	322	322	322	322	322	322	322	322	322
CP_least	Pearson Correlation	-.103*	-.364**	-.575**	1	.343**	-.600**	-.209**	.150**	.408**	-.551**	-.197**	-.130**
	Sig. (1-tailed)	.032	.000	.000		.000	.000	.000	.003	.000	.000	.000	.010
	N	322	322	322	322	322	322	322	322	322	322	322	322
DQ_most	Pearson Correlation	.175**	-.249**	-.280**	.343**	1	-.586**	-.286**	.244**	-.018	-.025	-.568**	.353**
	Sig. (1-tailed)	.001	.000	.000	.000		.000	.000	.000	.372	.325	.000	.000
	N	322	322	322	322	322	322	322	322	322	322	322	322
DQ_least	Pearson Correlation	.040	.054	.465**	-.600**	-.586**	1	.143**	-.316**	-.256**	.175**	.410**	-.196**
	Sig. (1-tailed)	.236	.168	.000	.000	.000		.005	.000	.000	.001	.000	.000
	N	322	322	322	322	322	322	322	322	322	322	322	322
ER_most	Pearson Correlation	-.359**	.391**	.121*	-.209**	-.286**	.143**	1	-.567**	-.549**	.311**	.039	-.018
	Sig. (1-tailed)	.000	.000	.015	.000	.000	.005		.000	.000	.000	.243	.371
	N	322	322	322	322	322	322	322	322	322	322	322	322
ER_least	Pearson Correlation	.298**	-.532**	-.137**	.150**	.244**	-.316**	-.567**	1	.416**	-.297**	-.225**	.061
	Sig. (1-tailed)	.000	.000	.007	.003	.000	.000	.000		.000	.000	.000	.137
	N	322	322	322	322	322	322	322	322	322	322	322	322
FS_most	Pearson Correlation	-.095*	-.301**	-.410**	.408**	-.018	-.256**	-.549**	.416**	1	-.505**	-.289**	.030
	Sig. (1-tailed)	.044	.000	.000	.000	.372	.000	.000	.000		.000	.000	.298
	N	322	322	322	322	322	322	322	322	322	322	322	322
FS_least	Pearson Correlation	.215**	.045	.391**	-.551**	-.025	.175**	.311**	-.297&&	-.505**	1	-.043	.181**
	Sig. (1-tailed)	.000	.208	.000	.000	.325	.001	.000	.000	.000		.219	.001
	N	322	322	322	322	322	322	322	322	322	322	322	322
GT_most	Pearson Correlation	-.329**	.403**	-.042	-.197**	-.568**	.410**	.039	-.225**	-.289**	-.043	1	-.599**
	Sig. (1-tailed)	.000	.000	.227	.000	.000	.000	.243	.000	.000	.219		.000
	N	322	322	322	322	322	322	322	322	322	322	322	322
GT_least	Pearson Correlation	.351**	-.394**	.136**	-.130**	.353**	-.196**	-.018	.061	.030	.181**	-.599**	1
	Sig. (1-tailed)	.000	.000	.007	.010	.000	.000	.371	.137	.298	.001	.000	
	N	322	322	322	322	322	322	322	322	322	322	322	322

* Correlation is significant at the 0.05 level (1-tailed).
** Correlation is significant at the 0.01 level (1-tailed).

Table 7.2 – Factor analysis: communalities

	Raw		Rescaled	
	Initial	Extraction	Initial	Extraction
BO_most	4.530	2.111	1.000	0.466
BO_least	13.513	7.996	1.000	0.592
CP_most	5.044	2.713	1.000	0.538
CP_least	19.214	16.360	1.000	0.851
DQ_most	14.136	10.358	1.000	0.733
DQ_least	11.442	8.389	1.000	0.733
ER_most	15.017	10.445	1.000	0.696
ER_least	8.220	4.376	1.000	0.532
FS_most	20.198	15.909	1.000	0.788
FS_least	4.520	2.342	1.000	0.518
GT_most	16.605	12.715	1.000	0.766
GT_least	4.489	2.341	1.000	0.521

Extraction Method: Principal Component Analysis.

Table 7.3 – Factor analysis: total variance explained

	Component	Initial eigenvalues[a]			Extraction sums of squared loadings			Rotation sums of squared loadings		
		Total	% of variance	Cumulative (%)	Total	% of variance	Cumulative (%)	Total	% of variance	Cumulative (%)
Raw	1	51.023	37.263	37.263	51.023	37.263	37.263	32.101	23.444	23.444
	2	24.327	17.766	55.029	24.237	17.766	55.029	30.329	22.149	45.593
	3	20.705	15.121	70.150	20.705	15.121	70.150	33.625	24.557	70.150
	4	12.485	9.118	79.268						
	5	10.929	7.982	87.250						
	6	5.449	3.979	91.229						
	7	3.602	2.630	93.859						
	8	3.520	2.571	96.430						
	9	2.486	1.816	98.246						
	10	2.402	1.754	100.000						
	11	2.19E-015	1.60E-015	100.000						
	12	-5.8E-015	-4.24E-015	100.000						
Rescaled	1	51.023	37.263	37.263	3.734	31.116	31.116	2.684	22.369	22.369
	2	24.327	17.766	55.029	2.027	16.889	48.005	2.596	21.634	44.003
	3	20.705	15.121	70.150	1.973	16.444	64.449	2.453	20.446	64.449
	4	12.485	9.118	79.268						
	5	10.929	7.982	87.250						
	6	5.449	3.979	91.229						
	7	3.602	2.630	93.859						
	8	3.520	2.571	96.430						
	9	2.486	1.816	98.246						
	10	2.402	1.754	100.000						
	11	2.19E-015	1.60E-015	100.000						
	12	-5.8E-015	4.24E-015	100.000						

Extraction method: principal component analysis.
[a] When analyzing a covariance matrix, the initial eigenvalues are the same across the raw and rescaled solution.

Notice that some variables are significantly correlated with more than one component. In particular, the *acceptance* of the BO system (BO_most) is *positively* and rather evenly correlated with all three factors (Table 7.4). Similarly, the *rejection* of the BO system (BO_least) is *negatively* correlated with the Factors 2 and 3, thus reinforcing the acceptance of the BO (tribalistic) psychosocial system in both these components. Further, the strongest correlation between BO_most and the component 2 is relatively instable, since it changes position as we vary the cases in the sample (Appendix 3). For these reason, the BO system does not offer an evocative basis for taking the components to mean. Yet, the correlations themselves are theoretically amenable, since BO is a rather low psychosocial system. As such, according to the nested structure of Graves's scheme, BO is pretty much enveloped, that is, transcended-and-included, by the higher levels of existence. As well, as we see, the acceptance of the BO system (BO_most) is manly associate with the component 2, which includes the likewise sacrifice-self system DQ, whereas it is less rejected in the component 3, which is marked by the acceptance of the receptive sacrifice-self system FS.

In the same vein, both the DQ and the FS psychosocial systems are significantly related with more then one component. This time, however, the signs of the significant correlations appear reversed, so that they help discriminating the meaning of the different factors. Accordingly, the variable-component pattern in the Factor 1 exhibit strong acceptance of the CP psychosocial system – appearing both as positive correlation with CP_most and negative correlation with CP_least – coupled with strong rejection of both the DQ and the FS systems – again appearing in the two forms. This pattern agrees with Graves's theory and allows a pretty clear interpretation of the Factor 1 as expressing the *egocentric wave of existence*.

In turn, the correlation pattern for the component 2 shows the acceptance of the DQ system paired with a strong rejection of the GT system. The positive correlation with the BO_most variable reinforces the sacrifice-self aspect in this component and pushes the developmental level somewhat downward. The correlation pattern is therefore quite sound and supports the interpretation of the

Factor 2 as the *absolutistic wave of existence*. Finally, the Factor 3 clearly expresses the *sociocentric wave*, indicated by the strong acceptance of FS value system coupled with a marked rejection of ER strategic, materialistic and cold rationality. Also, the significant negative correlation with the variable BO_least suggests a typically sociocentric (FS) rebuff of the rejection of other value systems.

At this point, it is worth noticing that these results from the Factor Analysis entail a remarkable accomplishment, for in a sense they already convey a (partial) cross-cultural test of the Gravesian system, given that there was nothing guaranteed regarding the likelihood of obtaining theoretically sound components out of our sample of Brazilian students. So, as the principal components are pretty meaningful, they are appropriate to be used as predictors in the subsequent analyses, so that the robustness of Graves's model will be further pushed through the tests of hypothesis involving them and the behavioral responses in the experimental settings. Moreover, in the very vein of Graves's model, each principal component takes into account the relationship between all the (selected) psychosocial systems, thus representing theoretical constructs, *abstracted* from and ultimately linked to the individual profiles, but which nonetheless do not provide a description of any particular individual. As well, we notice that the acceptance-rejection pattern exhibited in the components 2 and 3 are theoretically suitable in both directions (ER strategic thinking also rejects FS sensitivity and equalitarianism, whereas GT systemic-holistic perspective also rejects DQ absolutism and clear-cut moral judgments). So, even if we do *not* have any principal components representing either the strategic-materialistic (reverse of the component 3) or the systemic-holistic (reverse of the component 2) waves of existence, the basic traits of these two psychosocial systems may still help to interpret negative deviations of the corresponding factor scores, respecting the sample average.

Yet, the importance given to the egocentric wave apparently contradicts the sample profile presented in Figure 7.1. In fact, there is no contraction but only two different perspectives of looking at the data set. Figure 7.1 tell us only

that a very few percentage of the individuals in our sample are likely to be centered at the CP psychosocial system, whereas the Factor Analysis revealed that the *variance* relating to the CP value system comprises a significant amount (23.44%) of the total variance present in the entire sample. In other words, while only a few of our participants may be actually and individually navigating the egocentric wave of existence, the implications of the egocentric wave itself on the behavioral responses can be assessed out of its significant contribution explaining the overall psychosocial variability among the individuals, that is, as a theoretical construct, it has *significance* and *meaning* by itself. Of course, the same is true for other components.

Table 7.4 – Factor analysis: rotated component matrix[a]

	Raw			Rescaled		
	1	2	3	1	2	3
BO_most	**.786**	**.993**	**.713**	**.369**	**.467**	**.335**
BO_least	.008	-1.705	-2.256	.002	**-464**	**-.614**
CP_most	**1.621**	.161	-.244	**.722**	.072	-.109
CP_least	**-3.933**	.543	.773	**-.897**	.124	.176
DQ_most	**-1.302**	**2.943**	-.026	**-.346**	**.783**	-.007
DQ_least	**2.458**	**-1.531**	.066	**.727**	**-.453**	.020
ER_most	.355	-.201	**-3.206**	.092	-.052	**-.827**
ER_least	-.198	.706	**1.959**	-.069	.246	**.683**
FS_most	**-1.902**	-.432	**3.479**	**-.423**	-.096	**.774**
FS_least	**1.210**	.544	**-.763**	**.569**	.256	**-.359**
GT_most	.443	**-3.465**	-.716	.109	**-.850**	-.176
GT_least	.455	**1.444**	.221	.215	**.682**	.104

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization
[a] Rotation converged in 7 iterations.

7.3. A new look at the sample

Before briefly reviewing our expectations concerning the behavioral responses in the CPR experiment, let us give a new look at our sample considering the three principal waves of existence we just obtained. The aim is to show up the intricacy involved in the utilization of the groups that participated in the experimental sections as unity of analysis, for the matter of distinguishing the separate contribution of each wave of existence to explain the observed behavioral responses.

We pursue this by grouping the individuals according to their *factor scores* in the three principal components using Hierarchical Cluster Analysis. The cluster method employed was the *average linkage between groups* and the distance measure was the *squared Euclidean*. Among the numerous possible combinations of hierarchical clusters that can be formed on basis of the three principal components, we found that 9 clusters (plus one isolated individual) would supply the most insightful view of the variance structure present in the data set. The results are summarized in Figure 7.2.

In Figure 7.2, the vertical axis measures the *mean factor scores* in each cluster. Since the factors are normalized (zero mean and unitary standard deviation), the factor mean in each cluster reports average deviations from the corresponding mean in the sample. The percentages indicates the ratio between the number of individuals pertaining to each cluster and the total number of individual in the sample ($\underline{n} = 322$).

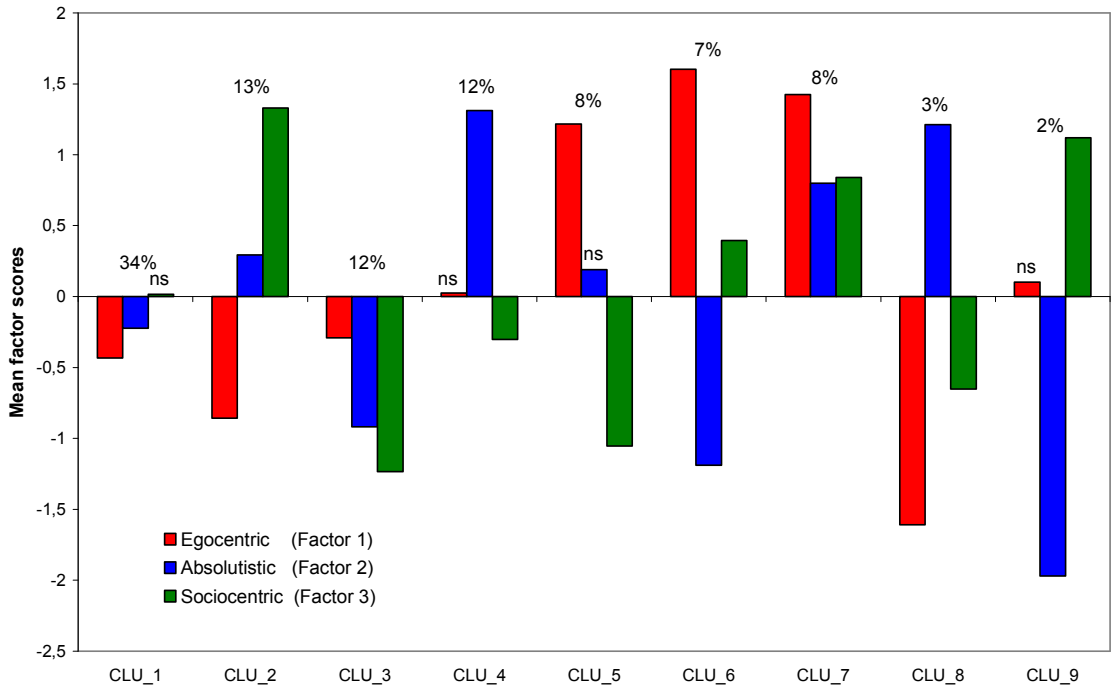


Figure 7.2 – Nine hierarchical clusters on basis of the three principal components.

The point we want to highlight is the difficulty of attaining well defined combinations of characteristics in order to test the implications of the levels of existence on the experimental settings using the groups as unity of analysis. Ideally, we would like to have homogeneous groups of people scoring high in only one factor and be able to control the combination of systems present in the experimental sections. Note that the Factor 1 (egocentric wave) already implies a rejection of both the DQ (blue) and FS (green) value systems. Still, as relative and aggregated measures all three factors may come together as deviations either above (Cluster 7) or bellow (Cluster 3) the sample mean. Of course, this reflects the fluid and flowing nature of the phenomenon at the individual level, but when it comes to aggregate characterizations, things become even cloudier.

Thus, on the basis of these 9 clusters, only 2% (n = 8) of the individuals are clearly sociocentric and anti-materialistic (Cluster 9), that is, individuals who score high in the Factor 3, low than average in the Factor 2 and at least non

statistically significantly (ns) above the sample mean in the Factor 1. Even so, the negative deviation relating to Factor 2 (the absolutistic wave) brings to the scene some affinity with the GT value system. As we saw, GT is also expected to cooperate, while not so unconditionally as FS. We have also 13% (n = 41) of individuals scoring high in sociocentrism and low in egocentrism, but this time their scores are significantly above the sample mean in absolutism (Cluster 2).

As well, the clear-cut absolutists (Factor 2) amount to only 3% (n = 9) in this sample (Cluster 8), while we have more 12% (n = 40) scoring high in absolutism and that score around the sample mean in egocentrism (Cluster 4). The clearest egocentric group is the Cluster 5 (n = 26), where sociocentrism (Factor 3) scores significantly below the sample mean and absolutism (Factor 2) goes around that mean. Again, we do well reminding that the negative deviation as to Factor 3 expresses some affinity with the strategic-materialistic wave or the ER value system (the reversed Factor 3). This is more so in Cluster 3, where the strategic stance is coupled with an affinity toward the holistic GT (negative deviation as to Factor 2). As we know, both are express-self systems but their positioning concerning opportunistic behavior is quite opposite. In the Cluster 6 (n = 22) the egocentrism opposes the absolutism, but the scores for sociocentrism stands above the mean of the sample.

As an obvious statistical event, the biggest cluster (n = 109) congregate the individuals scoring around the sample mean in the three principal components (Cluster 1). The isolated individual is basically of the competitive type, combining a high score in the Factor 1 with extremely low (more the three standard deviations below the average) score in sociocentrism (Factor 3). In relation to absolutism, this participant is no different from the average.

7.4. A first look at the groups characteristics

As we are saying, we would like to get to homogeneous as well as controlled combinations of levels of existence in the experimental sections, so that the groups themselves could work as the main unity of analysis in the

assessment of the interactions of the values systems with the experimental situations and treatments. However, owing to the difficulty just described, together with the logistic impossibility of scheduling preferred participants to take part into specific experimental sections, such aim ought to be abandoned. We thus had to work with groups with uncontrolled configurations relating to the values systems, produced according to the restrictions of our participants to take part in the experimental sections schedule.

The final configurations of the 25 groups (8 participants each) in the CPR experiment are presented in Figure 7.3. As we see, there are some few groups that casually approximate our initial intend, such as the Groups 1, 2, 4, and 16 (the group identification numbers do not reflect the experimental schedule). The Groups 8 and 9 also reasonably isolate the factor absolutism, but the factor egocentrism still remains above the mean. Interesting combinations of systems appear in the groups such as the Group 17 (high absolutism and sociocentrism with low egocentrism) and the Groups 3, 10, and 20 (high egocentrism and sociocentrism and low absolutism).

7.5. The levels of existence within the appropriation dilemma

In the last chapter we advanced the theoretical expectations about the behavioral responses associated to each separate psychosocial system. Now that we have obtained the three principal waves of existence present in our sample, we need to reconstruct those hypotheses according to the variable-component correlation pattern within each factor, in order to test their relationship with the experimental observations.

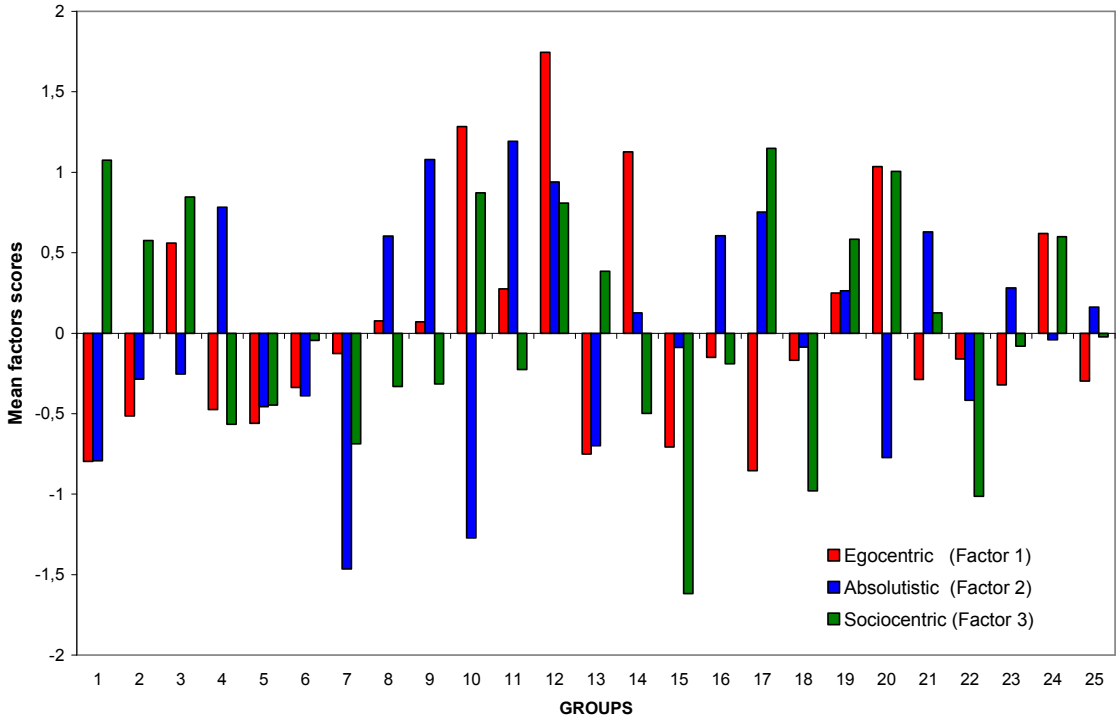


Figure 7.3 – Actual groups profiles on basis of the three principal components.

To recapitulate, let us remind that the *sacrifice-self systems* – BO (purple), DQ (blue) and FS (green) – as well as the GT (yellow) value system are expected to support the joint strategy (whilst for different reasons), and to be less go-getting in their individual appropriation behaviors than the *express-self systems* CP (red) and ER (orange). Further, though also being an express-self system, the theme of the (first *being level* or second tier) GT is *'Express self for what self desires, but never at the expenses of others and in a manner that all life, not just my life, will profit'*. Conversely, the theme of the (5th *subsistence level*) ER is *'Express self for what self desires, but in a fashion calculated not to bring down the wrath of others*'; and that of the (3rd *subsistence level*) CP is *'Express self, to hell with others and the consequences, lest one suffer the torment of unbearable shame'*. The themes of the 2nd, 4th and 6th *subsistence levels* and *sacrifice-self systems* are, respectively: *'Sacrifice self to the way of*

your elders' (BO), *'Sacrifice self now to receive reward later'* (DQ) and *'Sacrifice now in order for all to get now'* (FS).

Accordingly, the hypothesis **H1-CPR** posited that overinvestment in the unregulated conditions is higher for the express-self systems (CP, ER, and GT) than for the sacrifice-self systems (BO, DQ, and FS). Thus, given the variable-component correlations comprising the factors, we raise the following:

- **HF1-CPR**: the overinvestment in the unregulated condition (baseline) correlates positively with the scores in the Factor 1 (egocentric) and negatively with the scores in the Factors 2 (absolutistic) and 3 (sociocentric).

The hypothesis **H2-CPR** stated that the degree of cooperative behavior (i.e. restrained investment in the CPR) in the nonbinding communication setting is higher for the sacrifice-self systems (BO, DQ, and FS) and for the express-self systems GT, than for the express-self systems CP and ER. Now we have a difficulty in forming an unambiguous expectation about this behavioral response in relation to the Factor 2 (absolutistic), since we have declared that both DQ and GT are expected to cooperate and they appear inversely correlated in the variable-component structure of Factor 2 (Table 7.4). Moreover, we have also posited that both DQ and GT are expected retaliate cheaters, by skipping the norm themselves (even if DQ is more inclined to do so, out of taste for punishment). So, in the measure that both traits may be fairly influencing the observed response (as suggested by DQ and GT communalities in Table 7.4), we expect to find no significant statistical relationship between the level of investment and the scores in the Factor 2, for the influence in the same direction is balanced by the opposition between DQ and GT in the internal structure of Factor 2. Or, alternatively, both the positive and the negative correlations with the Factor 2 are consistent with the cooperative behavior, so that that these influences may offset one another. If statistical significance is eventually found, we cannot immediately tell which value system (DQ or GT) is associated to it without examining the relationship with the separate variables.

As we shall see, such configuration has practically negated explanation power to the Factor 2 for most of the behavioral responses examined in this

study. Notwithstanding, it should be reinforced that the variable-component correlation pattern constituting the Factor 2 is itself theoretically sound. The actual indistinctness results from the dramatic change of perspective stressed out by Graves himself when the individual leaves the *subsistence* group-level of existence (from AN to FS) and gets to the first *being* level of existence at GT. At this point of consciousness development, the way the individual expresses the self (i.e. its *agency* capacity) no longer disagree with the willingness to cooperate (the *communion* capacity), previously associated only with the sacrifice-self systems (BO, DQ, and FS). Yet, people centered at DQ strongly reject the GT value system, and this is what Factor 2 correlation pattern is expressing. So, resuming all the above clarification, the **HF2-CPR** ought to posit that:

- **HF2-CPR**: the level of investment in the CPF in the nonbinding communication setting correlates positively with the scores in the Factor 1 (egocentric) and negatively with the scores in the Factor 3 (sociocentric). No statistical relationship is expected between the level of investment in the nonbinding communication condition and the scores in the Factor 2 (absolutistic).

As for the frequency of primary and secondary defections, the reformulation of the hypothesis according with the structure of the principal components leads us to advance that:

- **HF2.1-CPR**: The frequency of defection (either form) correlates positively with the scores in the Factor 1 (egocentric) and negatively with the scores in the Factor 3 (sociocentric). No statistical relationship is expected for the frequency of defections and the scores in the Factor 2 (absolutistic), owing to the reasons outlined above.
- **HF2.2-CPR**: The frequency of cheating (primary defection) correlates positively with the scores in the Factor 1 (egocentric) and negatively with the scores in the Factor 3 (sociocentric). As explained, no statistical relationship is expected involving the frequency of cheating and the scores in the Factor 2 (absolutistic).

- **HF2.3-CPR**: The frequency of the "never defect" response correlates positively with the scores in the Factor 3 (sociocentric) and negatively with the scores in the Factor 1 (egocentric). Again, no statistical relationship is expected between the "never defect" response and the scores in the Factor 2 (absolutistic).

Finally, concerning the sanctioning condition, we have posited that the CP and ER systems are expected to pay more fines, due to tentative and immediately punished defections during this institutional setting. Accordingly, in terms of the principal components we expect:

- **HF3-CPR**: The expenditure with fines correlates positively with the scores in Factor 1 (egocentric) and negatively with the scores in the Factor 3 (sociocentric). No statistical relationship is expected involving the scores in the Factor 2 (egocentric), since this is just another expression of the cheating behavior.

As for the use of the sanctioning mechanism, assessed by the total expenses with fees, we have advanced that the BO, DQ and GT value systems are expected to expend more with fees, while CP, ER and FS are expected to expend less, albeit for different reasons. In particular, we stressed out that only for GT such behavior can be interpreted as authentic "strong reciprocity", since BO and DQ are supposed to attribute some rewarding value to the act of punishment itself. Conversely, while FS is expected to incur into personal costs for the sake of reinforcing the social norm, it dislikes punishing mechanism, while CP and ER are just not willing to incur into the costs of reinforcing the social norm (second order dilemma). We have also pondered that CP may discount such a cost out of "taste for punishment".

Note that, for this behavioral response, besides the ambiguity concerning the relationship involving the Factor 2, we also cannot unequivocally interpret the relationship between expenditures with fees and the scores in the Factor 3, because both FS and ER are expected to exhibit similar responses. On the one hand, when the scores for the Factor 3 are high, less expenditure is expected owing to FS kindness, but when the scores for the Factor 3 are low, also less

expenditure is expected out of ER greediness. In order to distinguish which would be the case we need to look at the disaggregate variables. Additionally, the same problem is about to happen respecting the Factor 1, given the negative correlation between CP and FS within this component and that both value systems are expected to respond in the same direction.

Given all the above, we do not expect to find significant associations between the principal components and the expenditures with fees, so that this relationship will be examined through the hypothesis **H4-CPR**, respecting the disaggregate variables.

Finally, in order to facilitate the interpretation of the aggregate observations in the CPR experiment, let us just briefly sketch some theoretical reasoning relating to the collective behavior. The general point is basically that the collective commitment with the joint strategy is hazarded by the presence of *individuals* scoring high in egocentrism (Factor 1) as well as low in sociocentrism (Factor 3). During the nonbinding communication setting, the stability of the collective strategy is particularly challenged if the group shows a configuration such that either high egocentrism or low sociocentrism appears together with high absolutism (e.g. Groups 8, 9, 11, 12, 14). This is because primary cheating, due to CP or ER, is likely to be promptly retaliated by DQ. Conversely, the stability of the joint strategy (and thus the efficiency of the collective appropriation) is expected to be favored in groups where sociocentrism dominates and egocentrism is low (e.g. Groups 1, 2 and 17 in Figure 7.3).

In general, it is expected that the imposition of the sanctioning institution will improve the efficiency of the collective appropriation strategy, through the reinforcement of the social norm. Nonetheless, depending on the frequency of defections and the intensity of the (costly) sanctioning, the economic performance may easily drop bellow the level observed during the communication condition and even the baseline setting. Again, such deterioration of the economic efficiency is more likely if the group presents the combination mentioned above (high egocentrism or low sociocentrism together with high absolutism). We expect that groups scoring high in absolutism (Factor 2),

especially if simultaneously scoring low in sociocentrism (Factor 3), will make a conspicuous use the sanctioning mechanism in order to punish defectors.

Yet, these expectations formed on the basis of aggregate information are acknowledge fragile, given the observations raised above, and since the groups have very heterogeneous composition. Moreover, the individual profiles are themselves formed by combinations of the various waves of existence. For this reason, the descriptions and interpretations regarding the groups' performance we are about to present comprise only a preliminary exploration of the explanatory power attributable to Graves's constructs. The results of the formal statistical tests will be analyzed after we have finished this exploratory appraisal. In the next section, we begin investigating the overall effects of the communication and the sanctioning treatments in relation to the baseline condition.

7.5.1. Aggregate effects of communication and sanctioning

To recapitulate, the matter of face-to-face communication is a theme outside the scope of the strict economic approach, for thus defined it deals exclusively with social interactions mediated through the markets. We saw that the adoption of noncooperative game theory represented a major transformation in these regard. Yet, noncooperative game theory also ruled out the significance of face-to-face communication. As noticed, the ability to communicate is indeed treated as superfluous in much of the contemporary noncooperative game theory (OSTROM et al., 2002).

In the Chapter 3 we argued that such inference is tied to the assumptions respecting *information* and *valuation* in the model of "rational egoists". We said that those assumptions have *already* ruled out the effect of communication, since it cannot add new information about how different players may evaluate a joint strategy eventually contracted through communication (i.e. given those assumptions, all players know that all players are "rational egoists" and that all players know that). Conversely, if agents are reendowed with some *conscience*

and *culture*, so that they can differ in type, the obvious importance of communication can again be captured by the game theoretical approach, since communication is now a source of information about player's types.

We acknowledged that the institutional setting often leads concrete individuals, who diverge in types, to behave similarly. In particular, the inability to communicate can indeed lead individuals who otherwise would be willing to cooperate to behave *as if* they were "rational egoists". Still, the point here is that the ability to communicate effectively does not depend only on the external circumstances, but entails a totally different intersubjective mind-set, departed from the instrumental conception of action and directed toward mutual understanding (HABERMAS, 1984 e 1987). As the likelihood of getting to mutual understanding, as prerequisite for improving the collective performance, depends on both the horizontal (shared horizons) and vertical (shared depth) dimensions involved in the phenomenon of communicative action (Wilber, forthcoming), we will examine this effect, first, by exploring the relationship between the effectiveness of communication and the average psychosocial profiles in the various groups, and latter through the statistical relationships between the factor scores and the behavioral responses following the communication opportunity.

In this section, we begin examining the most aggregate effect of communication as well as of the sanctioning mechanism considering the whole sample (25 repetitions of the complete experimental design). So, in Table 7.5 we present the descriptive statistics for the aggregate payoffs per round in each experimental condition. The Net_Yield refers to the aggregate gain per round in the Market 2 (the CPR) after discounting the opportunity cost of the corresponding investments in the Market 1 (the secure alternative); M2_Total and M1_Total are the gross aggregate gains in the Market 2 and Market 1, respectively. The GT_Total is the sum of the aggregated gross gains in both alternatives. Since the information conveyed by the M1_Total and GT_Total are redundant in respect to Net_Yield and M2_Total, in the remaining we will report only these last two measures.

Table 7.5 – CPR experiment: descriptive statistics

Treatment		N	Minimum	Maximum	Mean	Std. Deviation
Baseline	Net_Yield	250	-4,05	3,18	,6028	1,43729
	M2_Total	250	,45	5,29	3,9630	,99742
	M1_Total	250	1,5	4,0	2,640	,4515
	GT_Total	250	1,95	9,18	6,6028	1,43729
	Valid N (listwise)	250				
Communication	Net_Yield	250	-2,28	3,24	2,6381	,94808
	M2_Total	250	1,87	5,29	4,9751	,46539
	M1_Total	250	1,9	4,4	3,663	,5611
	GT_Total	250	3,72	9,24	8,6381	,94808
	Valid N (listwise)	250				
Sanctioning	Net_Yield	250	-13,44	3,24	2,2302	2,42851
	M2_Total	250	-10,59	5,29	4,3516	2,26104
	M1_Total	250	2,9	4,4	3,879	,3739
	GT_Total	250	-7,44	9,24	8,2302	2,42851
	Valid N (listwise)	250				

The aggregate net yield per round at the optimum (Pareto) level of investment (36 tokens in Market 2) is R$3.24 and the at the Nash equilibrium (64 tokens) is R$1.28 (40% of the optimum). While Net_Yield means are not much representative, owing to the large variance, the average gain in the baseline condition was only 19% of the optimum. On average, this performance got up to 81% in the communication treatment and recurred down to 69% in the sanctioning treatment, due to the economic lost with fines and fees.

Given that the treatments strongly affected the pattern of appropriation from the CPR experiment (i.e. the level of investment in Market 2), the *assumption* of equality of variances of the collective gains between the three conditions does not hold. In this case, either the Welch or the Brown-Forsythe statistics are preferable to the *F* statistic to test for the overall equality of means. As well, Tamhane supplies a conservative pairwise comparisons test which is appropriate when the variances are unequal.

The overall test of equality of means indicate that the sample's average aggregate gains per round in the three treatments are significantly ($p < .0001$)

different one from the others (Table 7.6). Tamhane's pairwise multiple comparisons test shows that the collective appropriation from the CPR (Market 2), measured either as net yield or gross gain, are statistically significantly higher in the communication treatment than in both the baseline and the sanctioning conditions. The sanctioning mechanism improved the efficiency of collective appropriation only in relation with the baseline condition (Table 7.7).

Table 7.6 - CPR experiment: robust tests for equality of means

		Statistic[a]	df1	df2	Sig.
Net_Yield	Welch	175,240	2	448,307	,000
	Brown-Forsythe	98,131	2	490,660	,000
M2_Total	Welch	110,143	2	399,810	,000
	Brown-Forsythe	30,919	2	366,456	,000

[a] Asymptotically F distributed.

Table 7.7 – CPR experiment: multiple comparisons

Dependent variable	(I) Condiction	(J) Condiction	Mean difference (I-J)	Std. error	Sig.	95% confidence interval	
						Lower bound	Upper bound
Net_Yield	Baseline	Communication	-2,03532*	,10890	,000	-2.2964	-1,7743
		Sanctioning	-1,62740*	,17848	,000	-2,0553	-1,1995
	Communication	Baseline	2,03532*	,10890	,000	1,7743	2,2964
		Sanctioning	,40792*	,16488	,041	,0122	,8037
	Sanctioning	Baseline	1,62740*	,17848	,000	1,1995	2,0553
		Communication	-,40792*	,16488	,041	-,8037	-,0122
M2_Total	Baseline	Communication	-1,01212*	,06961	,000	-1,1791	-,8451
		Sanctioning	-3,8860*	,15630	,040	-,7636	-0,136
	Communication	Baseline	1,01212*	,06961	,000	,8451	1,1791
		Sanctioning	,62352*	,14600	,000	,2727	,9743
	Sanctioning	Baseline	,38860*	,15630	,040	,0136	,7636
		Communication	-,62352*	,14600	,000	-,9743	-,2727

* The mean difference is significant at the .05 level.

Still, as we will discuss in greater detail below, this result can be partially attributed to the events observed in only two groups (Groups 21 and 22). We shall refer to these events as "institutional battles" since in each of these groups there was one participant who sturdily rejected the joint strategy and used the sanctioning mechanism as a weapon to hit back the players who were willing to coerce them to conform to the accord. As a result, it was observed a tremendous economic lost in the sanctioning condition for these groups. The total lost amounted to 244% of the optimum net yield, considering the 10 rounds of the sanctioning treatment, in the Group 21, and to 117% in the Group 22. If we exclude these groups from the test, under the claim that they show unrepresentative situations, then the efficiency of the communication and the sanctioning conditions become comparable *when* measured by the net yield in Market 2 (i.e. the difference between the means is not statistically significant at .05% of probability). Nonetheless, the difference remains significant, showing a worse performance in the sanctioning treatment, if we take into account only the

(gross) gain in Market 2 (Table 7.8). We thus regard that, in the overall, while the sanctioning institution did help to discourage individual cheating, the economic costs involved may not compensate the lost due to cheatings themselves.

Table 7.8 – CPR experiment: multiple comparisons (after taking groups 21 and 22 out of the sample)

Dependent variable	(I) Condiction	(J) Condiction	Mean difference (I-J)	Std. error	Sig.	95% confidence interval	
						Lower bound	Upper bound
Net_Yield	Baseline	Communication	-2,20692*	,10842	,000	-2,4670	-1,9468
		Sanctioning	-2,13639*	,12732	,000	-2,4415	-1,8312
	Communication	Baseline	2,20692*	,10842	,000	1,9468	2,4670
		Sanctioning	,07053	,09767	,852	-,1637	,3048
	Sanctioning	Baseline	2,13639*	,12732	,000	1,8312	2,4415
		Communication	-,07053	,09767	,852	-,3048	,1637
M2_Total	Baseline	Communication	-1,09888*	,07058	,000	-1,2684	-,9293
		Sanctioning	-,85813*	,09845	,000	-1,0941	-,6222
	Communication	Baseline	1,09888*	0,7058	,000	,9293	1,2684
		Sanctioning	,24075*	,07550	,005	,0594	,4221
	Sanctioning	Baseline	,85813*	,09845	,000	,6222	1,0941
		Communication	-,24075*	,07550	,005	-,4221	-,0594

* The mean difference is significant at the .05 level.

However, the comparison of the economic performance between these two settings depends on many other factors that we have not investigated. The first which is noteworthy mentioning is the initial endowment given to the participants. The higher the appropriation power of individual players the higher it will be the economic impact of individual defections. For this reason, as Ostrom et al. (2002) have shown, the increasing of the players' appropriation power can weaken the ability to sustain the joint strategy on bases of communication alone. The same could probably result had the players received

only aggregate information about outcomes. Finally, in our experimental design the two opportunities to communicate are costless, while the sanctioning mechanism is costly.

All of these uninvestigated factors render our comparisons of the economic performances between the two institutional conditions rather fragile. Actually, we have not designed the experiments with this intend but in order to test out the implications of the psychosocial levels of existence on the collective dilemma in interaction with alternative institutional settings.

A final observation concerning the most general sanctioning behavior is that defections in the last round were systematically fined, despite the fact that such attitude, likewise those just reported, confirm a disposition to incur into pure monetary lost. All this is of course in open contrast with the predictions of the "rational" choice model. Yet, as we have saying, it is not that such behaviors are irrational but that the additional abstractions involved in the conventional model, particularly with regard to the scope of social interactions and human concerns, simply rule out whatever other expression of rationality but the optimization of individual material payoffs. The experimental results show that, in general, the opportunity of face-to-face communication shifted the province of the rational choice from the individual to the collective domain. Our purpose now is to investigate to what extend this general effect is related with the levels of existence.

7.5.2. Value systems, rule following and economic performance

In order to visualize the hypothesized relationships between the disposition to cooperate and the psychosocial level of development as captured by the three principal components, we begin by handily clustering the groups according to the number of defections observed during the communication and sanctioning conditions. It should be noticed that the total number of defections include both primary deviations from the joint strategy, as established by the participants themselves, and secondary deviations in response to the initial defection. We did *not* count as defection the reactions in which the participant

decided to *reduce* his or her investment in the Market 2 (the CPR) after someone else have overinvested. However, as we shall see, it was observed a situation where the participants decided to fine an individual who invested *less* than the agreed upon during the sanctioning treatment. This occurrence, in spite of being the only one of its kind, indicates that the wiliness to incur into a costly activity to reinforce the behavior of rule following present a kind of autonomy (WITTGENSTEIN, 1980), with is not tied to the logic of maximizing the individual payoff.

Again, what follows is only a first exploitation of the results concerning the relationship between the waves of existence and the behavioral responses in the experimental settings. Acknowledged, the aggregation of the factor scores among the individual does not provide a fair description of the psychosocial profile of the groups, unless the variance within the groups were sufficiently small. Notwithstanding, for the purposes of exploiting the general trends and offering a visual description of them, the aggregation schema works well enough. The tests of the statistical significance of our working hypothesis will be pursued in the next section, using disaggregated information.

Meantime, we clustered the groups into 6 categories of increasing number of total observed defections: (1) no defection at all; (2) 1 and 2 defections; (3) 3, 4 and 5 defections; (4); 11 and 16 defections; (5) 19 and 22 defections, and (6) 27, 31, 39 and 42 defections. We then calculated the mean factor scores for egocentrism, absolutism and sociocentrism considering all the individuals pertaining to those clusters of groups. The results are summarized in the Figure 7.4.

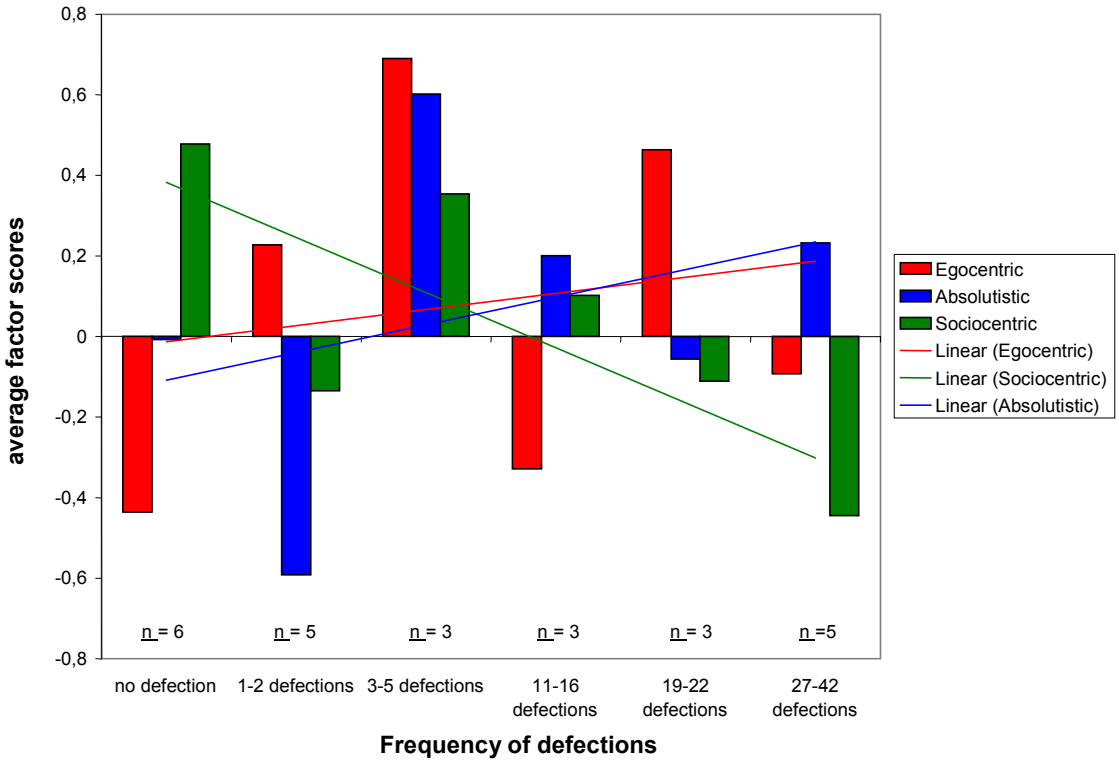


Figure 7.4 – Average factor scores and linear trend for increasing number of defections.

We will comment more on these patterns as we proceed with the descriptions of the performance observed in each cluster. The most general trait, however, is indicated by the linear trend line adjusted to each series of factor scores. The goodness of fitting ranges from very poor for egocentrism (R^2 = 0.03) to reasonable for sociocentrism (R^2 = 0.56). However, this is a function of the way we have clustered the groups and it passes over all the individual variation present in the sample. In any case, for illustrative purpose only, Figure 7.4 already offers an initial indication that *the relation between the value systems and the disposition to cooperate exists and runs in the expected direction*. Accordingly, the higher the scores in sociocentrism the more the individuals honored their commitment with the regulating norms as settled in each group. Conversely, the higher the scores in egocentrism and absolutism the more the individuals deviated from that norms. In order to distinguish the motive under the

deviations, in the next section we will isolate the primary defection from the total number of defections in the hope of showing that egocentrism is under the cheating behavior proper while absolutism correlates more with the retaliation of primary defections. The same aim is also attempted using the total expenditure with fines and fees, as we shall see.

But resuming with the description of the groups performance, Figures 7.5 through 7.23 present: (1) the evolution of the net yield per round as percentage of the optimum (the net yield at Pareto); (2) the aggregate profile of each group using the average factor scores for each of the three principal components; (3) the number of defections per round in the communication and sanctioning conditions, and (4) the total expenditure with fines and fees per round of the sanctioning treatment.

To remind, the optimum strategy consisted of an aggregate investment of 36 tokens in the Market 2. This could be achieved by a rotational scheme in which four participants invest 4 tokens and the other four invest 5 tokens in the Market 2. The strategies of investing either 4 or 5 tokens return the same total (Market 1 plus Market 2) payoff (R$ 1.15 per round), representing 99% of the optimum net yield. The Nash equilibrium is for each participant to invest 8 tokens in Market 2, resulting into an aggregate investment of 64 tokens. This gives an efficiency of 40% of the optimum.

The numbers above the bars in the two lower panels of the figures indicate the number of either defectors or punishers in the corresponding round. The dotted line in the lower left panel indicates the moment in which the participants had a second, costless, and optional (five minutes) communication opportunity. The absent of the doted line between the 10^{th} and 11^{th} rounds indicates that the group disregarded this second communication opportunity. As long as "time is money" we can see the possibility of giving out the second talk without disturbing the stability of the joint strategy as an improved performance.

We begin with the groups in which cooperation ran smoothly, that is, the ones where no a single defections was observed during both the communication and the sanctioning conditions. This effect was observed in 6 out of the 25

groups (Figures 7.5 through 7.10). The appropriation performance in Group 1 (Figure 7.5) was inferior because the participants did not find the optimal strategy. They settled a rotational schema where the individuals who had uneven ID numbers started investing 15 tokens in the Market 2 while the others, holding the even ID numbers, invested all their tokens in the Market 1. This gives an aggregate investment of 60 tokens per round in the Market 2, and pays a return corresponding to only 56% of the optimum net yield as indicated in the Figure 7.5. The participants did use the second (five minutes) communication period but they reiterate the original proposed strategy.

The Group 3 established a strategy of investing between 4 and 6 tokens in the Market 2. Again, the agreement was fully respected but the economic performance started to fall as the participants gradually converged toward an investment of 6 tokens each. Yet, they considered things were going well enough and dispensed the additional five minutes of talk. The participants in the Groups 4 and 25 limited the appropriation from the CPR by settling a norm of investing 5 tokens each while those of Group 13 agreed investing 4 tokens. Either of these strategies returns 99% of the optimum net yield. The participants in the Groups 13 and 25 dismissed the additional five minutes of communication.

In the Group 17 the participants set up a more complicated strategy in which in each round one subject was allowed to invest 14 tokens in the Market 2 while the others should invest four tokens. This gives a total of 42 tokens per round and results into a net yield corresponding to 97% of the maximum. But this scheme could not be completed for a third rotation due to the limit of 30 rounds. Since they knew the total number of rounds, the strategy involved a new schema entering in operation in the 27^{th} round. According to it, in the 27^{th} and 28^{th} rounds the participants who were given ID numbers from 1 to 4 should invest 10 tokens in the Market 2 and the others invest 5 tokens. In the two next and final rounds they inverted the rule. This strategy was briefly ratified during the second talk. As a result, they invested a total of 60 tokens in the Market 2 in each of the four last rounds and the efficiency dropped to 56% of optimum net yield.

Concerning the value systems, the averaged pattern shown in Figure 7.4 is exactly what we should expect as correlative to perfect cooperation. In all of these groups, whenever it was observed a less than optimum economic performance the issue is totally due to errors in finding the optimum appropriation strategy (Groups, 1, 3 and 17). In the Group 17, the complication of the whole strategy that involved two different rotational schemas makes the steadiness of cooperation particularly remarkable. As well, except for the Group 1, the near to collective optimum performances observed in all of these groups makes the extra profits accruable from an isolate defection more significant and thus more tempting. In spite of this, not a single defection occurred in a total of 120 rounds of the communication and sanctioning treatment for these six groups taken together.

Part of the explanation, we maintain, must be found in the psychosocial composition of these groups. On the one hand, the lowly factor scores for egocentrism are critical for preventing primary defections. On the other hand, the sociocentric wave of existence (with the accompanying anti-ER stance) hallmarks the inclination toward consensus during the communication period – an undertaking evidently facilitated by the absent of both egocentric and absolutistic positions. Moreover, in three of these groups the participants did not need to expend time (transaction costs) with a second round of negotiation.

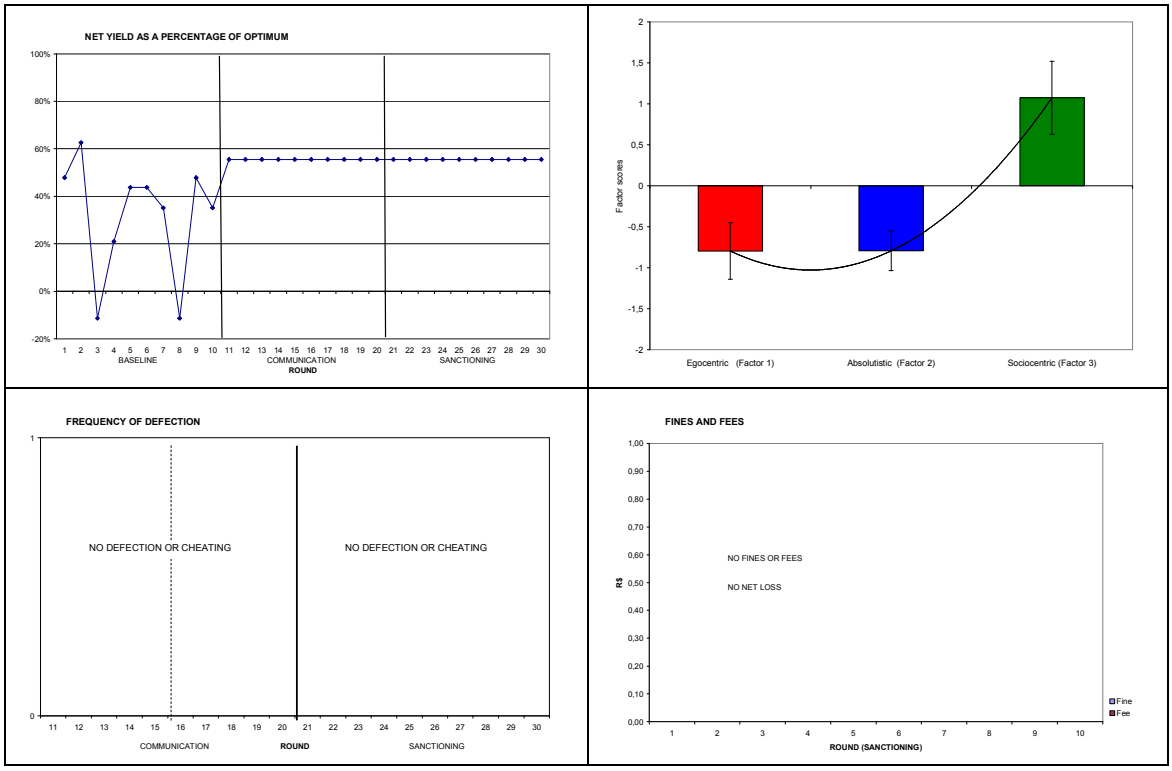


Figure 7.5 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 01.

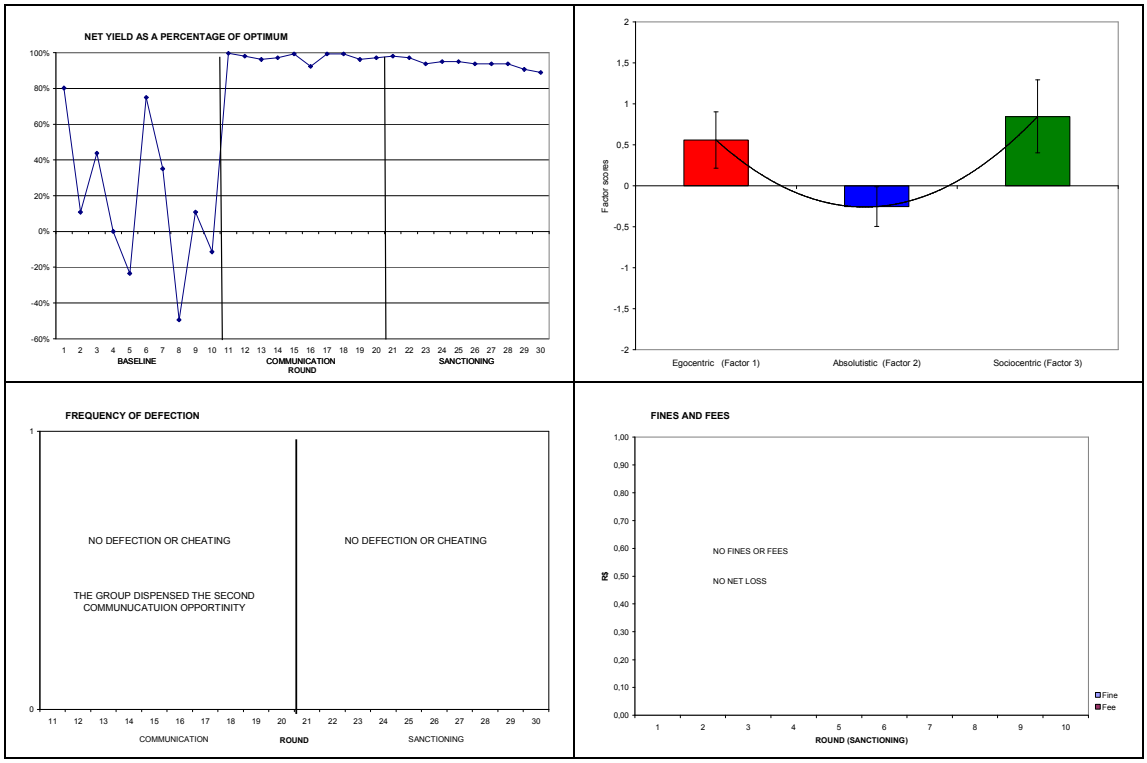


Figure 7.6 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 03.

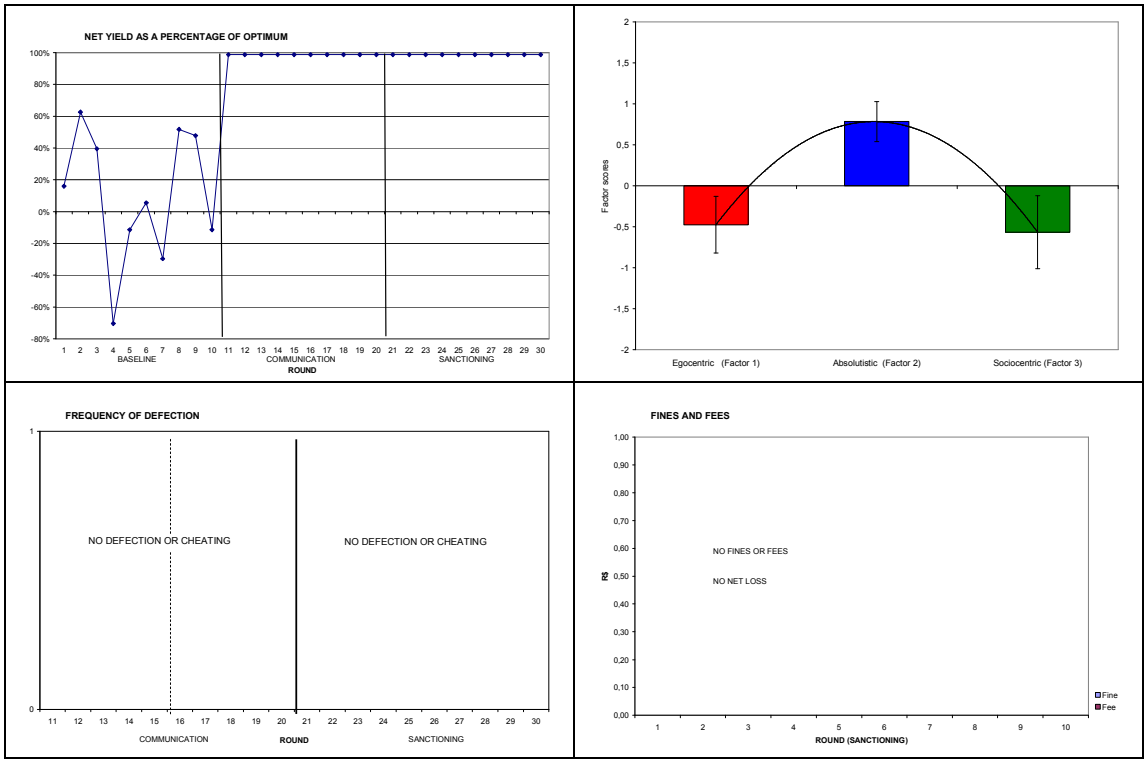


Figure 7.7 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 04.

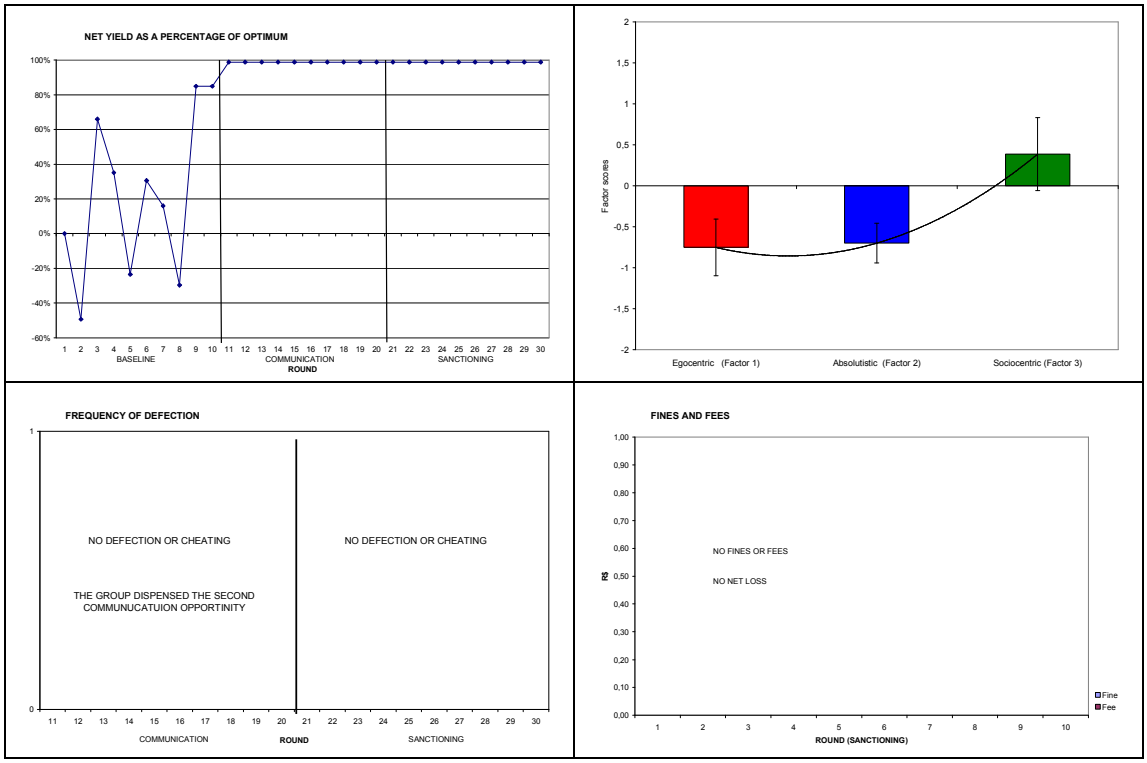


Figure 7.8 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 13.

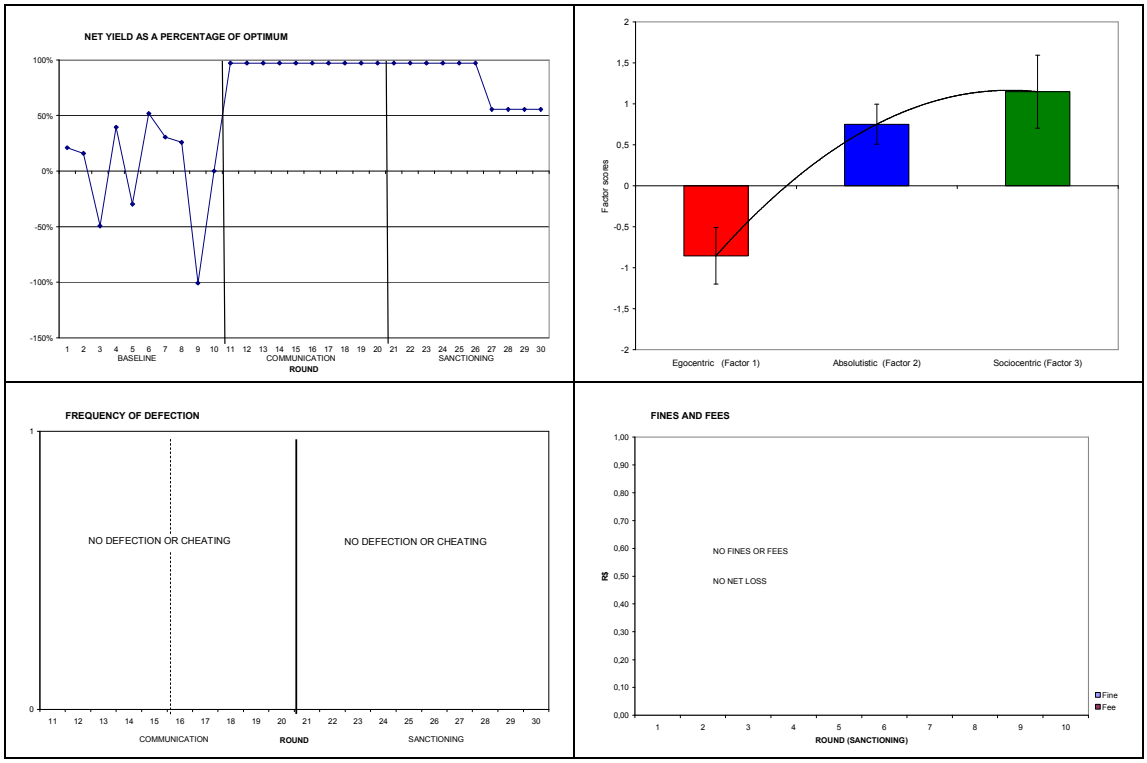


Figure 7.9 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 17.

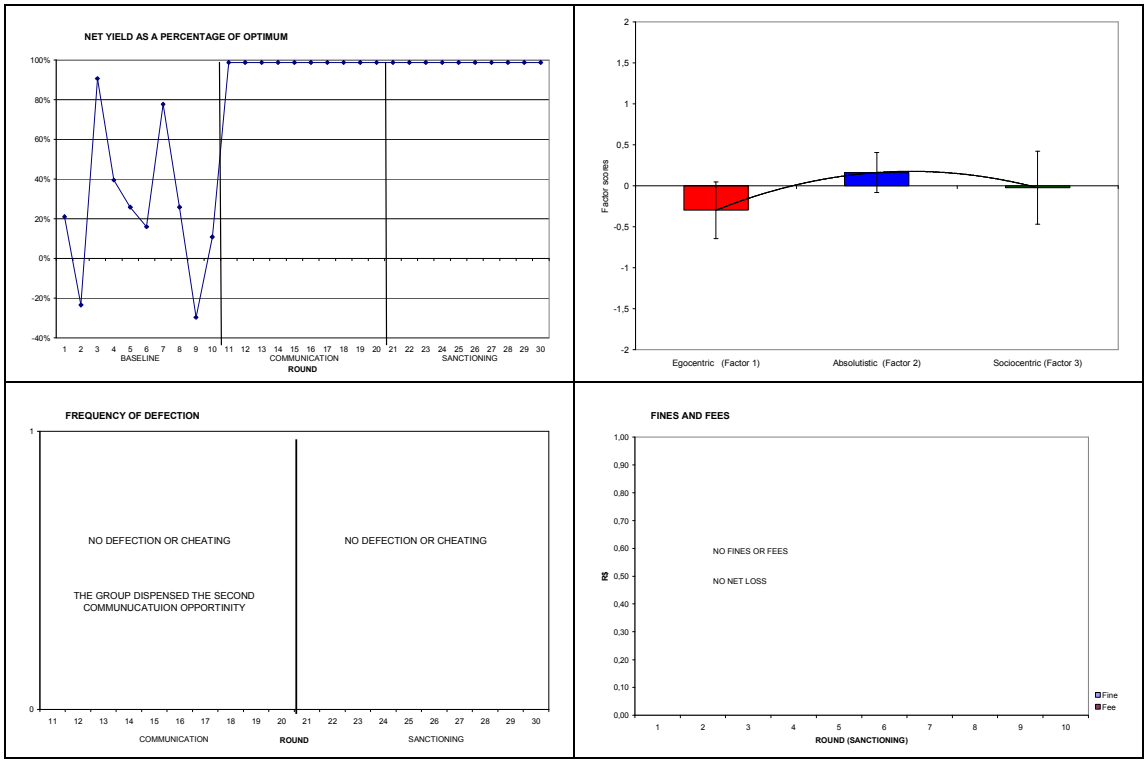


Figure 7.10 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 25.

In the next cluster we come together the 5 groups in which it was observed either 1 or 2 defections (Figures 7.11 to 7.15). In terms of performance, the participants in all these groups were able to find the near to optimum strategy of investing 5 tokens in the Market 2 (99% of the maximum net yield). Only in the Group 14 this conclusion need to wait until the second communication opportunity. The initial agreement, which held for the first five rounds of the communication condition, was to invest a maximum of 8 tokens each. As a result, the performance in those rounds varied between 52% and 69% of the optimum (Figure 7.13). The participants in Group 7 dispensed the additional communication opportunity.

As already indicated, the defections during the sanctioning condition resulted into a much larger aggregate lost owing to the additional costs of fines

and fees. As a general observation, the participants preferred to pay the highest fee (R$ .20) to impose the highest fine (R$ .80) on defectors, including when penalizing multiple defectors. This choice was perhaps induced by the better ratio fine/fee among the alternatives, even if the theoretical equilibrium predicts no sanctions.

To illustrate the additional lost, starting from an agreement to invest 5 tokens in the CPR, a single defection in which the defector invest 15 tokens (i.e. the total endowment) results into a decline of the aggregate efficiency from 99% to 85% of the maximum net yield (see Group 5, Figure 7.15), if no sanctioning is imposed. If there are two defections of this kind, the collective efficiency falls down to 56% of the maximum (Group 20, Figure 7.14). When sanctioning is allowed, the additional lost will depends on the disposition of the participants to pay the fees. In the Group 5, where a new 15 tokens defection occurred in the 24th round, the total fine amounted to R$ 2.80 and the corresponding total fee was of R$ .75. As a result, the efficiency drooped to -25% of the maximum. In the Group 7 (Figure 7.12), the same situation led the efficiency to fall down to -29% of the optimum and in the Group 14 (Figure 7.13) it cut down to -71% of the maximum net yield.

In the Group 6 (Figure 7.11) the collective efficiency dropped from 99% to -61% due to a defection in the last round. The defector invested 12 tokens, instead of the "rational" 15 tokens we could expect, and was punished with a total fine amounting to R$ 3.90. This fine was imposed by six out of the other seven players. Four of them incurred in the highest fee (R$ .20) to impose a fine of R$ .80, one paid R$ .14 to fine the defector in R$ .50 and the others chose to pay R$ .08 to fine her in R$ .20. And all of this even though they knew the game was over. Besides, that was a behavioral regularity whenever a last round defection occurred.

As for the psychosocial profiles, the aggregate scores for these five groups taken together show that egocentrism is above the sample mean while absolutism and sociocentrism are bellow the average. In particular, the aggregate score for absolutism is the lowest observed in Figure 7.4, suggesting a bent

toward the systemic-holistic (GT) wave of existence (the reversed Factor 2). Since, as we said, the total number of defections includes both primary defections (cheating) as well as reciprocating defections, we suggest that the low scores in absolutism helps to explain the little number of defections in these groups despite the scores for egocentrism. In other words, we suggest that the primary defections (cheating) are mainly due to the egocentric wave while the absolutists generally reciprocate as a form of punishment. This hypothesis will be formally tested in the next section.

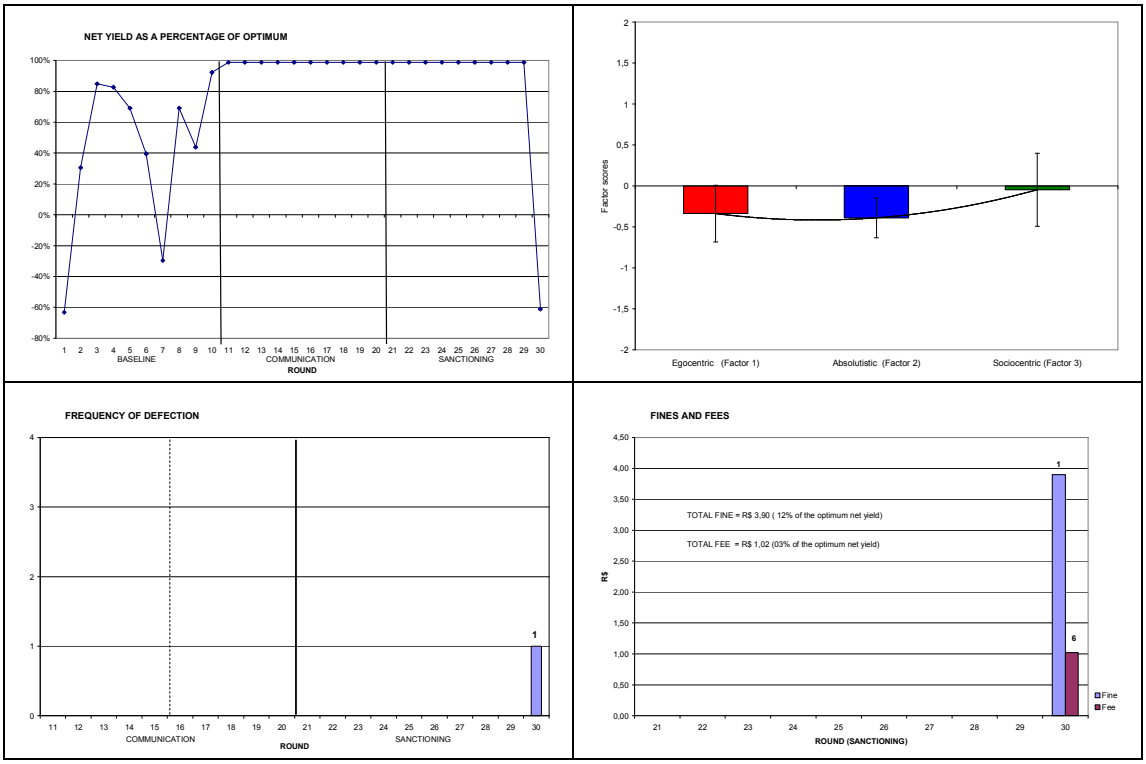


Figure 7.11 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 06.

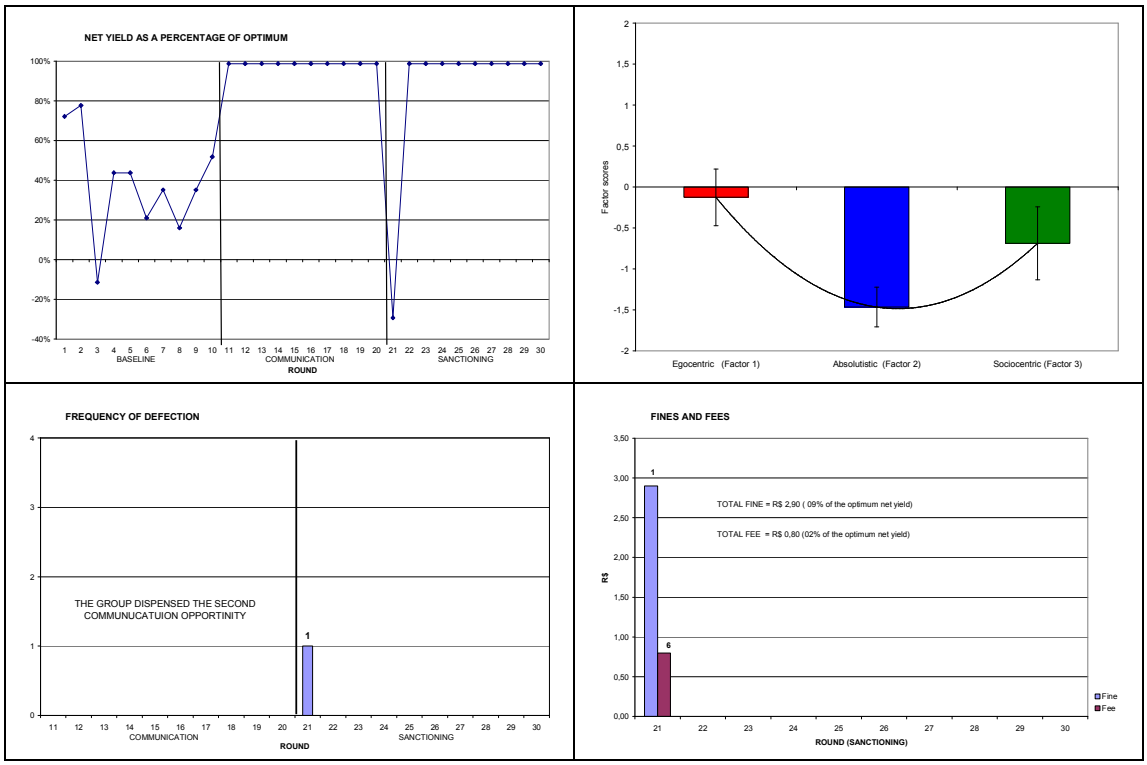


Figure 7.12 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 07.

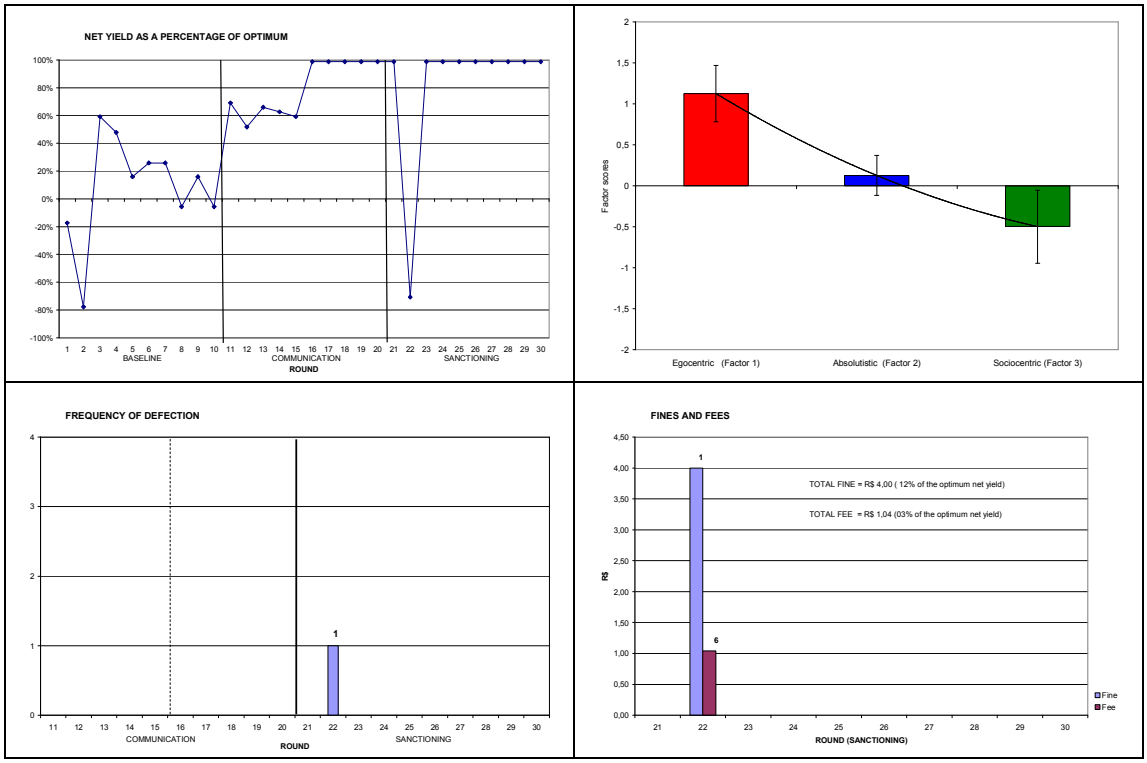


Figure 7.13 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 14.

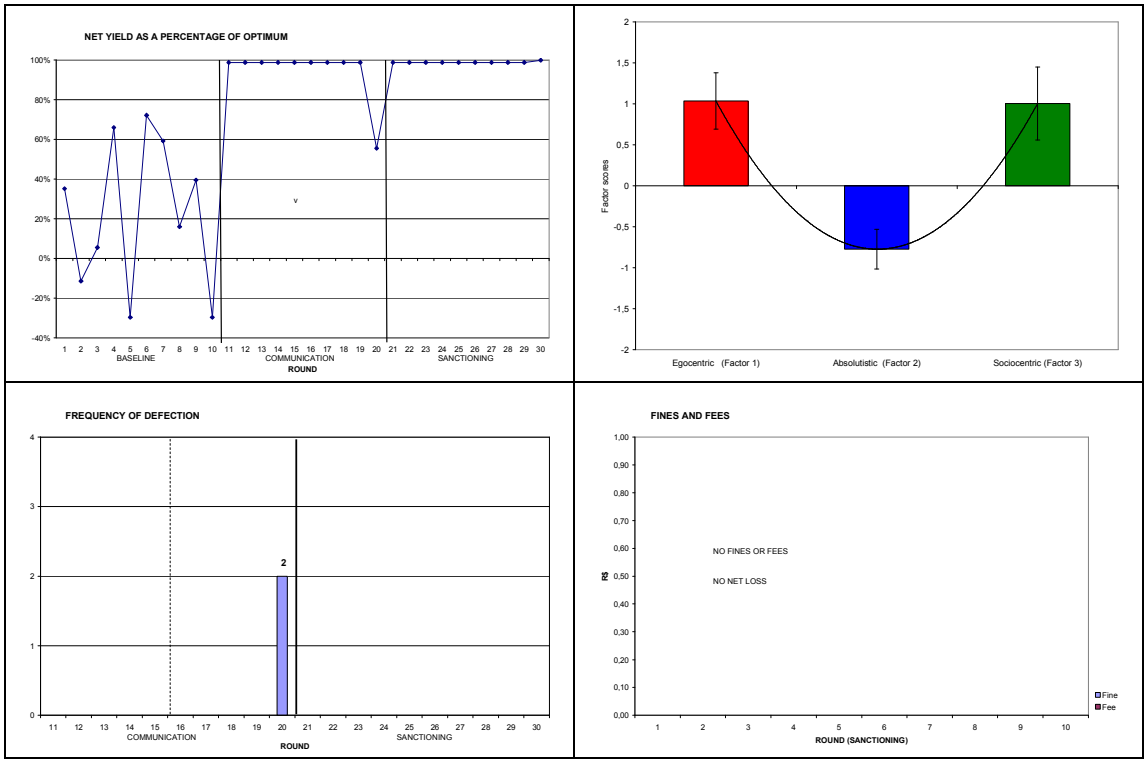


Figure 7.14 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 20.

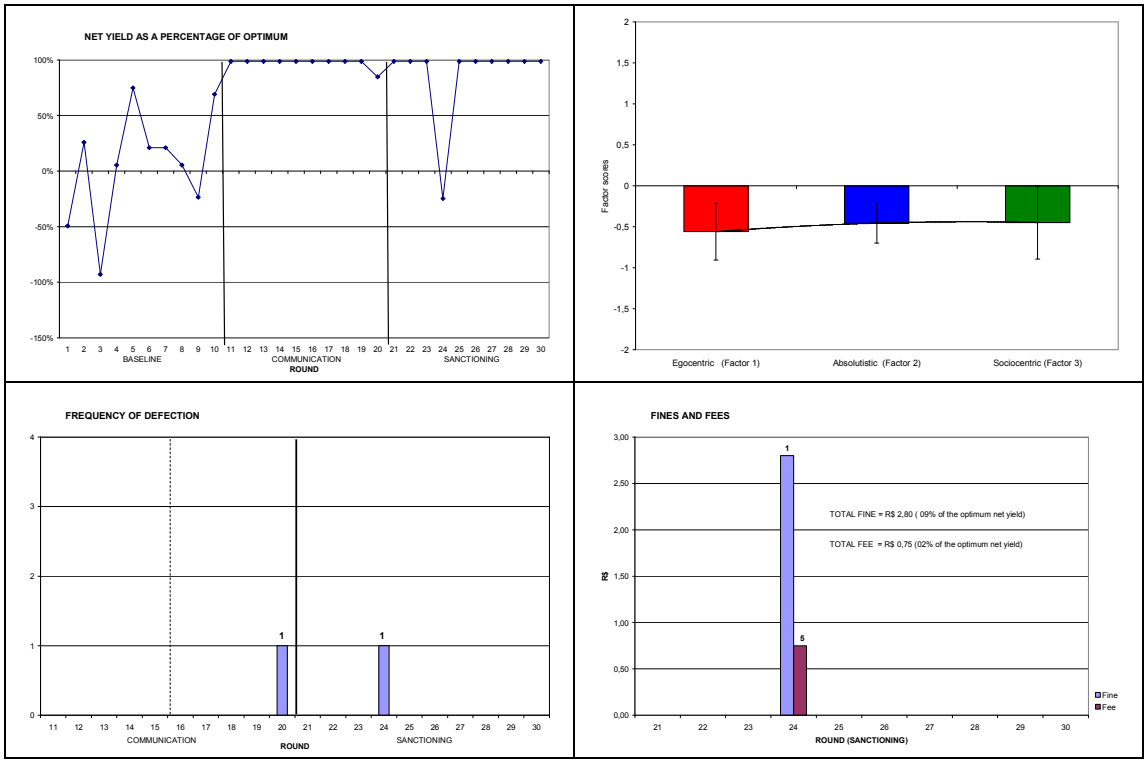


Figure 7.15 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 05.

Apart from the delimited range of defections (3 to 5) there is not much in common among the Groups 8, 12 and 19. Beginning with the appropriation strategies and economics performances, the optimal rotation (mixed) strategy of investing 36 tokens (Pareto) in the Market 2 was found by the participants in the Group 8 (Figure 7.16). However, they decided to set aside the 15th, 20th, 25th and 30th rounds to apply a more flexible rule according to which investments up to 8 tokens (Nash) was allowed. As a result, they achieved 100% of the maximum collective net yield in the rounds where the optimum strategy was followed. This rule was disrespected only in the 19th round, when a player invested 15 tokens and caused the aggregate efficiency to fall down to 91%. The other four defections observed in this group occurred in the rounds 15th, 20th and 25th, that is, the ones in which the rule allowed investing up to 8 tokens. The defection in

the 25th round (sanctioning) was heavily fined by all the other players. The player was fined a total of R$ 4.20 and the total fee amounted to R$ 1.20. As a result the efficiency cut down to -120% of the optimum net yield. In the last round, the 75% of efficiency was due to the "up to 8 tokens" strategy and the fact that two individuals decided to *reduce* their appropriation from the CPR by investing 3 and 4 tokens each. A third player invested 7 tokens while the others followed the rule of investing 8 tokens.

In the Group 12 (Figure 7.17) the participants agreed investing 5 tokens in the Market 2, obtaining 99% of the collective optimum. The norm was followed without defection during the entire communication treatment. Indeed, the participants considered unnecessary to spent additional time discussing. Still, one player defected three times during the sanctioning condition despite of being fined. A possible explanation should associate this behavior with the high scores in egocentrism found in this group, since the uprising against the coercive power of majority is indeed a distinctive trait in the CP (red) value system. Unfortunately, we cannot make this association because precisely this subject is one of the four cases for which we have no psychosocial information.

In the Group 19 (Figure 7.18) the first 10 minutes communication resulted into a rule of investing 7 tokens in the Market 2. This leads to an efficiency of 69% of the optimum, but a defection (15 tokens) at the 14th round cut the collective performance down to 40%. In the 15th round another player skipped the norm and invested 12 tokens, but the earlier defector turned back and invested only 4 tokens. Given that another player also reacted by *reducing* the investment to 4 tokens, the aggregate performance increased to 72% of the optimum. The group then took advantage of the additional five minutes of talk and changed the rule limiting the investments to 5 tokens. Again, the same individual who had defected in the 14th round repeated this action and caused the collective performance to fall from 99% to 85% of the optimum net yield. This time she repeated the action in the next round and was followed by the same individual who had defected before. With two defections of 15 tokens each, the

collective performance dropped to 56%. In this group the sanctioning mechanism solved the problem and no more defections where observed after its institution.

As presented in Figure 7.4, the aggregation of the values systems profiles for these groups resulted into an uninteresting pattern. The combination of positive scores in the three components precludes any noteworthy behavioral prediction. By inspecting the groups separate profiles the only indication suggests simply that the middling factor scores are compatible with the ordinary observation, that is, occasional defections usually discouraged by either the threat of or effective sanctioning (Groups 8 and 19). We will have better indications on that as we proceed.

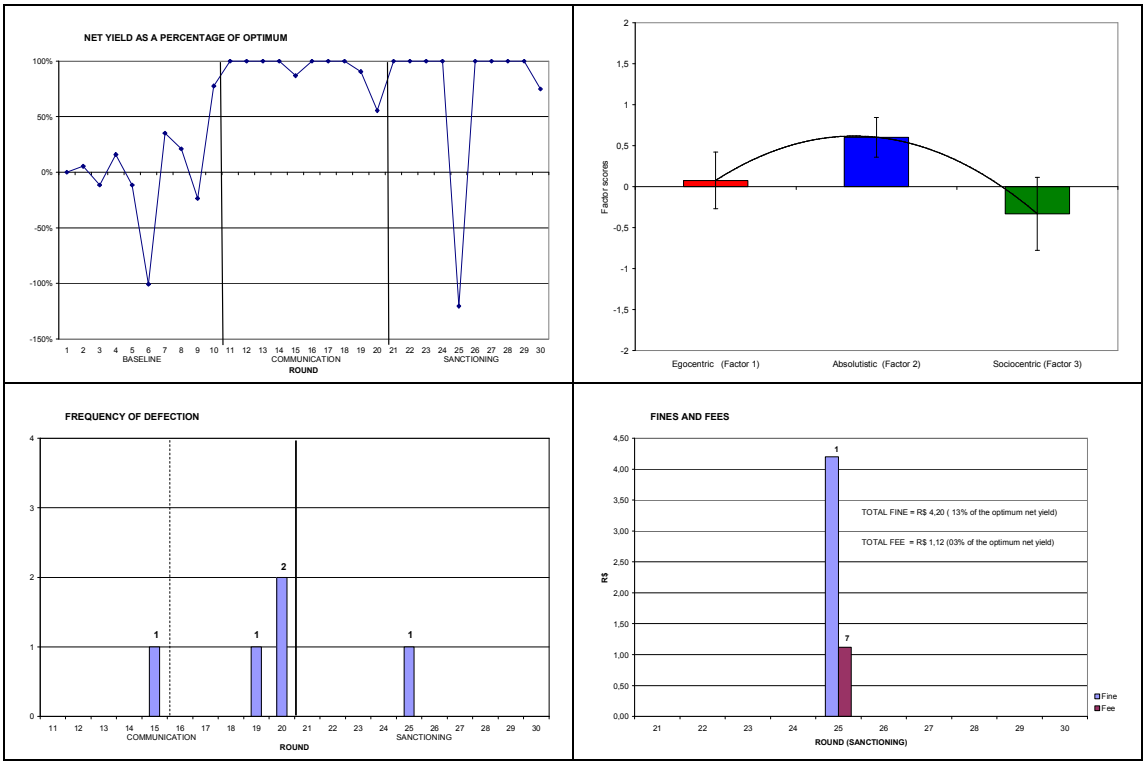


Figure 7.16 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 08.

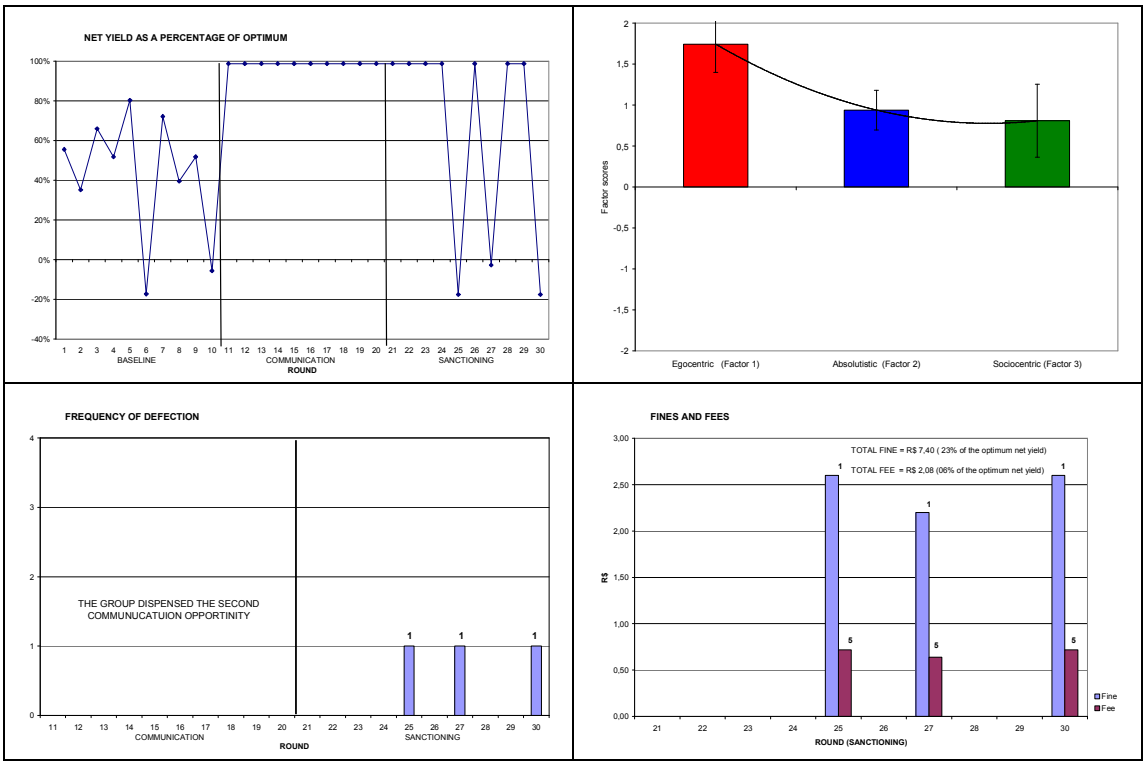


Figure 7.17 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 12.

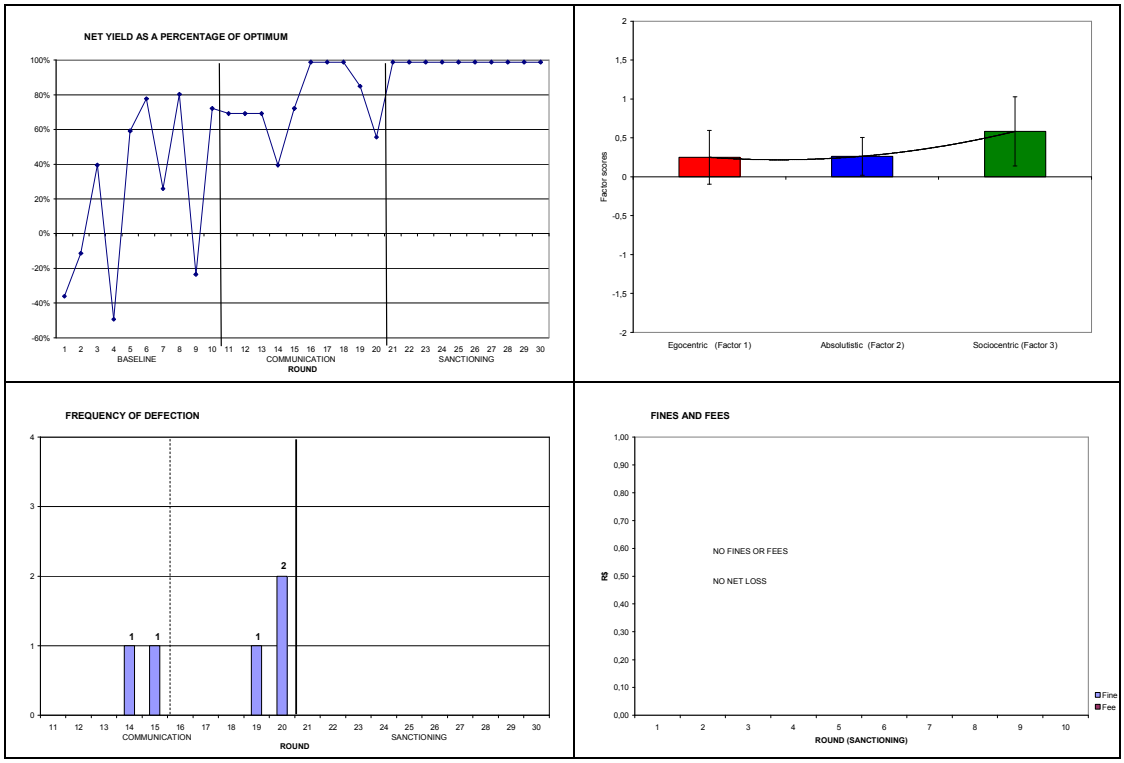


Figure 7.18 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 19.

Indeed, in the next cluster (11, 12 and 16 defections) the mean factor scores for the three principal value systems components run nearer the sample average. In the Group 16 there were 11 defections (Figure 7.19) and in the Group 23 they were 12 (Figure 7.20). In both cases the defections occurred during the communication treatment and the deterioration of the rule following behavior was promptly corrected by the imposition of the sanctioning mechanism. In both groups the participants initially agreed to invest 5 tokens in the CPR. In the Group 23 the participants dispensed the second talk opportunity while in the Group 16 they used it to change the norm and decided to invest 7 tokens in the Market 2 from that point on. This decision, which represents a decline of the collective performance from 99% to 69%, truly upset the proponent of the earlier strategy who refused to keep with the others until the imposition of the

sanctioning mechanism. He had, notwithstanding, already cheated at the 15th round, that is, prior to the changing of the joint strategy. As he continued cheating for more two rounds, the others sustained the norm of investing 7 tokens in the Market 2. Only at the 18th round another player defected by investing 10 tokens and the joint payoff dropped to 40% of the optimum net yield. At the 19th round this new defector resumed to the 7 tokens norm but two other skipped from it; one invested 12 and the other 8 tokens. However, the collective performance rose because the primary defector decided to put all his tokens in the Market 1 and other three players also *reduced* their investments in the CPR (two invested 6 and one invested 5 tokens). In the last round of the communication condition, again three players put only 6 tokens in the Market 2, but the primary defector resumed investing 15, another player invested 12 and other two invested 8 tokens. As a result, the collective performance cut down to 21% of the optimum. As we said, the imposition of the sanctioning mechanism solved the trouble and the collective payoff returned to the 69% corresponding to the 7 tokens norm. The small peak at the 26th round was due to one player who put only 3 tokens in the CPR (Figure 7.19). When asked whether the players considered that a defection they answered that while it was actually a deviation of the norm they were not willing to impose any fine for that kind of choice.

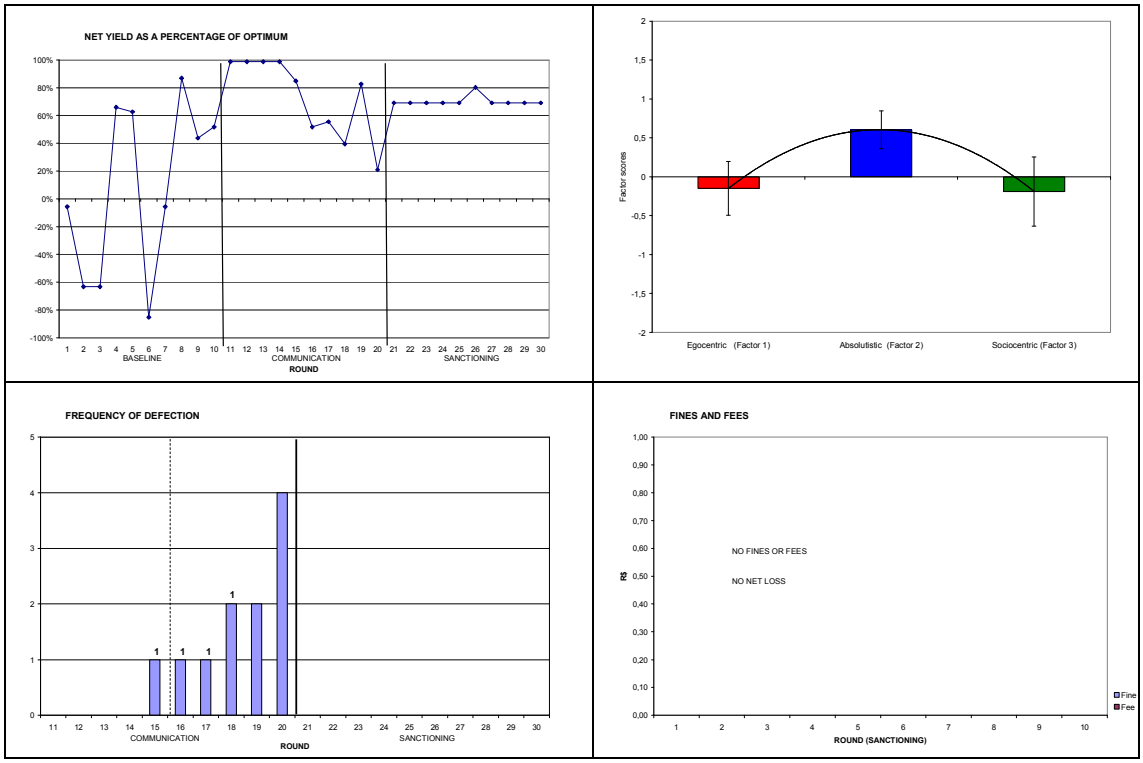


Figure 7.19 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 16.

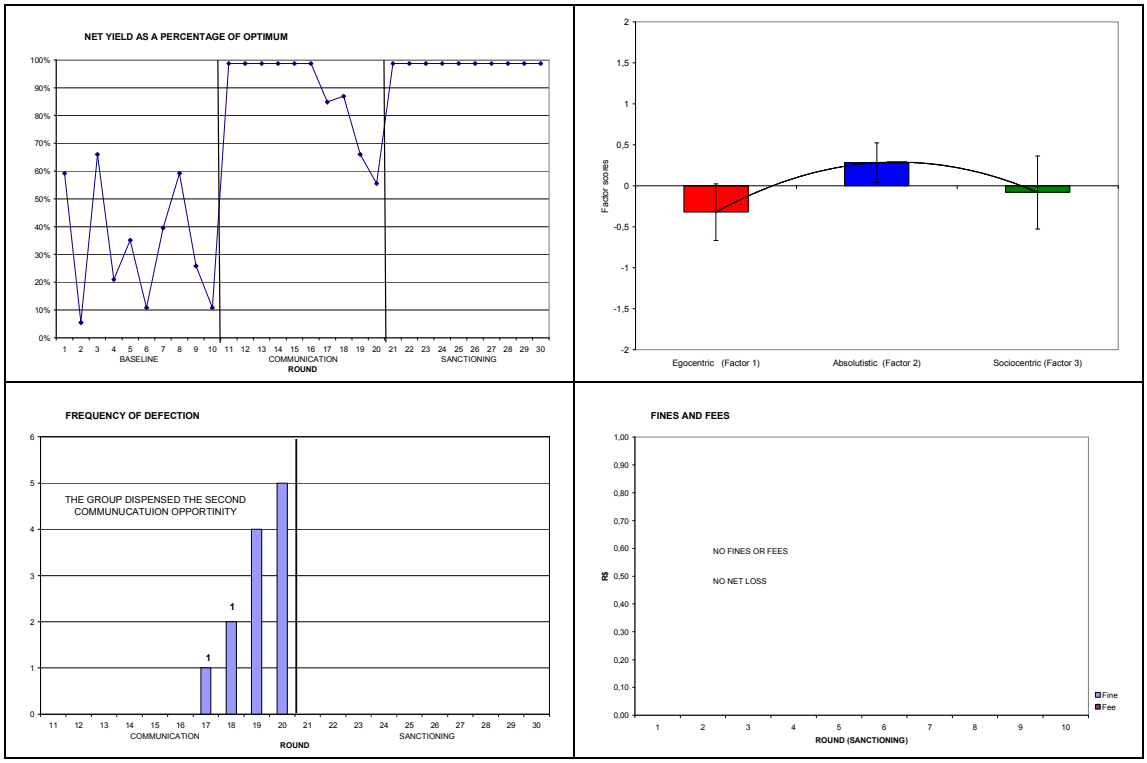


Figure 7.20 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 23.

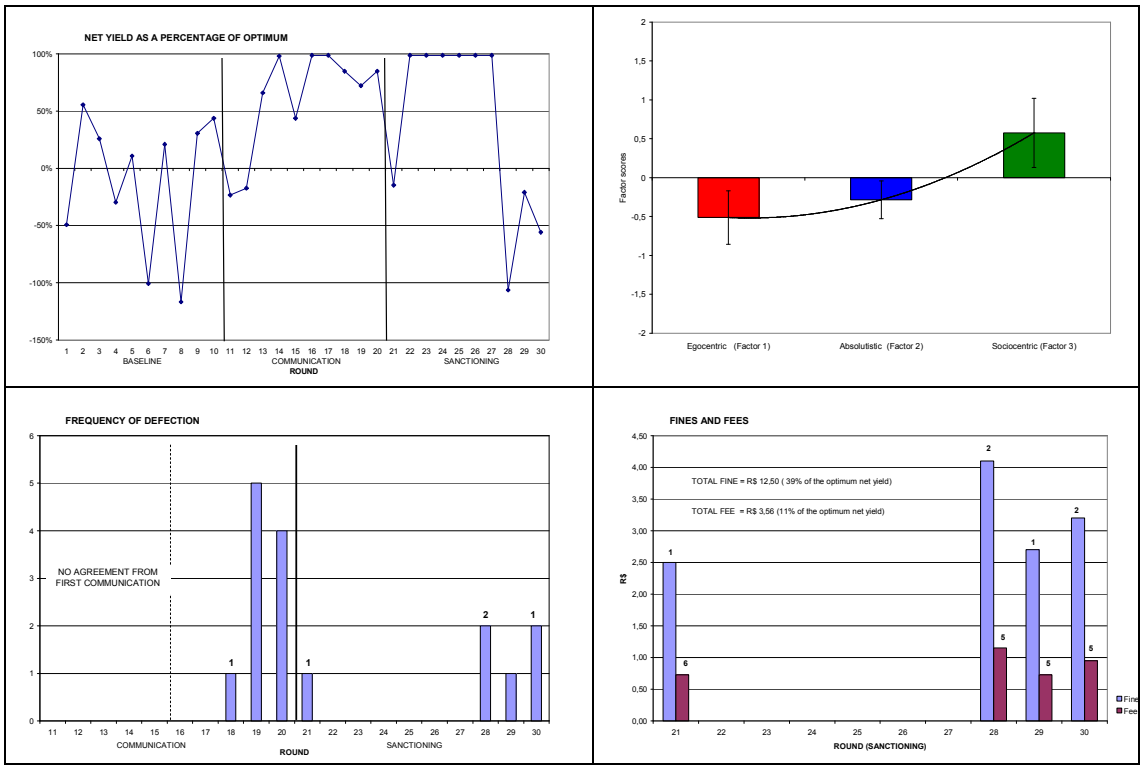


Figure 7.21 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 02.

In the Group 23 (Figure 7.20), the deterioration of the level of cooperation is more typical since the norm was not changed for an inferior strategy as in the Group 16. After 6 rounds without defections, the collective performance fell from 99% down to 56% of the optimum net yield in the path of an increasing abandoning of the norm. As shown in Figure 7.20, the threat of punishment due the imposition of the sanctioning mechanism restored the rule following conduct.

These results approximate the typical expectation when human rationality is identified with a particular form of reasoning, that is, the principle of maximization of the individual utility without much concern for others or collective outcomes. The fact that we are observing this in the domain where the value systems factor scores are near the sample mean suggests that such

somewhat crude view of human nature may actually provide a reasonable approximation of the ordinary human behavior, at the current configuration of levels of existence. We will return to this point as we have finished the description of the performance in the remaining groups.

In particular, the Group 2 presents a less typical situation. The complication begins by the fact that the group did not get to a collective strategy during the 10 minutes communication period. Yet, since they had a chance of sharing their understanding of the game, the average performance increased in relation to the baseline condition, getting as high as 98% of the optimum net yield in the round number 14 (Figure 7.21). The group then used the additional 5 minutes of communication and finish with the previously considered norm of investing 5 tokens in the Market 2. They thus enjoyed a 99% collective efficiency for the two next rounds when the situation deteriorated owing to a defection that caused an almost complete departure from the norm till the imposition of the sanctioning mechanism (only two player stuck with the norm). The efficiency did not fell too much not only because the reacting defections were not extreme (two player invested 10, one invested 8 and two other 7 tokens in the 19th round) but also because one player reacted by *reducing* his investment in the Market 2 to only 3 tokens. He repeated this play in the 20th round. The two players who had put 10 tokens also repeated this level of investment, two others reduced the deviation from the norm by investing 6 tokens and the other resumed to the norm in the 20th round.

After a defection, promptly fined, in the first round of the sanctioning treatment, things seemed to return to the ordinary expectation discussed above. However, at the 28th round the same subject decided to defect again and insisted in continuing doing so until the end of the game, despite the net lost due to the fines. Moreover, in the 28th round another player also put 15 tokens in the Market 2 and a third one invested 10 in the last round. Both of them were fined. As a result the economic performance fell as down as -120% of the optimum net yield in the 28th round and remaining negative until the end of the game.

The post game interview revealed a disagreement about the validity of the norm and, by extension, concerning the legitimacy of the fines imposed on defectors. The subject argued that the norm settled between the 15th and 16th rounds predicted its own extinction were it not followed in the next rounds of the communication treatment. Since, as reported, the norm was overtly abandoned after the 18th round, the subject believed that there was no more any rule to be enforced. He then invested 15 tokens in the first round of the sanctioning treatment and was surprised by being fined. The shock made him resume to the norm but eventually he "realized", as he said, that the fines were "illegal". He then decided to seek his own appropriation strategy in the hope of not being forced to pay for the "illegal fines".

The records of the communication periods showed that he was right concerning the content of the discussion. Moreover, the instructions explicitly pointed out that the sanctioning mechanism were offered as an instrument for rule enforcement in the cases in which the group had settled one, and that the mechanism would not be operational otherwise. This was put this way as an attempt of not inducing the participants to believe that they necessarily should conform to whichever joint strategy, since the logistic conditions of our study forced us to reveal the three existing treatments in advance. Nonetheless, we explained, whether or not the group has a norm is something up to the participants themselves. We had not taken part in the discussion and could not decide about the "legality" of the sanctions imposed. Presumably, we said, if the others are sanctioning they probably agree they have a norm. We thus subtracted the fines from his payoffs under protests.

In the next cluster, the Groups 10 and 11 present 19 defections each and in the Group 18 there were 22 defections. In all of these groups the participants found the near to optimum joint strategy, which get to 99% of the maximum net yield. In the Groups 10 and 11 it were accepted to limit the investment in the Market 2 to 4 tokens while in the Group 18 the norm was set at 5 tokens. In the Group 10 the norm was sustained along the first five rounds of the communication treatment and, while the participants reiterated the agreement

during the second talk, the first defection occurred in the 17th round. The next round three other players reacted by putting 10 tokens in the Market 2 as well as did the subject who have cheated before. In the 19th round he and the two others reduced their investment to 9 tokens while the third repeated the 10 tokens in the Market 2. At the same time two other players got out of the norm; one investing 6 and the other investing 8 tokens. In the last round of the communication period one player resumed to the norm but another skipped out and put 7 tokens in the Market 2. The total investment amounted to 60 tokens so that the group performance fell down to 56% of the maximum net yield. As shown in Figure 7.22, there were two more defections (15 tokens each) during the sanctioning treatment. Both of them were not so heavily fined in comparison which other cases we have seen before. The first defection, together with the fines and fee, led the performance down to -17% of the maximum net yield; the second was lighter fined and the performance returned to 25% of the optimum. After that, the joint strategy was resumed and the collective performance returned to the 99%.

In the Groups 11 (Figure 7.23) and 18 (Figure 7.24) the pattern of the instability suggests that the group did not achieve a consensus during the 10 minutes of communication. In the Group 11 the norm was skipped in the second round after the talk whereas in the Group 18 it was just in the first round. In both case the rule was initially broken by only one subject, but this caused an avalanche of defections which led the performance drop down to 26% of optimum the in the Group 11 and to 40% in the Group 18. The joint strategy was resumed in both groups after the second talk but it was also abandoned immediately before the imposition of the sanctioning mechanism, that is, at the 20th round. In the Group 11 three players cheated and in the Group 18 they were four. The overinvestment in the first case led the performance fall down to 66% and in the second it got to 26%. These defections at the last round of the communication period express an opportunist behavior that was attempted again at the last round of the game. Still, as commented before, this stratagem never paid since that last-round defections were regularly fined. The participant in the Group 11 was fined a total of R$ 3.90 while that of Group 18 paid R$ 1.70.

Including the fees, the collective performance get down to -41% in the Group 11 and to 27% in the Group 18. The defector in the first round of the sanctioning treatment (Group 18) was fined in R$ 2.60 and the aggregate performance was of -18% in that round (Figure 7.24).

As for the values systems profiles, the three groups have not much in common. The aggregate pattern shown in Figure 7.4, particularly for absolutism (Factor 2) and sociocentrism (Factor 3), results from a balancing of positive and negative deviations around the sample mean. Tentatively, we observe that the opportunism just reported appeared in the groups were sociocentrism scores lower (Groups 11 and 18). Beside, this also conveys an indication of the affinity with the strategic-materialistic wave of existence. Also, the pattern of the defections in the Group 10 is different from those observed in the other two groups. In the Group 10, where we have a combination of high egocentrism and low absolutism, the pattern shows an abandonment of the rule following behavior. Conversely, where either absolutism is higher (Group 11) or egocentrism is lower (Group 18) the pattern shows a convergence from an initial disagreement concerning the rule toward a more stable conformance, despite some individual opportunism. Finally, in each of these three groups there was only one participant who stuck with the rule from the first to the last round regardless the turbulence caused by the others. As we shall see, this "never cheating" behavior correlates positively with the scores in sociocentrism, as it should be, since, although the rule following behavior is typical of the absolutistic value system, DQ is also typically punitive. In the experimental setting, prior to the institution of the sanctioning mechanism, reciprocal defection is the only way of expressing reproving toward cheating.

Once again, the formal test of the conjectures we are rising in order to interpret the collective performances will be accomplished in the next section. To complete description, let us move to the last cluster of groups.

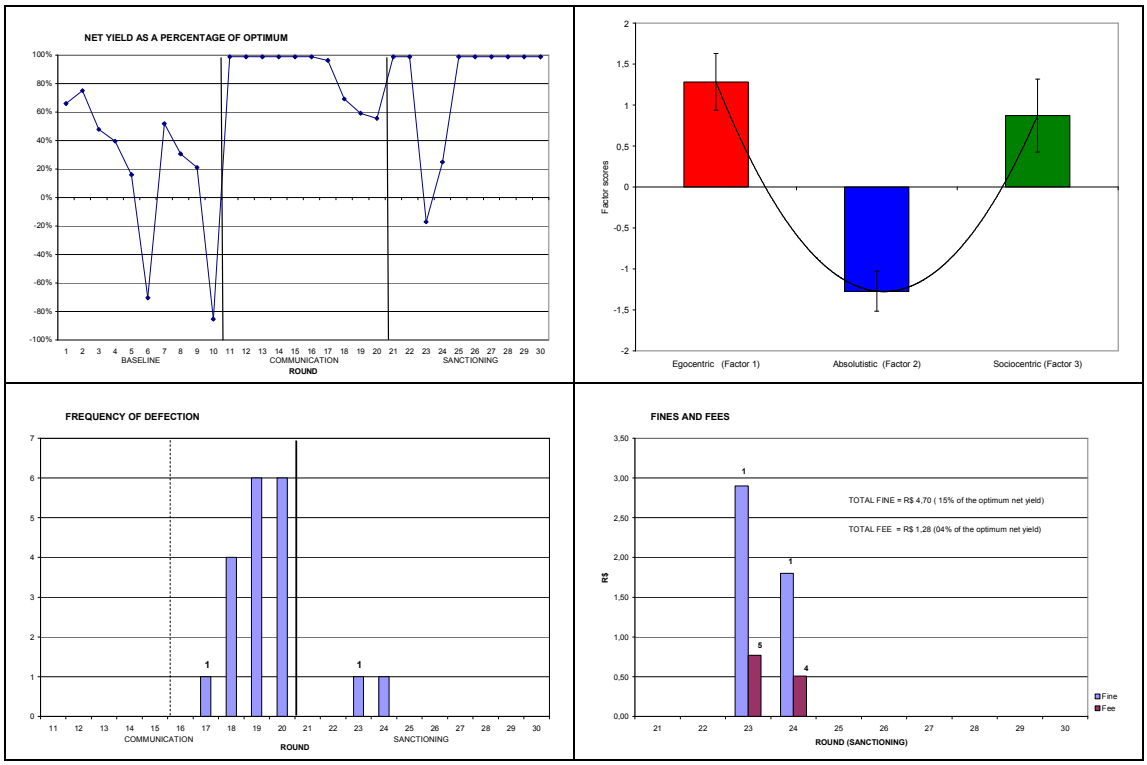


Figure 7.22 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 10.

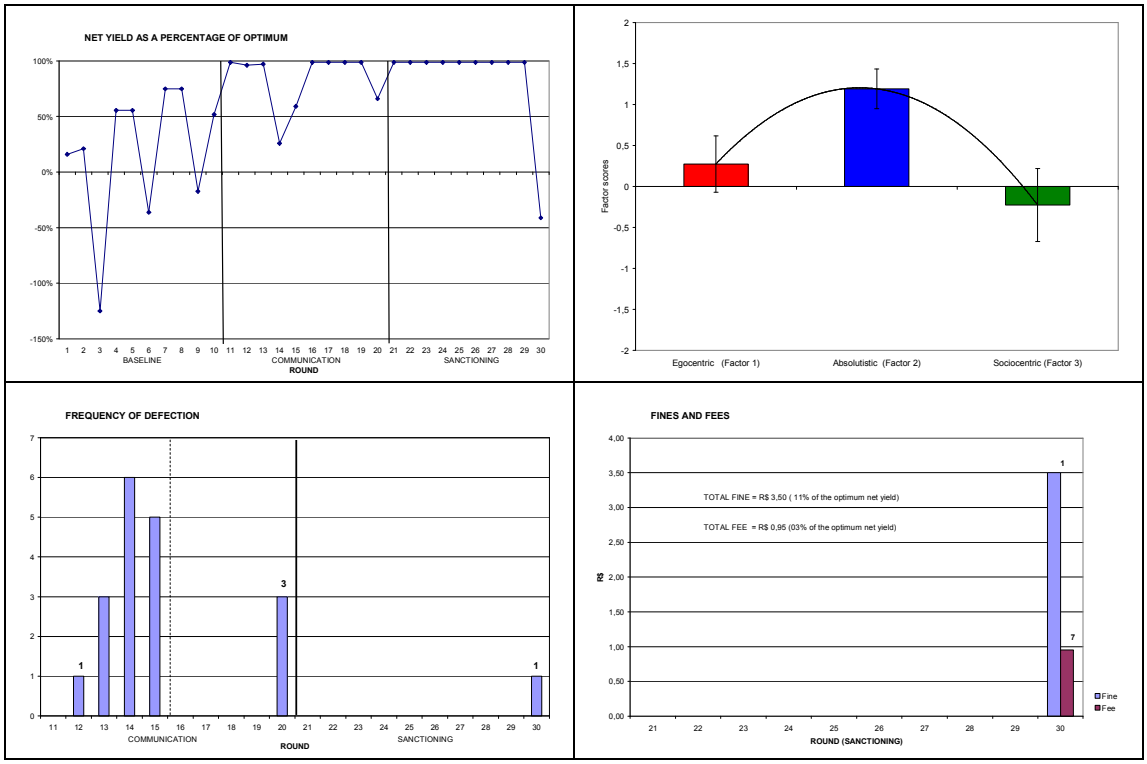


Figure 7.23 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 11.

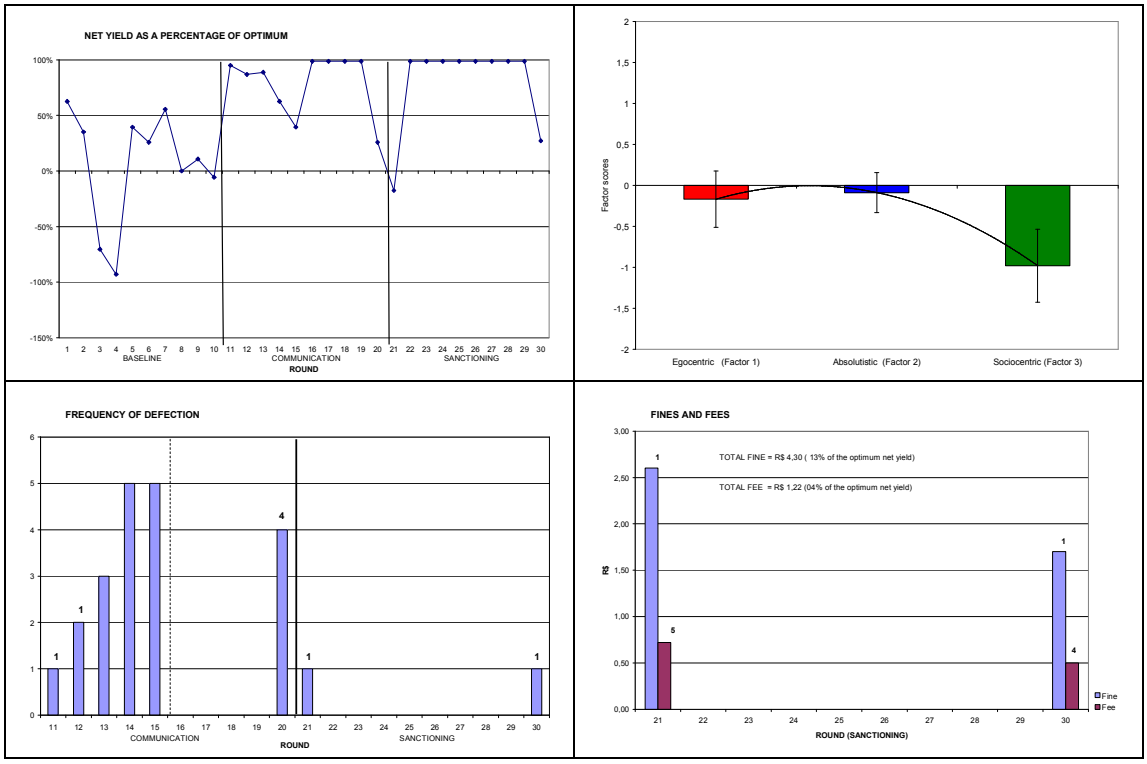


Figure 7.24 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 18.

The last cluster brings together the groups presenting the highest levels of defection. Though we have grouped them so to better expose the most general association between the frequency of defection and factor scores, the occasions observed in the Groups 21 and 22 are noteworthy for a separate consideration. These are the groups in which we observed the mentioned "institutional battles". We shall thus let them to the end.

The three other groups tell a story similar to the ones just reported, beginning by the heterogeneity of the value systems profiles. The participants in these three groups found the near to optimum (99%) joint strategy and set up the norm of investing 5 tokens the Market 2. Still, the players in the Group 24 agreed also with a "free play" at the 20th round, prior to the imposition of the sanctioning institution.

In the Group 9 (Figure 7.25), one only defection in the second round after communication (the 12th round) caused the avalanche that pushed the performance down to -6% of the optimum. The cooperation resumed after the second talk but a new defection by a different player reinitiated the process. This time the deviations were lighter so that the collective efficiency fell just to 48% of the optimum net yield. Characteristically, if egocentrism is not too prevailing, the imposition of the sanctioning mechanism sufficed to put an end into the turbulence.

The pattern of defections in the Group 15 (Figure 7.26) resembles the one saw in the Group 10, above (Figure 7.22). This time, however, we cannot rely on the above than average scores for egocentrism to explain the deterioration of cooperation. Still, the scores for sociocentrism (Factor 3) are lower than average, suggesting a higher influence of the ER achieving-self in this Group. This may also be related with the opportunistic last-round defection observed in the Group 15. In terms of economic performance, the defections during the communication treatment led the efficiency down to only 16% of the optimum net yield. In the last round, the weight of fines and fees pressed the performance down to -52% of the maximum net yield.

Finally, in the Group 24 (Figure 7.27) we can say that the norm was almost completely ignored. A defection in the first round following communication started an avalanche that the second talk was not able to stop. Of course, we don't mean that the second talk had no effect at all. In fact, after the additional five minutes of face-to-face communication 7 players resumed to the norm, but one invested 11 tokens in the CPR. In the next round he increased the investment to 15 tokens and five other participants reacted to the earlier defection by investing more than agreed upon. At the 18th round the aggregate investment amounted to 72 tokens and the collective performance got to 0% to the optimum net yield. The absent of defections marks in the 20th round is due to the nature of the norm itself, which settled this round for a free play. In this round the aggregate investment was of 61 tokens, resulting into a performance of 52% of the optimum. The imposition of the sanctioning mechanism again solved the

problem, while it was necessary to fine the individual who have initiated the second avalanche of defections and that insisted in cheating again in first round of the sanctioning treatment. He invested 15 tokens in order to appropriate extra R$ .93 from Market 2, but was fined a total of R$ 3.50. Also, though the instructions drew attention to the character of the institutional change, establishing that defections in the communication period were not subject to sanctioning, we received three impositions of fines (totalizing R$ 1.70) upon the individual who have cheated in the first round after the 10 minutes communication period. As explained in the previous chapter, this particular instruction was necessary in order to preserve the non enforceable character of the joint strategy during the communication treatment, since the participants knew that a sanctioning mechanism might be operational from the 21st round on. Again, this was due to the logistic situation of our experimental condition.

Concerning the connections with the value systems, the pattern drawn by high egocentrism, low absolutism and high sociocentrism suggests a potentially explosive combination, for the low absolutism together with high sociocentrism point to an unwilling disposition to discipline egocentric rebelliousness. Still, the mentioned explosiveness requires that this pattern be expressing a combination of individuals centered at those levels of existence instead of an aggregate pattern perhaps associated to not well defined individual psychosocial profiles. Also, such pattern is clearer in the Groups 10 (Figure 7.22) and 20 (Figure 7.14) than in the current Group 24. Here, in fact, the absolutism runs around the average. We suggest that the encounter of the egocentric and the sociocentric waves may explain the difficult of get to a consensus about the joint strategy during the communication periods while the ordinary level of absolutism helps to understand the avalanches that followed primary defections.

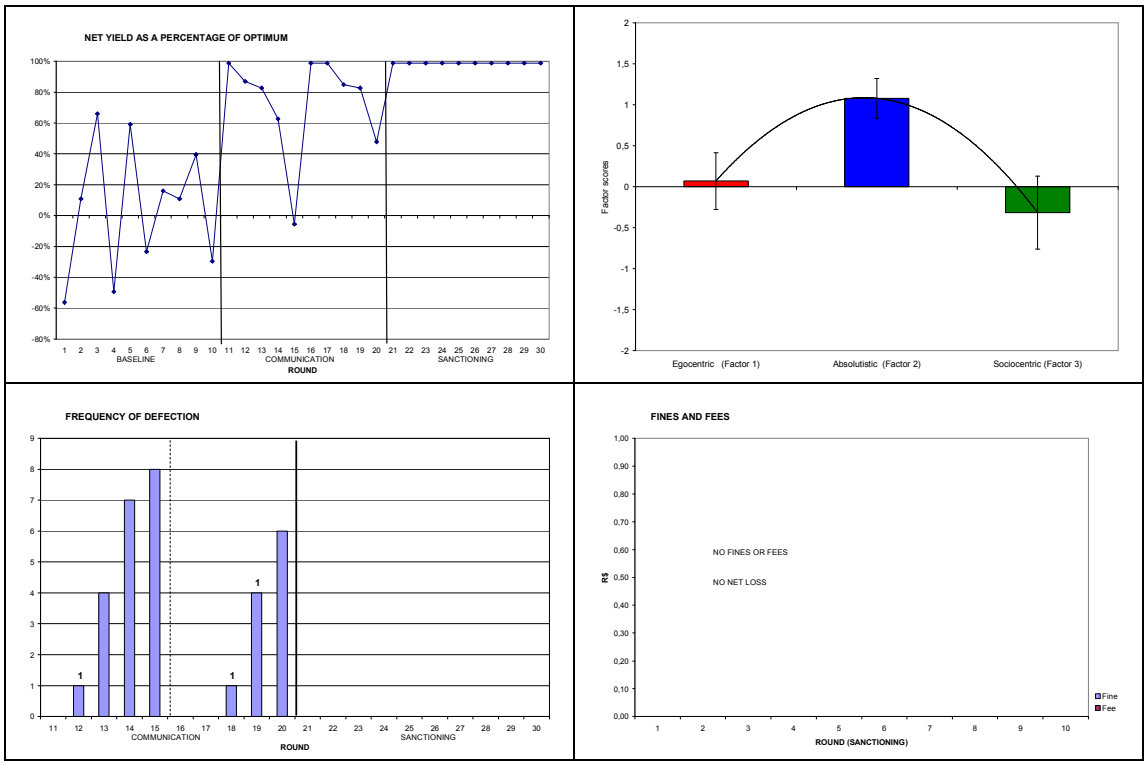


Figure 7.25 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 09.

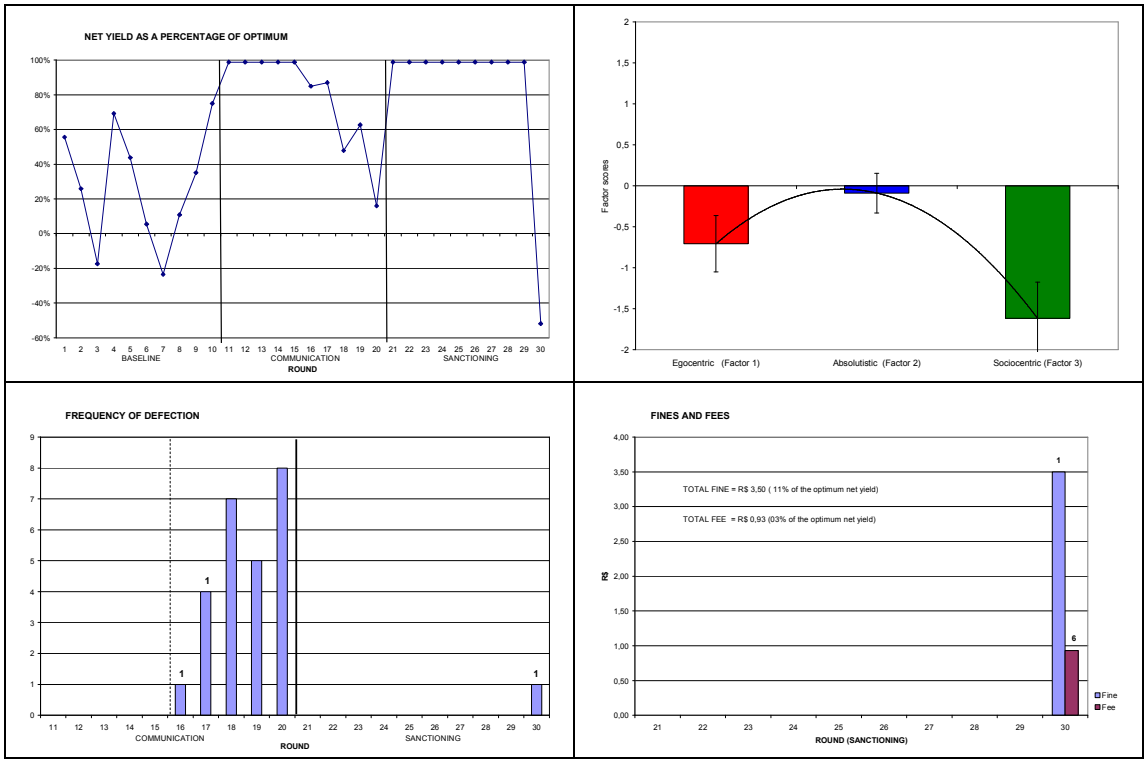


Figure 7.26 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 15.

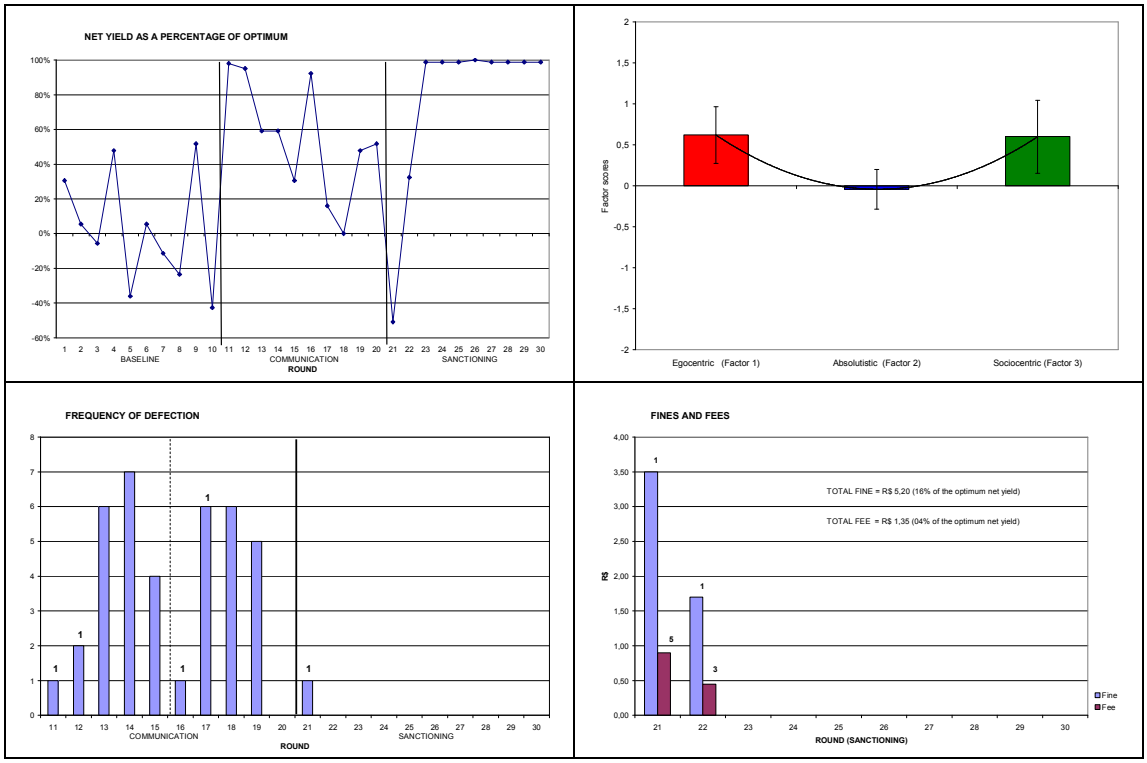


Figure 7.27 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 24.

We now address the mentioned "institutional battles" observed in the Groups 21 and 22. We saw that something of the kind occurred also in the Group 2 where one participant regarded the fines imposed on him as "illegal". As mentioned, on the occasion we augmented with him that while the instructions indeed established that the sanctioning mechanism would not be operational if the group had not set a collective norm, we (the researchers) could not decide whether the group had or not a norm. Such interpretation should be a matter up to the participants themselves and, in that case, the group certainly believed they had a norm.

In the Groups 21 and 22, the issue of the participants' autonomy to decide upon their own agreement unexpectedly extended to the use of the sanctioning mechanism. In each of these groups there was one participant who

strongly refused the very idea of settling a joint strategy. So, as shown in Figures 7.28 and 7.29, the ten rounds of the communication treatment are full of defections in both groups. Still, in the Group 21, when it came to the sanctioning treatment, after the participants have read the instructions, we were asked about the issue of the effectiveness of the mechanism. As usual, when asked on this point, we answered that they should decide by themselves whether or not they want to use the mechanism, and that the issue pointed out in the writing instructions just meant that they were not supposed to necessarily get to any joint strategy. But as we said that, the subject who was foreseeing what should probably happen if he continued defecting from that point on, asked us whether he also could use the mechanism to eventually hit back on others. On this we responded that the sanctioning institution was supposed to be a mechanism for norm enforcement, and that the players who were conforming to the rule were supposed not to be fined, as well as nobody was supposed to be fined in case the group had not instituted any norm to regulate the investments. Still, we added, we were not the interpreters of the rule so that they themselves should decide how to use the sanctioning institution.

The subject looked as having got the answer he was waiting for. Thus, in the first round of the sanctioning treatment, after investing 15 tokens in the Market 2, he presented his "sanctioning card" imposing the highest fine on all the other players. We then checked with him whether he was attentive to the fact the total fees corresponding to the fines he was willing to impose would be subtracted from his final payoff. As he ascertained he was aware of that and sustained his decision, we registered the fines and fees. As indicated in Figure 7.28 the battle of defections and fines continued all through the 10 rounds of the sanctioning treatment. In the total, this participant spent R$ 9.80 with fees and was fined a total of R$14.50. His final total (Market 1 plus Market 2) payoff amounted to only R$ 8.12 whereas the average payoff ran around R$ 29.34 (we paid only 1/3 of the total payoff plus R$ 5.00). The aggregated outflow with fines and fees amounted to R$ 79.04 or 244% of the maximum net yield in the 10 rounds of the sanctioning condition. The lowest point in Figure 7.28, at the 27th

round, indicates a collective appropriation corresponding to -415% of the optimum net yield in that round.

In the Group 22 there was no open question concerning the proper use of the sanctioning mechanism, and while it was observed a defection, promptly fined, in the first round of the sanctioning treatment, things seemed to return to the normal from the second round on. Surprisingly, though, in the 26th round we were presented a "sanctioning card" imposing a fine on the other seven players. The subject showed us the card and privately asked us whether he could do that play since there was no deviation from the 5-token rule registered on the board. Owing to the events in the Group 21, which have happened some sections before, we felt compelled to accept his play. We then announced to the group that they should fill out the "sanctioning card" and this induced another subject to impose a new fine (R$ .10) on the player who had cheated at the first round of the sanctioning treatment. The other participants just left their cards blank. But as they received the R$ .80 fines the battle of fines popped out, creating an odd situation where the payers continue to fine one another without any apparent reason, for they all continued to play as established by the joint strategy.

The post-game interview revealed that the player who started this battle in the Group 22 felt coerced to behave against his "principles". According to his view, the formation of a joint strategy had the character of collusion and was contrary to the "spirit of the free enterprise".

As explained in the last chapter, we followed the orientations found in the literature in the sense that the players were not given clues neither to naturally occurring parallel decision environment nor to what we expect of them. In particular, we did not call Market 2 a fishery or forestry and consciously neutralized the expectations concerning what they should or shouldn't do during the communication period. Still, we did refer to the alternatives as *markets*, and used the term *investment* to refer to the choice behavior during the *game*. Moreover, in order to be sure that the participants had clearly understood the distinct characteristics of the two "markets", we actually explicated that Market 1

offered a fixed and *secure* return while the returns from Market 2 were variable and *uncertain*, due to the interdependence of the individual decisions.

In fact, the issue of non-contextualization seemed to us as an elusive aspiration, for the terminology just mentioned – together with the heading of Applied Economics Research – already create a general context that may rise up issues of ideological nature. So, while the majority of the participants regarded the formation of a collective investment strategy as mutually advantageous, this participant strongly refused this "solution". He pleaded with others that those who were not willing to face the *risk* inherent to Market 2 should invest their tokens in Market 1, the *secure* alternative.

When the sanctioning mechanism was imposed, this "risk-pro investor" naturally felt coerced by the others. He presumably consumed himself during the first five rounds of the sanctioning treatment until he finally realized a possible way out of his frustration, by fining his oppressors. He continued to play according to the rule in the hope of not being identified, for the individual decisions were been publicized on the board. Unfortunately, owing to the limitations of the experimental environment (the class room) the other players noticed when he consulted us about his plan and easily deduced the origin of the fines.

As for the economic performance, this battle in the Group 22 cost a total of R$ 37,78 (117% of the maximum net yield in the 10 rounds). In the lowest point, at the 27th round, the joint appropriation corresponded to -238% of the optimum. In the last round the players stopped fire and the collective payoff returned to 99% of the maximum net yield (Figure 7.29).

Concerning the value systems, the aggregate profile of the Group 21 indicates predominance of absolutism (Factor 2) whereas in the Group 22 the chiefly characteristic is the negative mean factor score for sociocentrism (Factor 3). Both profiles are generally consistent with high frequency of defection.

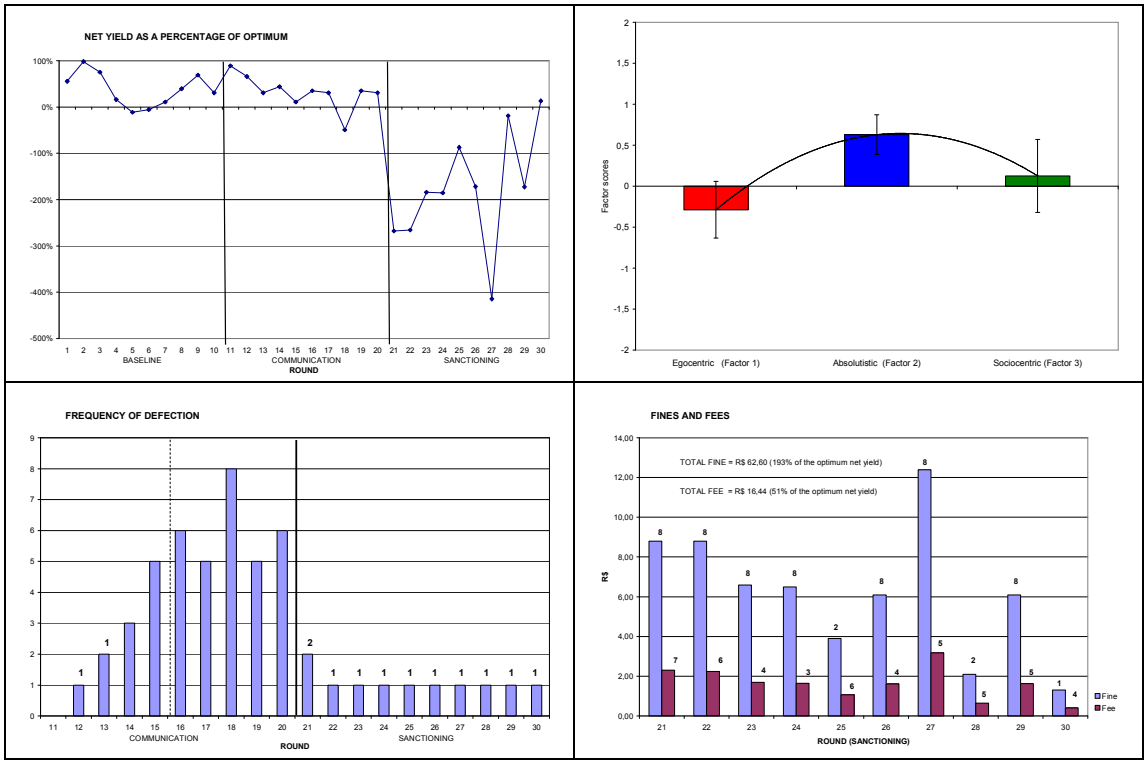


Figure 7.28 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 21.

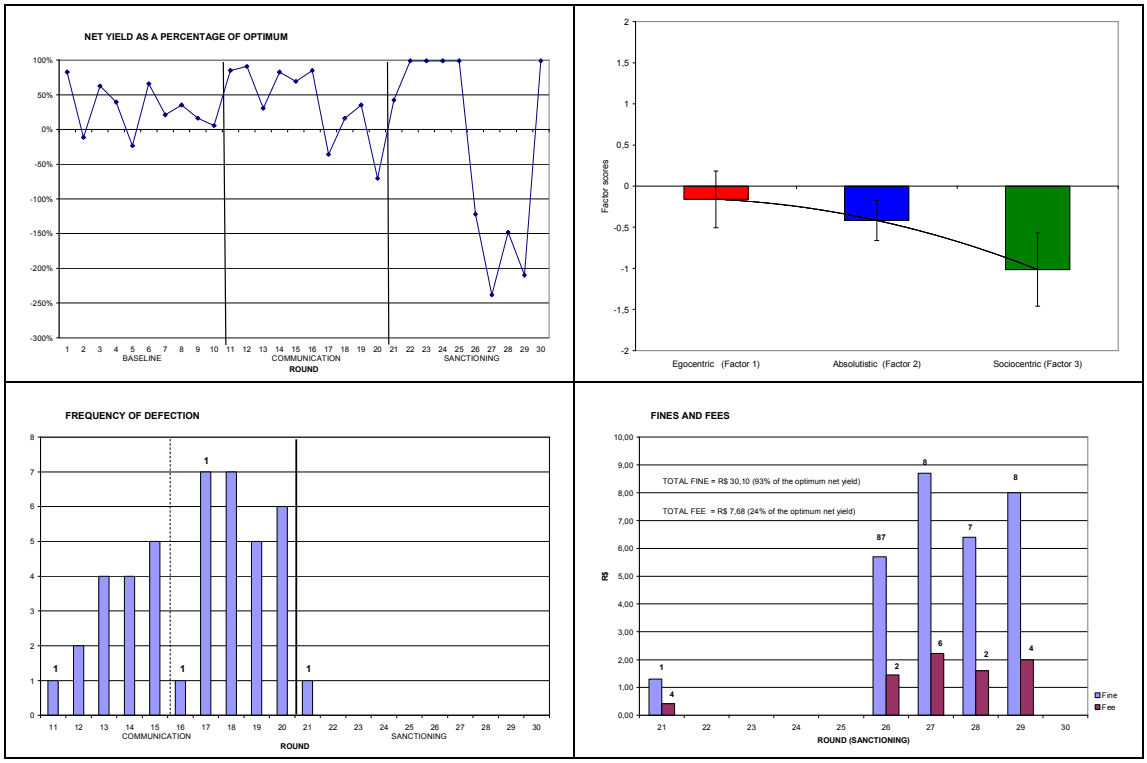
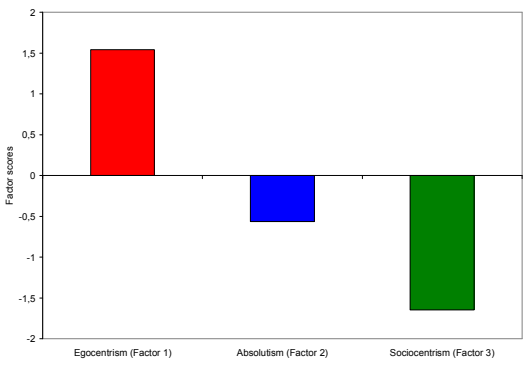


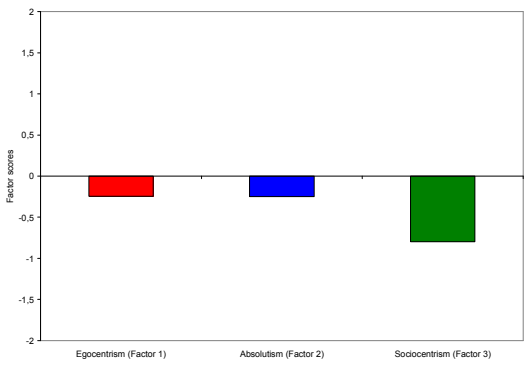
Figure 7.29 – Efficiency of the collective appropriation, group's average psychosocial profile, frequency of defections, and expenditures with fines and fees – Group 22.

Still, given the somewhat unusual behaviors of the players who initiate those "institutional battles" we present in Figure 7.30 their individual scores for the three principal components. The psychosocial profile of the player in the Group 21 (Figure 7.30, panel [a]) helps much to explain the overwhelming expression of his individual will and his strong repudiation of the coercive power of the majority (egocentric drives) together with his indifference in relation to the opinion of others (negative sociocentrism). The impact of this kind of personality into a discussion group marked by positive degree of absolutism (see Figure 7.4) can explain the difficulty of achieving an agreement toward a joint strategy as well as the avalanches of punishing defections observed in the Group 21 during the communication treatment.

The situation in the Group 22 cannot be explained by this explosive combination of egocentrism with absolutism, though. Yet, the Group 22 present one of lowest aggregate scores in sociocentrism (Figure 7.4). The profile of the unusual player in the Group 22 (Figure 7.30, panel [b]) shows that his individual characteristics conform to those of the whole group. The negative sociocentrism, which means, moreover, affinity with the ER strategic thinking, is consistent with his decision of fining the entire group while attempting to hide his identity during the sanctioning treatment. The near to average scores for egocentrism and absolutism indicates that the ER value system indeed predominates.



(a) Atypical player in the Group 21



(b) Atypical player in the Group 22

Figure 7.30 – Two atypical players.

7.5.3. Disentangling the connections

Except for the references concerning the unusual occasions in the Groups 21 and 22, just discussed, the analysis of the relationship between the value systems and the behavioral responses in the CPR experiment have relied thus far exclusively on aggregate psychosocial profiles. We have first averaged the individual profiles to produce a synthetic psychosocial characterization of each group participating in the experimental sections. We then clustered these groups

according to approximate ranges of defection rate and took the mean scores of the individuals in those clusters to generate the corresponding value systems profiles. Using this procedure we were able to indicate the general agreement between the defection rate and that which we should expect from the Levels of Existence theory (Figure 7.4). The observations, moreover, already offer a first indication that the disposition to cooperative increases together with the levels psychosocial development, out of the *decreasing egocentrism*.

As commented, this procedure is limited because the relationship between the behavioral responses and the value systems profiles, however being ultimately a collective phenomenon, is strongly affected by the particular combination of the individual characteristics present in the social space of each group. When the groups are relatively homogeneous in relation to the value systems profiles this effect is minimized. However, when the variability is large, the explanation of the collective performance derived from the aggregate profiles may impart some spurious associations.

We now address this limitation by disaggregating the individual information concerning both the value systems profiles and the behavioral responses in the experimental settings. We first explore the data set using bivariate correlation analysis and then linear multivariate and binomial logistic regressions analyses. As for the differences observed during the sanctioning treatment in the Groups 21 and 22, we repeated the analyses both including and excluding these groups from the sample. It was found that, while the statistical significance of some specific correlations and coefficients does change, the overall conclusions drawn from the whole set of analysis remains pretty much the same. We thus decided to preserve the full experimental information except for excluding the outlier defections during the sanctioning treatment in the Group 21 (the nine single defections shown in Figure 7.28). So, the following results are grounded on this information base. The correlation matrix produced when the Groups 21 and 22 are excluded is presented in the Appendix.

7.5.4. Bivariate analysis

Table 7.9 presents the Pearson bivariate correlations between the values systems/levels of existence and the behavioral responses including: (1) the appropriation strategies in the three treatments, (2) the wiliness to run off the joint strategy, and (3) the wiliness to pay for enforcing the joint strategy. Our interest is to check for the consistency of these behaviors in the experimental settings with the theoretical expectations resumed in our set of hypothesis. Given the fluid and flowing nature of the interior (psychological developmental) phenomenon under investigation, the correlations that can be found on basis of two or three snapshots of its outward appearance in the experimental conditions are unsurprisingly low to moderate. Still, a reasonable number of them are statically significant and, as we shall see, they are in wide agreement with the theoretical expectations.

Before begin, we notice that the variable *fines*, measuring the total payment with fines, imparts only primary and punished defections during the sanctioning condition, since in this institutional setting participants do not rely on secondary defections as a form of norm enforcement. On the other hand, the variably *frequency of defection* conveys other types of defection, such as primary, secondary imitative and secondary punitive defections. Also, the level of investment in the baseline conditions is more expressive of the overall character of the appropriation behavior, whereas the investments in the communication and sanctioning conditions convey somewhat redundant information concerning the trend to run out of the norm. The strong reciprocity behavior is approximated by the wiliness to pay the fees in order to sanction noncooperators. We remind that "strong reciprocity" is not the only interpretation for this behavior, for even if sanctioning is costly, the underlining motivation may also include some "taste for punishment". Such possibility should be taken into account when the behavior is associated with the CP or DQ value systems.

7.5.4.1. The ER system reveals its go-getting character in the unregulated and nonbinding communication settings

Commencing with the appropriation behavior, the only significant correlations in the baseline condition are with the ER (orange) value system. These correlations tell us that the higher the identification with ER (ER_most), the more the go-getting character of the appropriation behavior under the unregulated setting. This relationship is confirmed by the negative and significant correlation between the total number of tokens invested in the Market 2 and the *rejection* of the ER values system (ER_least). This result supports the part of the hypothesis **H1-CPR** respecting the ER value system, while there is no significant indication relating the other levels of existence.

We regard these correlations with ER as remarkable because only in the baseline condition the players are free to seek and behave according their own individual bent. The pattern of the correlations with ER (significant positive when ER is accepted and significant negative when ER is rejected) provides an empirical evidence of it's highly achievement oriented character. Moreover, it is an experimental behavioral response that is predicted by Graves's theory.

The strategic and self-oriented aspect of ER appropriation behavior is further characterized by the correlations of ER with the levels of investment in the communication and sanctioning conditions. The results show that ER continues to overinvest as long as no sanction can be imposed. This result supports the corresponding parts of hypotheses **H2-CPR** as well as **H3.1-CPR**. This is shown through: (1) the significant positive correlation of the total investment during the communication condition with ER_most, (2) the significant negative correlation of those investments with ER_least, together with (3) the significant *positive* correlation with ER_*least* in the sanctioning condition (i.e. the less strategic the thinking the more the individual invest during the sanctioning institutional setting). The correlation with the ER_most in the sanctioning treatment also goes in the "right" direction, but it is not statistically significant. Note that the correlation of ER_most with the frequency of defection

is positive and statistically significant, as advanced in the hypothesis **H2.1-CPR**. The opposite and also significant correlation with ER_least confirms this trend.

Putting this information together with those concerning the levels of investment in the communication and sanctioning conditions we can conclude that most of the defections due to ER occurred during the rounds of the communication condition. This is suggested as well as by the fact that the ER_most does not pay *significantly* more fines, even if the correlation is positive, as we posited by the hypothesis **H3.2-CPR**. The sense of our hypothesis, nonetheless, is indicated by the observation that the *rejection* of the ER achieving orientation (ER_least) does lead people to defect less and thus to pay significantly *less* fines.

Finally, relating to the sanctioning behavior, the *rejection* of ER (ER_least) correlates negatively with the outflow with fees, while ER_most present no significant correlation with these expenditures. Since FS antagonizes with ER (Tables 7.1 and 7.4), the first correlation is possibly reflecting FS unwillingness to penalize defectors, as we shall verify ahead. The second and no significant correlation indicates that ER is not particularly bent toward bearing the costs of reinforcing the joint strategy, which does not contradict the hypothesis **H4-CPR** in this regard.

Table 7.9 – Pearson correlations

Value systems (levels of existence)	n = 196	Individual total investment per treatment (tokens)			Individual totals		
		Baseline	Communication	Sanctioning	Defection (frequency)	Fine (R$)	Fee (R$)
BO_most	Pearson Correlation	0,014	-0,076	-0,088	-0,068	**-0,105***	-0,079
	Sig. (1-tailed)	0,423	0,145	0,111	0,170	0,072	0,135
BO_least	Pearson Correlation	0,064	**0,128****	-0,036	**0,155****	**0,122****	**0,117***
	Sig. (1-tailed)	0,188	0,037	0,310	0,015	0,044	0,051
CP_most	Pearson Correlation	0,042	0,040	**-0,130****	**0,118****	**-0,106***	-0,054
	Sig. (1-tailed)	0,280	0,288	0,034	0,050	0,069	0,225
CP_least	Pearson Correlation	-0,029	**-0,116***	0,044	**-0,140****	**-0,111***	**-0,152****
	Sig. (1-tailed)	0,346	0,052	0,270	0,025	0,061	0,017
DQ_most	Pearson Correlation	-0,049	-0,056	0,015	0,003	0,057	0,047
	Sig. (1-tailed)	0,247	0,219	0,415	0,486	0,216	0,256
DQ_least	Pearson Correlation	0,007	-0,005	**-0,160****	0,036	-0,060	0,005
	Sig. (1-tailed)	0,459	0,472	0,012	0,309	0,202	0,470
ER_most	Pearson Correlation	**0,093***	**0,121****	-0,021	**0,233*****	0,065	0,033
	Sig. (1-tailed)	0,097	0,045	0,385	0,001	0,184	0,323
ER_least	Pearson Correlation	**-0,134****	**-0,146****	**0,116***	**-0,311*****	**-0,096***	**-0,126****
	Sig. (1-tailed)	0,031	0,021	0,053	0,000	0,090	0,039
FS_most	Pearson Correlation	-0,085	**-0,134****	**0,100***	**-0,259*****	-0,074	**-0,112***
	Sig. (1-tailed)	0,118	0,030	0,081	0,000	0,152	0,060
FS_least	Pearson Correlation	0,049	**0,198****	**0,105***	**0,277*****	**0,226*****	**0,295*****
	Sig. (1-tailed)	0,247	0,003	0,072	0,000	0,001	0,000
GT_most	Pearson Correlation	0,015	0,091	0,010	0,016	0,068	**0,109***
	Sig. (1-tailed)	0,415	0,101	0,447	0,413	0,171	0,064
GT_least	Pearson Correlation	0,069	0,025	-0,039	**0,111***	0,013	-0,022
	Sig. (1-tailed)	0,168	0,363	0,293	0,060	0,426	0,379
Egocentric	Pearson Correlation	0,024	0,063	**-0,092***	**0,104***	0,015	0,076
(Factor 1)	Sig. (1-tailed)	0,369	0,190	0,100	0,074	0,417	0,145
Absolutistic	Pearson Correlation	-0,004	-0,038	-0,003	0,055	0,000	-0,018
(Factor 2)	Sig. (1-tailed)	0,477	0,299	0,482	0,220	0,499	0,401
Sociocentric	Pearson Correlation	**-0,105***	**-0,159****	0,044	**-0,295*****	**-0,125****	**-0,119****
(Factor 3)	Sig. (1-tailed)	0,072	0,013	0,268	0,000	0,040	0,048

*** Correlation is significant at the 0.01 level (1-tailed).
** Correlation is significant at the 0.05 level (1-tailed).
* Correlation is significant at the 0.15 level (1-tailed).

7.5.4.2. Egocentrics defect significantly more while can be put in line through sanctioning

The correlations show significant positive relationship between the frequency of defections and the scores in egocentrism (Factor 1), in agreement with hypothesis **HF2.1-CPR**. This also appears in the significant positive correlation between the frequency of defections with CP_most and significant negative correlation with CP_least (the original hypothesis **H2.1-CPR**).

We also posited the CP (as well as ER) would respond to the impositions of the sanctioning mechanism by reducing investment in that treatment (hypotheses **H3.1-CPR**). On the other hand, in the hypothesis **H3.2-CPR** we speculate that the CP would at least try to skip the norm during the sanctioning condition and that this could cost them more than average. We indeed found that both CP_most as well as the scores in egocentrism (Factor 1) correlates significantly negatively with total investment in the sanctioning condition, supporting hypothesis **H3.1-CPR**. We did not find direct evidence that the egocentrics pay more fines, though. On the contrary, CP_most correlated significantly negatively with the total payment of fines, reinforcing the previous idea that they can be put in line by the imposition of the sanctioning mechanism. Yet, the *rejection* of the CP value system (CP_least) also correlates significantly negatively with the total expenditures with fines.

In the hypothesis **H4-CPR** we posited that CP_most is not bent to incur into costs in order to reinforce the collective norm. The negative correlation sign suggests indicates in this direction but this correlation is small and not statically significant. Moreover, again the unwillingness to pay the fees is significantly correlated with the *rejection* of the CP value systems (CP_least). While we have not put forward explicit hypothesis referring to the *rejection* of any value system, we suggest that these significant correlations with CP_least are possibly reflecting the FS value systems, which antagonizes CP in the component egocentrism (Tables 7.4; see also Table 7.1).

Accordingly, the remaining correlations with CP_least show that the more CP is *rejected* the more the player cooperates with the joint strategy. This is shown by: (1) CP_least significant *negative* correlation between with the total investment during the communication setting, (2) the significant negative correlation with the frequency of defection, and (3) the significant negative correlation with the payment of fines, as already mentioned.

7.5.4.3. Sociocentrics cooperate spontaneously and don't like to impose fines

Conversely, for the sociocentric and anti-strategic component (Factor 3) we note that: (1) the total investment is significantly lower during both the baseline and the nonbinding communication conditions (as suggested by hypotheses **HF1-CPR** and **HF2-CPR**), (2) the frequency of overall defection (whatever reason) as well as primary defections (payment of fines) are significantly fewer (concurring with hypotheses **HF2.1-CPR** and **HF2.2-CPR**); and, (3) the application of the sanctioning mechanism is significantly less frequent (in accordance with the implication of hypothesis **H4-CPR**).

Clearly, the same ideas is conveyed by the correlations with FS as separate variable, except for the positive association of FS_most with the total investment during the sanctioning condition. This is probably due to the prevalence of the FS_most responses in the total sample variability (Figure 7.1), and does not mean that the FS centered individuals are running out of the norm precisely when sanction is imposed.

As we see, the FS_most correlates significantly *negatively* with the total investment during the communication period (agreeing with hypothesis **H2-CPR**) and with the frequency of defections (in accord with hypothesis **H2.1-CPR**). Moreover, the *rejection* of the FS value system (FS_least) correlates *positively* and significantly with: (1) the total investment in both the communication and the sanctioning institutional settings, (2) the frequency of defection, and (3) the payment of fines.

As for the use of the sanctioning mechanism (hypothesis **H4-CPR**), the *rejection* of FS is *positively* and significantly associated with the application of fines whilst its *acceptance* is significantly *negatively* correlated with it. All these are in wide concurrence with Graves's theory.

In particular, these results indicate that the sociocentrics cooperate spontaneously, i.e., during the nonbinding communication setting, and that they are not prone to impose penalty on others. While FS are indeed supposed to be willing to bear the cost of reinforcing the collective strategy, the penalizing and coercive character of the sanctioning mechanism at hand in the experimental setting is contrary to the compassionate, conciliatory, and anti-coercive character of the FS value system. We thus did not expect to find FS making extensive use this coercive instrument (as set forth in hypothesis **H4-CPR**).

7.5.4.4. No significant relationships with absolutism (Factor 2) but GT strongly reciprocates

As anticipated, the similarities of the behavioral responses involving the antagonists DQ and GT in the interior of the Factor 2 make difficult to find significant relations concerning the component absolutism. Moreover, we found there are also few significant correlations involving both the DQ and the GT value systems separately.

Despite the statement in hypothesis **H1-CPR**, the lack of significant statistical correlations concerning the investment in the baseline setting is not surprising. In fact, aside CP and ER, there is nothing special about the other value systems to indicate that either over or sub investment could get out the range of the overall variability. Similarly, the behavioral response which is expected to deviate from the general pattern during the communication condition is that of overinvestment, which are not expect be observed neither for DQ nor GT (as in hypothesis **H2-CPR**). Conversely, the *rejection* of the absolutist stance (DQ_least), which is associated with the egocentric wave (Tables 8.4), does

correlate significantly with the *reduction* of the level of investment during the sanctioning treatment (as posited by hypothesis **H3.1-CPR**).

In turn, GT_most correlates *positively* and significantly with the application of the sanctioning mechanism, as advanced by hypothesis **H4-CPR**. That this is truly an expression of strong reciprocity, as we suggested, is indicated by the *positive* and significant correlation between the frequency of defection and the *rejection* of the GT value system (GT_least). This relationship shows that those who defect more are *not* those centered at the GT level of existence (as posited by the hypothesis **H2.1-CPR** and **H2.2-CPR**).

7.5.5. Multivariate analyses

The bivariate correlation analyses provided us with some indication as to how the varied behavioral responses measured in the experimental conditions relates with the value systems and the three principal components or levels of existence. Yet, the bivariate correlations tell us this story only through the examination of pairwise relations between the behavioral variables, on the one hand, and each of the separate value systems or principal components, on the other hand. Mostly, the explanation of a behavior involves the intercourse of multiple interrelated variables that may have distributed and combined influences upon the behavioral responses under investigation. In some extend, such intercourse is already captured in the linear combinations that constitute each of the three principal components or factors. Still, the factors themselves may offer partial and combined influences upon the behavioral responses.

We now examine this multivariate relationship by means of multiple regressions analysis using the three principal components as predictors of the various behavioral responses produced during the CPR experiments. We should remind that the factor scores are, first, normalized descriptors (zero mean and unitary variance), and, second, that they present minimal correlation one with the others (orthogonal rotation). The first point means that the individual scores indicate standard deviations from the sample mean, so that they offer a

description which is relative to the variability present in our sample. The second point means that each principal component describes a distinct dimension of the interior (Left-Hand quadrants) phenomenon under investigation.

Still, the wave-flowing and spiraling dynamic of the interior phenomenon we have taken some pictures of implies that our snapshots cannot offer particularly high quality images. We already saw that reflected in the low to moderate magnitudes of the bivariate correlations. Here, the original blurriness of the images conveyed by the principal components (see the moderate communalities in Table 7.2) is added to out attempt to capture their relationship with the behavioral responses using linear regression analysis. For all that we have been discussing on the dynamic of the psychological development; we should not expect to obtain well fitting adjustments using linear regressions models. As we shall see, the goodness of fit vary indeed from very poor to just moderate. Nonetheless, it is not the case, either, of searching for more flexible specifications, such as quadratic or higher order polynomials, because our purpose is neither to model the psychosocial development mathematically nor to establish quantitative relations between Graves's constructs and our selected behavioral responses. Instead, we are primary concerned with the *qualitative meaning* of at least statistically significant coefficients and in producing some experimental evidence that the likelihood of the cooperative response increases with the levels of existence.

In other words, our application of the multivariate regressions aims just to submit our data to a more rigorous statistical inspection, in order to make sure of the robustness of the relationships just examined via bivariate correlations. Also, even if we are not looking for quantitative effects, the explanatory connotation implied by the positioning of the independent and dependent variables in the regression analysis is preserved.

7.5.5.1. The higher the score in sociocentrism, the lower the investment after communication (evidence for hypothesis HF2-CPR)

Beginning again with the appropriation behavior, the only statistically significant ($p < .10$) linear regression model in which the three waves of existence are used as predictors is that for the communication condition. The *F* statistic for the whole model is 2.245 and the significance level is .084 ($R^2 = .034$). Among the predictors, the only significant ($p < .05$) coefficient is that of the sociocentric level (Factor 3). The result confirms the negative association between the score in sociocentrism and the total investment during the communication condition, even after discounting for the partial (though not statically significant) effects of both other components (Table 7.10).

Table 7.10 – Linear multivariate regression dependent variable: total investment (communication) coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	58,606	,967		60,610	,000
	Egocentric	1,222	1,028	,085	1,188	,236
	Absolutistic	-,534	,959	-,040	-,558	,578
	Sociocentric	-2,272	,958	-,170	-2,373	,019

7.5.5.2. Egocentrics defect more and sociocentrics defect less (evidence for hypothesis HF2.1-CPR)

Table 7.11 present the regression for the frequency of defection. The model using the three levels as predictors presents *F* statistic of 7.883 and is significant at $p < .0001$ ($R^2 = .11$). The results corroborate and strengthen the findings that, on the one hand, the stability of the joint strategy is threatened by

the presence of egocentrism ($p < .05$) and, on the other, it is underpinned by sociocentrism ($p < 0001$). These associations remain significant even considering the partial effects of all three components. The scores on absolutism (Factor 3) had no significant effect in the explanation of the frequency of defection. We remind, nonetheless, that the bivariate correlation analysis shows significant positive correlations between the *rejection* of GT (GT_least) and the frequency of overall defections, which is a result that agrees with hypothesis **HF2.1-CPR** and is internal to Factor 2 variable-component structure (Table 7.4). The lack of statistical significance for the factor absolutism is thus possibly owing to the non significant association between defections and the DQ value system (Table 7.9). Again, we remind that DQ is expected to defect more especially due to secondary (retaliatory) defections, so that this is an observation which do not depends only on DQ own characteristic but also on the contingencies (more primary defections) of the experimental situation.

Table 7.11 – Linear multivariate regression dependent variable: frequency of defections coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	1,526	,145		10,533	,000
	Egocentric	,315	,154	,140	2,044	,042
	Absolutistic	,111	,144	,053	,774	,440
	Sociocentric	-,653	,143	-,312	-4,554	,000

7.5.5.3. Egocentrics cheat more, sociocentrics cheat less (evidence for hypothesis HF2.2-CPR)

In order to examine the initiative of cheating during the nonbinding communication settings we isolated the primary defection from the defections that follow prior deviations in earlier rounds. We did that by creating a binary variable (cheating) which indicates whether the player had or not being the first to deviate from the joint strategy (whether alone or simultaneously with someone else). The association was examined through binary logistic regression analysis.

On the total of 196 valid observations (4 are missing case respecting the SD questionnaire) there were only 35 (17.9%) primary defectors (cheaters) so that the model with no explanatory variables but the constant is able to classify correctly a percentage as high as 82.1% of the alternatives without predicting one single cheating (cut value = .50). As a result the explanatory power left for the principal components is necessarily small. Indeed, the inclusion of the three factors increases the percentage of correctly predicted categories just up to 82.7%. Despite this, the coefficients for egocentrism and sociocentrism are statistically significant (Table 7.12) and in agreement with hypothesis **HF2.2-CPR**. Accordingly, the higher the individuals' scores in egocentrism (Factor 1) the greater the probability of being a cheater. Conversely, the higher the individuals' scores for sociocentrism (Factor 3) the smaller the probability of cheating. The association between the scores for absolutism (Factor 2) and the probability of being a cheater is not statistically significant. We note that this is also a predictable result (hypothesis **HF2.2-CPR**), given that the antagonists in the variable-component structure of Factor 2 (DQ and GT) are both expected to behave similarly with reference to the initiative of cheating (i.e. none of them were expected to cheat).

Table 7.12 – Binary logistic regression dependent variable: initiative of cheating (primary defection) coefficients

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Egocentric	,404	,207	3,816	1	,051	1,498
	Absolutistic	-,142	,194	,536	1	,464	,868
	Sociocentric	-,579	,199	8,454	1	,004	,560
	Constant	-1,645	,206	63,727	1	,000	,193

[a] Variable(s) entered on step 1: Egocentric, Absolutistic, Sociocentric.

7.5.5.4. Sociocentrics support the collective strategy regardless the behavior of others (evidence for hypothesis HF2.3-CPR)

We now test the steadiness of the cooperative disposition associated to the three principal waves of existence. We approached that by creating a new binary variable indicating whether the player has stuck with the joint strategy without any single deviation all through the 10 rounds of the communication setting, regardless of the behavior of others.

This time the alternative categories are more balanced. There were a total of 109 (55.6%) "never cheating" behavioral responses out of 196 observations. The model without explanatory variables predicted correctly 55.6% of the occurrences (cut value = .50). The overall percentage of correctly predicted categories was 59.7% for the model including the three levels as predictors. As for the coefficients which are statistically significant (Table 7.14), the result shows that the probability of finding an individual who have never deviated from the joint strategy, regardless the behavior of others, is significantly ($p < .01$) higher if the individual scores higher in sociocentrism (Factor 3).

This result, however, is influenced by fact that the average scores in sociocentrism are higher precisely in the six groups in which no defections were observed at all (Figure 7.4). We thus retested this persistency in keeping with the collective agreement considering only the groups in which at least one defection

was observed. We thus excluded the six groups in which there is no defection and also the Group 21, which presents too many. Keeping the cutting point at .50, the overall percentage of correctly predicted categories is now 61.7%, for the model with the three principal components. The new coefficient is still significant ($p < .10$) and indicates that the higher the scores in sociocentrism the higher the probability of finding an individual who stuck with the collective norm regardless the occurrence of at least one defection on part of others (Table 7.14).

Table 7.13 – Binary logistic regression dependent variable: never deviate from the group's norm (all groups) coefficients

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Egocentric	-,119	,159	,561	1	,454	,888
	Absolutistic	-,015	,147	,010	1	,920	,985
	Sociocentric	,523	,156	11,281	1	,001	1,686
	Constant	,220	,148	2,203	1	,138	1,247

[a] Variable(s) entered on step 1: Egocentric, Absolutistic, Sociocentric.

Table 7.14 – Binary logistic regression dependent variable: never deviate from the group's norm (at least one defection) coefficients

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Egocentric	,175	,194	,816	1	,366	1,191
	Absolutistic	,058	,173	,113	1	,737	1,060
	Sociocentric	,324	,191	2,883	1	,090	1,383
	Constant	-,245	,179	1,869	1	,172	,783

[a] Variable(s) entered on step 1: Egocentric, Absolutistic, Sociocentric.

7.5.5.5. Sociocentrics pay less fines (evidence for hypothesis HF3-CPR)

Given the evidences already reported as to the behavior of the sociocentrics during the nonbinding communication setting, we no wonder discovering that they also pay fewer fines during the sanctioning condition. This is, nonetheless, a remarkable result if we take into account that the behavioral variability (i.e. decision to defect) is greatly reduced after the imposition of the sanctioning mechanism. This fact challenges the likelihood of finding significant statistical relationships.

Indeed, the multivariate linear regression model explaining the payment of fines (including all three principal components as explanatory variables) is not statistically significant (F = 1.84; p = .357). Yet, the bivariate model including only the constant and the factor sociocentrism is significant (F = 3.097; p = .080; R^2 = .16) and agrees which hypothesis **H3.2-CPR**. It indicates that participants scoring higher in sociocentrism pay fewer fines (Table 7.15).

While this variable is influenced by the wiliness to pay fees on the part of the punishers, we noticed that the total expenditures with fines is a *proxi* for the initiative of cheating, since no one single defection passed without being fined during the sanctioning condition and no defection in this setting is secondary, i.e., none has the character of punition, even if some may be imitative. Typically, defections during the sanctioning conditions occurred either at the first, second or at the last rounds.

Since defections occurred, and they cannot be attributed to the sociocentrics, they ought to be attributable to the egocentrics, the absolutists, or both, even if this does not appear as significant statistical associations. They can also be attributed to the go-getting and strategic thinking of ER, especially the last round defections, since the "rational" choice would recommend no punishing at the last round, though the punishers apparently reasoned otherwise. For all the previous results, we can infer that such defections are not due to BO, DQ or GT. It thus seems just natural to infer that they come from the CP and ER value

systems, as suggested by the previous findings. We now turn to examine who were the punishers.

Table 7.15 – Linear multivariate regression dependent variable: total outflow with fines coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,746	,142		5,258	,000
	Sociocentric	-,245	,139	-,125	-1,760	,080

7.5.5.6. The punishers were not sociocentrics but DQ and GT (evidence for hypothesis H4-CPR)

As for the sanctioning behavior, we argued that there was not possible to advance instantly recognizable hypothesis as regards to the relationship with the principal components. This was so because the value systems which antagonizes in the variable-component structure internal to each factor are expected to behave similarly relating to this particular behavioral response. Concerning the Factor 2 (absolutism), both DQ and GT are expected to serve from the sanctioning mechanism, while, as commented, DQ present some "taste for punishing" whereas GT express better the idea of "strong reciprocity". The same counteract interaction works internal to Factor 1 (egocentrics), since we have posited the CP also would sanction defectors, out of "taste for punishing", and CP antagonizes DQ, while not with FS, internally to Factor 1 (Table 7.4). Finally, relating to Factor 3 (sociocentrism), the antagonists FS and ER are both expected to do not use the sanctioning mechanism. FS wouldn't do it out of sensibility, kindness and

distaste for punishing, whereas ER is expected to "free-ride" on this second-order dilemma as much as in the first-order one.

Given these, we said that this hypothesis should be better examined by disaggregating the explanatory variables. We have, nonetheless, tried first to find some relationship with the principal components. We discovered that, as expected, the multivariate regression which includes all three principal components is not statistically significant (F = 1.504; p = .215). Yet, the bivariate model relating the total expenses with fees with the scores in the Factor 3 (sociocentric) does is statistically significant (F = 2.804; p = .096). The negative coefficient (Table 7.16) indicates that the sociocentrics are not willing to make use of the sanctioning mechanism, but we still don't know whether this result conforms to our expectations regarding FS or contradicts our expectation as to ER (since what we have in the Factor 3 is a *rejection* of ER value system). So let us see what we find by disaggregating the explanatory variables (Tables 7.17 and 7.18).

Table 7.16 – Linear multivariate regression dependent variable: total expenditure with fees (model 1) coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,209	,059		3,547	,000
	Sociocentric	-,097	,058	-,119	-1,675	,096

For this purpose, we have to test some alternative specifications since not all value system variables can remain in the model at the same time, due to the occurrence collinearity (individuals ought to distribute their "most like me" and "least like me" choices amongst a fix number of 40 questions). Given that the BO value system imparts less information, for it is somewhat evenly represented in all the three principal components (Table 7.4), we decided to take it out and run the model using only the other systems as predictors. The resulting model (model 2) is statically significant ($F = 3.978$; $p < .0001$; $R^2 = .18$) and the coefficients are presented in Table 7.17. Also, with this specification we are able to get to a new interesting observation.

Note that the coefficient for the variable DQ_most is statistically significant ($p < .10$) and indicates that the absolutistic stance is positively related with the punishing behavior, in agreement with our reasoning (hypothesis **H4-CPR**). The ability to show this depends on the balance of the correlations involving CP, DQ and the FS value systems (see Table 7.1). With BO already excluded, the FS value system draws off the statistical significance of CP, and this lets the significance of the behavioral response associated with DQ to express itself. If we take the variable FS_most out of the model (model 3)[59] the significance of DQ_most is drawn by CP_most, and now we can see that the affinity with the representative statements of CP relates significantly negatively with the expenses with fees (Table 7.18). This result disagrees with hypothesis **H4-CPR** because we assumed that CP's taste for punishing would be strong enough for pushing the association positively. Yet, we must concede that CP is definitely not "strong reciprocator", and since the sanctioning mechanism is costly, this result does not really disagree with Graves's characterization of CP. A deeper examination, by contrasting costly and costless mechanisms, would be required to disentangle such imprecision.

[59] Model 3 is also statistically significant, with F statistic = 4.234 ($p < .01$).

Table 7.17 – Linear multivariate regression dependent variable: total expenditure with fees (model 2) coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
2	(Constant)	-1,542	1,654		-,932	,352
	CP_most	-,032	,054	-,081	-,599	,550
	CP_least	-0,26	,023	-,131	-1,119	,265
	DQ_most	,079	,043	,361	1,859	,065
	DQ_least	-,003	,028	-,013	-,118	,906
	ER_most	,020	,035	,099	,578	,564
	ER_least	-,027	,028	-,092	-,976	,330
	FS_most	,046	,037	,238	1,247	,214
	FS_least	,144	,039	,350	3,713	,000
	GT_most	,085	,038	,413	2,198	,029
	GT_least	,018	,038	,046	,490	,625

In both models we see that both the *rejection* of the FS representative statements (FS_least) and the *acceptance* of those representing GT (GT_most) are positively and significantly related with the total expenditure with fees during the sanctioning treatment. According to the characterizations of these value systems, the interpretation is that, on the one hand, the affinity with GT expresses the wiliness to bear the imposed costs to apply fines on defectors and thus reinforce the joint strategy for the collective good, despite the fact that, in order to do this, they are imposing a lost on those who have defected, because for GT functionality and good governance have the highest priority (hypothesis **H4-CPR**). On the other hand, while the FS system is also communitarian, the focus goes on reconciliation and consensus. As we saw, from the FS point of view, feeling and caring must supersede cold rationality. The FS value system is strongly egalitarian, antihierarchy, anticoercion. This may explain why FS_most is not significantly related with the application of the sanctioning mechanism, whilst its application is significantly more supported by those who *reject* the characteristic gentility of FS value system (the FS_least variable). We finally

note that this is what explains the significant negative coefficient for sociocentrism in the model 1 (Table 7.15).

Table 7.18 – Linear multivariate regression dependent variable: total expenditure with fees (model 3) coefficients

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
3	(Constant)	,253	,815		,310	,757
	CP_most	-,083	,035	-,209	-2,413	,017
	CP_least	-,032	,023	-,163	-1,428	,155
	DQ_most	,034	,023	,157	1,500	,135
	DQ_least	-,013	,027	-,049	-,473	,637
	ER_most	-,017	,019	-,082	-,910	,364
	ER_least	-,031	,028	-,106	-1,129	,260
	FS_least	,129	,037	,313	3,493	,001
	GT_most	,046	,023	,223	2,023	,045
	GT_least	,011	,037	,027	,295	,769

7.6. Levels of existence, motivation and the provision dilemma

We now examine how our levels of existence behaved in the public good experiment. To remind, the experiment was a step-level givesome dilemma (Poppe and Zwikker, 1996) where enough participants had to contribute in order to obtain outcomes for all. A total of 127 participants (62 female and 65 male) took part in the experiment from which only 4 had not participated in the earlier phase of the study. There is one missing case concerning the psychosocial profile.

The experiment consisted of nine trials in each of which the participants had 0.50 reais (about US$ 0.25) at his or her disposal. The participants were asked to decide whether they would keep the 50 cents or would contribute them

to a common pool. ("I do/do not contribute"). If at least X percent of the participants contributed the 50 cents to the pool, all 127 participants – those who had not contributed included – would receive 1.00 real. Discussion between the participants was not allowed. If a participant contributed him – or herself and not enough others contributed, s/he lost his or her contribution. So, in each trial the participant could get 0 cents (s/he contributed but not enough others contributed); 50 cents (s/he did not contribute and not enough others contributed); 1.00 real (s/he contributed and enough others contributed) or 1.50 real (s/he did not contribute but enough others contributed).

The percentage (X) of the 127 participants who had to contribute in order to receive 1.00 real from the pool varied. In the nine trials it was 40%, 70%, 10%, 60%, 90%, 30%, 50%, 80% and 20%, respectively. Before every trial the participant was asked to indicate which percentage of all participants s/he expected to contribute ("I expect that… percent of all participants contribute"). To stimulate the participants to mention accurate expectations, each of the 9 participants (one per trail) who had made the most accurate expectations would receive a bonus of 5.00 reais (about US$2.50).

Following Poppe (2005), we identified the motive under the participants' decision whether to contribute or not by comparing his or her decision with his or her expectation regarding the decision of others. If participants expected enough others to contribute and did not contribute him – or herself, the motive was *greed*. If participants expected enough others to contribute and did contribute him – or herself, the motive was *solidarity*. If participants expected not enough others to contribute and did not contribute him – or herself, the motive was *fear* (of incurring a loss or being gypped). If participants expected not enough others to contribute and did contribute him – or herself, the motive was sense of *duty* (POPPE; ZWIKKER, 1996).

The nine trials were divided into three contribution levels: Low Contribution (the trials in which it was required that 10%, 20% or 30% of the participants contributed in order to obtain the R$ 1.00), Medium Contribution (40%, 50% or 60% required) and High Contribution (70%, 80% or 90%

required). For each participant and for each contribution level the percentage of the three trials was computed in which the participant expected that enough others would contribute, contributed him- or herself and showed *greed*, *fear*, *solidarity* and a sense of *duty*.

The results show that, considering all trails, 51.4% of the participants expected that enough (other) participants would contribute. The actual percentage of contributors amounted to only 26.0%. Given this, the public good was provided only up to the 40% contribution level. In the nine trials, 28.6% of the participants showed *greed*, 45.1% *fear*, 22.8% *solidarity* and 3.4% a sense of *duty* (Table 7.19). On general, Poppes' participants contributed more than ours. In his experiment, 57.9% of the participants expected that there would be enough contributions on the average of all trails, and 48.0% actually contributed for the common pool. As for the motives, 15.0% of Poppes' participants showed greed, 36.8% fear, 42.9% solidarity and 5.2% a sense of duty (POPPES, 2005).

Table 7.19 presents the distribution of expectations, actual decisions and motives according the three grouped levels of necessary contribution. Following Poppes, we performed a series of ANOVA with contribution level as within-subject factor. If an ANOVA showed a significant effect of contribution level, it was followed by a linear trend analysis. It was found that the contribution level had significant effect both on expectations and decisions whether to contribute or not, as well as on the motives under the participants' decisions. In all cases, the testes of homogeneity of variances indicated that the variances are unequal. We then draw once more on Tamhane's test for the multiple comparisons of equality of means. The results of ANOVA are presented in Tables 7.20, 7.21 and 7.22. The model summaries and parameters estimates for the linear tend analyses are presented in Table 7.23.

Table 7.19 – Percentages of participants who did (not) expect enough others to contribute, did (not) contribute themselves and showed fear, greed, solidarity and a sense of duty

	Participant expected that enough participants contributed					
	No		Yes		Both	
		Fear		***Greed***		
Participant did not contribute	Low	12,6	Low	57,0	Low	69,6
	Medium	40,9	Medium	23,4	Médium	64,3
	High	81,9	High	5,5	High	87,4
	All	45,1	All	28,6	All	73,8
		Duty		***Solidarity***		
Participant contributed	Low	1,0	Low	29,4	Low	30,4
	Medium	3,4	Medium	32,3	Médium	35,7
	High	5,8	High	6,8	High	12,6
	All	3,4	All	22,8	All	26,2
Both	Low	13,6	Low	86,4		
	Medium	44,4	Medium	55,6		
	High	87,7	High	12,3		
	All	48,6	All	51,4		

Note: Low, Medium, High: trials in which a low (10–30), medium (40–60) or high (70–90) percentage of the participants had to contribute to obtain R$1.00. All: all trials.

The results indicate that the higher the percentage of participants that had to contribute in order the public good could be provided the less participants expected enough participants to contribute and the less they contributed themselves. Therefore, the more contributions were required the more *fear* and sense of *duty* and the less *greed* and *solidarity* the participants showed (Table 7.23). The multiple comparisons show that the expectations regarding the contribution of others cut down significantly from the low to the medium and from the medium to the high contribution levels (Table 7.20), whereas the percentages of actual contributions fell down from the medium to the high requirements but actually rose from the low to the medium level (Table 7.21). The ANOVA results tell us that such increasing is not statistically significant but,

in fact, had not it occurred the public good wouldn't be provided at the 30% and 40% levels of required contributions. Accordingly, the true relation between the levels of necessary contribution and the percentage of actual contributions is not linear. We then adjusted a third degree trend line in order to show this as well as that the percentages of actual contributions are similar among the three higher levels of necessary contribution (Figure 7.31 to the right). Since the expectations are also statistically similar between the 10% and 20% as well as between the three upper levels, we adjusted a cubic trend line to the expectations as well (Figure 7.31 to the left).

Table 7.20 – ANOVA: multiple comparisons (Tamhane) dependent variable: expectation (yes or no)

(I) Contribution level	(J) Contribution level	Mean difference (I-J)	Std. error	Sig.
Low (10-30)	Medium (40-60)	0,307*	0,031	0,000
	High (70-90)	0,740*	0,024	0,000
Medium (40-60)	Low (10-30)	-0,307*	0,031	0,000
	High (70-90)	0,433*	0,031	0,000
High (70-90)	Low (10-30)	-0,740*	0,024	0,000
	Medium (40-60)	-0,433*	0,031	0,000

* The mean difference is significant at the .05 level.

Table 7.21 – ANOVA: multiple comparisons (Tamhane) dependent variable: actual contributions (yes or no)

(I) Contribution level	(J) Contribution level	Mean Difference (I-J)	Std. error	Sig.
Low (10-30)	Medium (40-60)	-0,052	0,034	0,328
	High (70-90)	0,178*	0,029	0,000
Medium (40-60)	Low (10-30)	0,052	0,034	0,328
	High (70-90)	0,231*	0,030	0,000
High (70-90)	Low (10-30)	-0,178*	0,029	0,000
	Medium (40-60)	-0,231*	0,030	0,000

* The mean difference is significant at the .05 level.

As for the motives, the differences between the three levels of required contribution are significant for fear and greed. The percentage of contributions which are motivated by sense of duty increases statistically from the low to the high level of contribution requirement, but not so from the low to the medium level. With reference to the motive solidarity, the pattern resembles that of actual contributions, so that the percentages first get higher and than fell down. The increasing shift is not statistically but consequentially significant, as pointed out. The decreasing of the contribution is significant both statistically as well as in its practical consequences. In order to illustrate the trends we adjusted a cubic trend line for each motive except for sense of duty. In this case we preserved the linear trend.

To summarize the relevant behavioral information, note first that the participants overestimated the willingness to contribute at the low and medium levels of necessary contribution, while the expectations match the actual levels of contributions much better at the three higher levels of necessary contribution (Figure 7.31). As for the motivations, an interesting point is that the overestimation of the expectancy at the low (10-30) contribution level seems to be owing to a subestimation of *greediness* among the participants. As shown in Table 7.19, about 82% of the total of noncontributors at the low contribution

level did not contribute due to *greed* (57% of the 69.3%). At the medium (40-60) contribution level the motive *fear* explains more of that overestimation, and at the high (70-90) level it explains the short of sufficient contribution together with the fact that not many individuals are motivated by a sense of *duty*. Still, at the high level of contribution requirement the sense of *duty* explains 46% of the percentage of actual contributions (5.8% of the 12.6%).

Table 7.22 – ANOVA: multiple comparisons (Tamhane) dependent variable: motive

(I) Contribution level	(J) Contribution level	Mean difference	Motive			
			Fear	Greed	Duty	Solidarity
Low (10-30)	Medium (40-60)	(I-J)	-0,283*	0,336*	-0,024	-0,029
		Std. Error	0,030	0,033	0,011	0,033
		Sig	0,000	0,000	0,080	0,772
	High (70-90)	(I-J)	-0,693*	0,514*	-0,047*	0,226*
		Std. Error	0,026	0,028	0,013	0,027
		Sig	0,000	0,000	0,001	0,000
Medium (40-60)	Low (10-30)	(I-J)	0,283*	-0,336*	0,024	0,029
		Std. Error	0,030	0,033	0,011	0,033
		Sig	0,000	0,000	0,080	0,772
	High (70-90)	(I-J)	-0,409*	0,178*	-0,024	0,255*
		Std. Error	0,032	0,025	0,015	0,027
		Sig	0,000	0,000	0,318	0,000
High (70-90)	Low (10-30)	(I-J)	0,693*	-0,514*	0,047*	-0,226*
		Std. Error	0,026	0,028	0,013	0,027
		Sig	0,000	0,000	0,001	0,000
	Medium (40-60)	(I-J)	0,409*	-0,178*	0,024	-0,255*
		Std. Error	0,032	0,025	0,015	0,027
		Sig	0,000	0,000	0,318	0,000

* The mean difference is significant at the .05 level.

Table 7.23 – Linear trend: independent variable is contribution level (10-90)

Dependent variable	Model summary		df1	df2	Parameter estimates			Sig.
	R Square	F			Constant	B	Std. Beta	
Expectation	0,390	729,149	1	1141	1,119	-0,012	-0,624	0,000
Contribution	0,027	31,553	1	1141	0,402	-0,003	-0,164	0,000
Fear	0,343	595,425	1	1141	-0,113	0,011	0,586	0,000
Greed	0,235	351,131	1	1141	0,711	-0,008	-0,485	0,000
Duty	0,013	14,985	1	1141	-0,006	0,001	0,114	0,000
Solidarity	0,049	58,688	1	1141	0,408	-0,004	-0,221	0,000

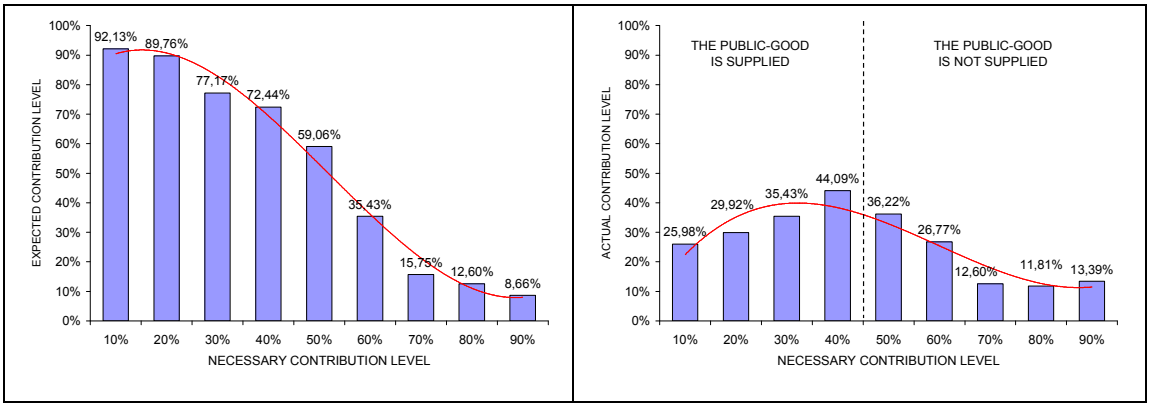


Figure 7.31 – Expected and actual levels of contribution per minimum necessary contribution.

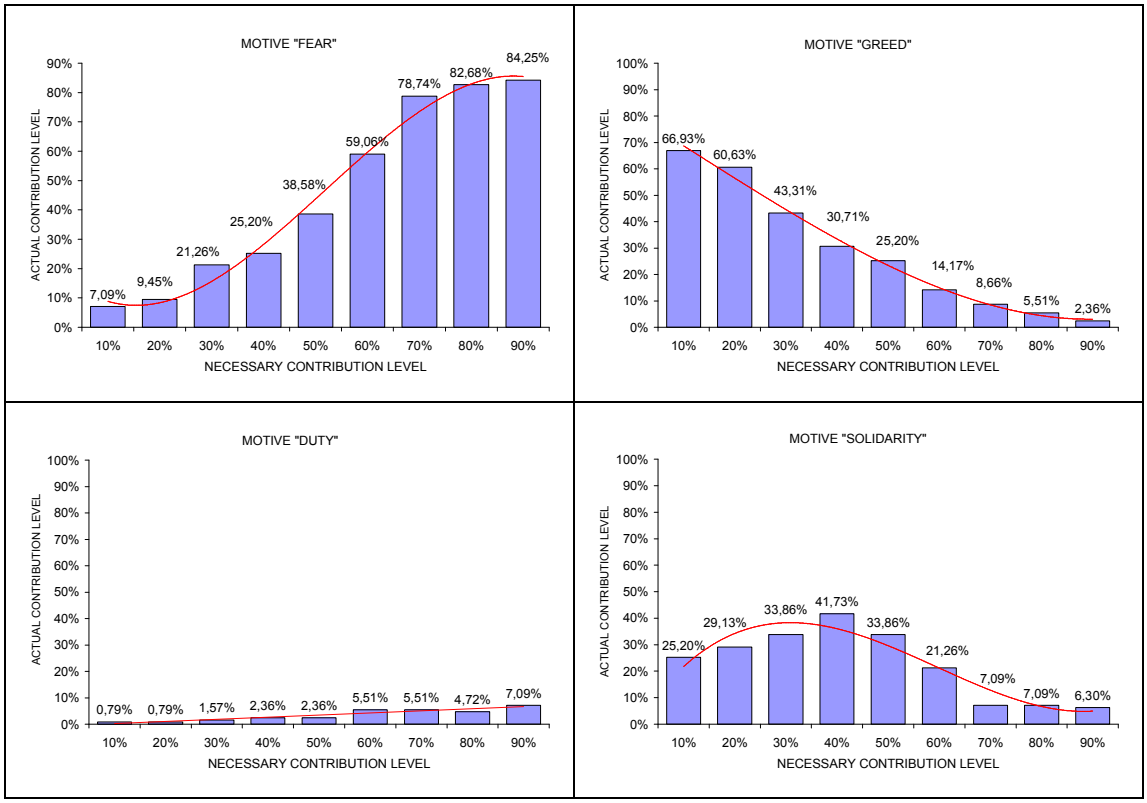


Figure 7.32 – Percentage of appearance of each motive per necessary contribution level.

Given these general observations, we want to know how the behavioral responses in the public good experiment relate with adult psychosocial development, as set forth by Graves's ECLET conception, and captured by the Spiral Dynamics assessment of the corresponding value systems and our three principal components. However, before proceeding with the statistical analysis, we ought to review our working hypothesis under the light of the variable-component correlations that characterizes the principal components.

7.6.1. Reframing the working hypotheses

Hypothesis **H1-PG** posits that FS is more optimistic in relation to the contribution of others, in comparison with the other value systems. So,

congruently with the characteristics expressed by the principal components, we can restate that:

- **HF1-PG**: If some significant positive relationship between the principal components and the expectations as to the contribution of other exists, it might be with the factor sociocentrism.

Relating the choice whether or not to contribute, hypothesis **H2-PG** just says that the express-self systems (CP, ER, and GT) are expected to contribute less frequently, while the sacrifice-self systems (BO, DQ, and FS) are expected to contribute more frequently. Given that such isolation agrees with the internal structure of the principal components, we can unambiguously posit that:

- **HF2-PG**: The scores on egocentrisms (Factor 1) correlates negatively with the frequency of contributions, while the scores on absolutism and sociocentrism correlates positively with the frequency of contributions, in the average of the nine trails.

As for the underlining motivations, hypothesis **H3.1-PG** maintains that: BO, DQ, and FS cooperate out of *solidarity* and sense of *duty*; ER and GT cooperate out of *solidarity* but not sense of *duty*; and CP does not cooperate. Taking the factorial structure, hypothesis **H3.1-PG** implies that:

- **HF3.1-PG**: Scores on absolutism (Factor 2) and sociocentrism (Factor 3) correlates positively with the motive sense of *duty*.

We judge that the correlations with the motive *solidarity* are undermined owing to the antagonism between DQ and GT, in the interior of Factor 2, and that between FS and ER within the Factor 3.

Considering the motives to do not contribute, hypothesis **H3.2-PG** says that when BO, DQ, FS and GT do not cooperate this might be out of *fear* bur not *greed*, while the noncooperative decision of CP and ER are due to both *fear* and *greed*. Accordingly, the new hypothesis suggests that:

- **HF3.2-PG**: The motive *greed* correlates positively with the scores on egocentrism (Factor 1) and negatively with the scores on sociocentrism (Factor 3).

Finally, as for the interactions with the contribution levels, the hypothesis **H4-PG** guesses that BO, DQ and FS contribute more at the high contribution level, and CP and ER contribute less at the low contribution level. Accordingly, the hypothesis for the principal components ought to be:

- **HF4-PG**: Scores on absolutism correlates positively with contributions at the high contribution level, while scores on sociocentrism (Factor 3) correlates positively with contributions at the both the low and the high contribution levels.

7.6.2. Bivariate analysis

As before, we begin by examining the Pearson bivariate correlations and then proceed with the multivariate regression analysis. In what follows we have only 126 instead of 127 cases, since one participant did not attend our claim to fill out the SD questionnaire. Also, in view of the results examined above we decided to employ a different classification of the contribution levels from that of Poppes. As we saw, in our experiment the decisive level of necessary contribution for the provision of the public good was 40%. Moreover, the multiple comparisons of differences of means suggested that a better grouping of requisite contribution levels would deem the low level as being that of 10-20%; the medium as 30-40%; high as 50-60% and the 70-80-90% as very high.

7.6.2.1. Different waves of existence form different expectations regarding the contribution of others (hypothesis H1-PG and HF1-PG)

Beginning with the expectations as to the contribution of others, if we consider the nine trails taken together the value systems do not add any statistically significant information (Table 7.24). This is not surprisingly; in view of our discussion regarding the similarity out the general expectation as the required contribution level increases (all values system were posited to expect that the contributions of other would fall as the requisite contribution increases).

Yet, if we disaggregate the levels of necessary contribution, as suggested above, we see that the ER value system (ER_most) reasons that others will contribute less at low (10-20) level of necessary contribution, but those who reject the ER system (ER_least), like the sociocentrics (Factor 3), reason just otherwise. This observation agrees with the conservative hypothesis **HF1-PG**, if not directly with **H1-PG**. Still, the scores in sociocentrism, as well as those in egocentrism, correlate significantly negatively with the expectations at the high (50-60) contribution level. This can be attributed the general trend, posited common to all value systems, concerning the skepticism as to the contributions of other as the required contribution increases.

But taking the very high contribution level (70-90), those who reject ER (ER_least) expect others will contribute *less* while those who reject FS (FS_least) and BO (BO_least) imagine others will contribute *more*. This result is hard to extricate. Anyway, it seems just like the general skepticism as to the cooperation at the very high contribution levels is being reversed out of sort strategy thinking (ER) who speculate that *others* (but not themselves) will be induced to contribute in view the very high requirement. This idea agrees with both the negative expectation of those rejecting ER as well as with the positive expectations tied to those who reject BO and FS (Table 7.1).

Finally, we have one more statistical significant correlation between BO_most and the expectations at the medium (30-40) contribution level. The coefficient tells that DQ_most believes that others will contribute less at this level of required contribution. This is an uninteresting result which does not demand further speculations.

The point here is that our results show that expectations vary, and this something minimized by the conventional approach, which assumes a unitary cognitive foundation (e.g, rational expectation model). Additionally, the significant associations support the basic inference concerning the “rational” (i.e. strategic) expectation formed by ER, particularly relating to the low level of required contribution, and aggregate the information that others (e.g. sociocentrics) may think just otherwise.

Table 7.24 – Pearson bivariate correlations: expectations

Value systems levels of existence	n =126	Expectation of contributions of others				
		All (10-90)	Low (10-20)	Medium (30-40)	High (50-60)	Very high (70-90)
BO_most	Pearson Correlation	-0,040	0,114	0,023	-0,111	-0,072
	Sig. (1-tailed)	0,329	0,102	0,399	0,109	0,211
BO_least	Pearson Correlation	0,107	-0,080	-0,008	**0,181**	**0,125**
	Sig. (1-tailed)	0,116	0,187	0,466	0,022	0,081
CP_most	Pearson Correlation	0,047	0,069	0,036	0,039	-0,005
	Sig. (1-tailed)	0,301	0,220	0,345	0,334	0,477
CP_least	Pearson Correlation	0,008	-0,028	0,018	0,053	-0,036
	Sig. (1-tailed)	0,463	0,379	0,420	0,278	0,346
DQ_most	Pearson Correlation	-0,098	-0,049	**-0,156**	0,010	-0,071
	Sig. (1-tailed)	0,138	0,292	0,041	0,458	0,216
DQ_least	Pearson Correlation	-0,047	0,040	0,080	**-0,173**	-0,028
	Sig. (1-tailed)	0,302	0,330	0,186	0,026	0,379
ER_most	Pearson Correlation	0,057	**-0,122**	0,081	0,089	0,044
	Sig. (1-tailed)	0,264	0,087	0,184	0,161	0,311
ER_least	Pearson Correlation	-0,054	**0,148**	-0,022	-0,064	**-0,135**
	Sig. (1-tailed)	0,275	0,050	0,402	0,238	0,066
FS_most	Pearson Correlation	0,099	0,094	0,082	0,025	0,077
	Sig. (1-tailed)	0,134	0,148	0,181	0,390	0,196
FS_least	Pearson Correlation	0,055	-0,034	-0,035	0,018	**0,168**
	Sig. (1-tailed)	0,271	0,351	0,349	0,419	0,030
GT_most	Pearson Correlation	-0,085	-0,024	-0,058	-0,099	-0,029
	Sig. (1-tailed)	0,173	0,395	0,258	0,135	0,372
GT_least	Pearson Correlation	-0,105	-0,043	-0,085	-0,071	-0,073
	Sig. (1-tailed)	0,120	0,317	0,171	0,216	0,209
Egocentric	Pearson Correlation	-0,044	0,067	0,009	**-0,128**	-0,020
	Sig. (1-tailed)	0,314	0,227	0,460	0,076	0,413
Absolutistic	Pearson Correlation	-0,043	-0,015	-0,065	0,019	-0,053
	Sig. (1-tailed)	0,317	0,432	0,233	0,418	0,278
Sociocentric	Pearson Correlation	-0,013	**0,161**	0,030	**-0,115**	-0,040
	Sig. (1-tailed)	0,440	0,036	0,369	0,099	0,329

In fact, in the comparison with the actual contributions, we commented about the sizeable overestimation of the actual contributions at the lower levels, and suggested that this seemed to be associated to with a subestimation of *greediness* among the participants. Now we see that this can be attributed to those less akin to the ER strategic way of thinking. We see that in the optimistic expectations of the sociocentrics (Factor 3). Again, this is an indication of the reasonableness of our hypothesis **HF1-PG**. We will have more on this when examining the correlations with the actual decisions and the underlining motives.

7.6.2.2. Sacrifice-self systems contribute more, express-self systems contribute less (hypothesis H2-PG, HF2-PG, H4-PG and HF4-PG)

As suggested above, if we look at the correlations with the actual decision to contribute (Table 7.25) we notice that it were the ER themselves who contributed less at the low requisite condition. As posited, in their expectations they were just generalizing their own bent ("self-mirroring" expectations). In doing so, ER predicted correctly the behavior of CP_most (as well as BO_least).

Note that ER strategic thinking also is also associated with fewer contributions at the very high (70-90) contribution level (significant negative correlation with ER_most and significantly positive correlation with ER_least). Likewise, those who were expecting higher contributions at the low (10-20) requisite level where the same who have contributed more: the FS_most and the sociocentrics wave of existence (Factor 3). The sociocentric level as well as the FS_most and ER_least separate variables also contributed more at the very high (70-90) requisite levels. The BO_most supported the provision of the public good at medium (30-40) requisite level, while those rejecting DQ (DQ_least) contributed less at the high (50-60) and very high (70-90) levels of necessary contribution.

Table 7.25 – Pearson bivariate correlations: actual contributions

Value systems levels of existence	n =126	Actual contribution				
		All (10-90)	Low (10-20)	Medium (30-40)	High (50-60)	Very high (70-90)
BO_most	Pearson Correlation	0,056	-0,046	**0,144**	-0,051	0,034
	Sig. (1-tailed)	0,265	0,304	0,054	0,285	0,352
BO_least	Pearson Correlation	-0,087	**-0,121**	-0,044	-0,004	-0,044
	Sig. (1-tailed)	0,167	0,088	0,313	0,482	0,312
CP_most	Pearson Correlation	-0,045	**-0,129**	-0,019	0,091	-0,047
	Sig. (1-tailed)	0,310	0,074	0,418	0,155	0,301
CP_least	Pearson Correlation	0,111	0,042	0,058	0,088	0,068
	Sig. (1-tailed)	0,108	0,322	0,259	0,164	0,224
DQ_most	Pearson Correlation	-0,007	-0,051	-0,030	0,077	0,018
	Sig. (1-tailed)	0,471	0,284	0,371	0,196	0,420
DQ_least	Pearson Correlation	**-0,148**	-0,045	-0,048	**-0,148**	**-0,133**
	Sig. (1-tailed)	0,049	0,309	0,297	0,050	0,069
ER_most	Pearson Correlation	**-0,152**	**-0,157**	0,014	-0,034	**-0,141**
	Sig. (1-tailed)	0,045	0,040	0,440	0,352	0,058
ER_least	Pearson Correlation	**0,189**	0,113	0,101	0,081	**0,136**
	Sig. (1-tailed)	0,017	0,105	0,131	0,184	0,064
FS_most	Pearson Correlation	**0,238**	**0,302**	0,029	0,039	**0,177**
	Sig. (1-tailed)	0,004	0,000	0,374	0,333	0,024
FS_least	Pearson Correlation	-0,028	0,025	-0,027	-0,050	0,050
	Sig. (1-tailed)	0,378	0,392	0,382	0,288	0,290
GT_most	Pearson Correlation	**-0,119**	-0,049	-0,083	-0,107	-0,070
	Sig. (1-tailed)	0,092	0,294	0,178	0,117	0,218
GT_least	Pearson Correlation	-0,071	0,020	-0,078	0,003	-0,084
	Sig. (1-tailed)	0,216	0,414	0,193	0,489	0,174
Egocentric	Pearson Correlation	-0,103	-0,051	-0,028	-0,089	-0,077
	Sig. (1-tailed)	0,126	0,285	0,376	0,160	0,196
Absolutistic	Pearson Correlation	0,017	-0,049	0,034	0,081	0,011
	Sig. (1-tailed)	0,424	0,293	0,351	0,184	0,449
Sociocentric	Pearson Correlation	**0,201**	**0,244**	0,035	0,012	**0,148**
	Sig. (1-tailed)	0,012	0,003	0,347	0,446	0,049

Thus, on the overall (considering all contribution levels), the provision of the public good was underpinned mainly by the contributions of the sociocentrics, that is, those who identify with the FS (green) and that reject the ER (orange) value system at same time. They offerings ought to counterbalance the snubbing of ER_most and DQ_least, as well as the GT_most unwilling to contribute in this experiment. These observations support hypothesis **H2-PG** and **HF2-PG**. They offer evidence that the sacrifice-self systems are more supportive to the provision of the public-good than the express-self systems.

The disaggregation of the contribution levels show up some additional significant correlations that support part of our statements in the hypotheses **H4-PG** and **HF4-PG**. As posited, the sociocentrics (and FS_most) contributed significantly more at the (very) high (70-90) contribution level, as well as in the low (10-20) contribution level. Accordingly, the positive correlation with BO_most (medium level) and the negative correlation with BO_least (low level), and DQ_least (high and very high levels), all them reinforce the evidence that the sacrifice-self systems are more cooperative. Conversely, ER_most contributed significantly less both at the low and very high contribution levels, and CP_most also contributed significantly less at the low contribution level.

7.6.2.3. Evidences showing the underlining motivations agree with the psychosocial characterizations as set forth by the ECLET conception (hypothesis H3.1-PG, H3.2-PG, HF3.1-PG, and HF3.2-PG)

The match of individuals' expectations as to the contributions of others and their actual decision whether or not to contribute themselves for the provision of the public good revel the motives underlining the observed behavior. Table 7.26 presents the Pearson bivariate correlations between such motivations and the frequency of responses relating the values system as well as with our three principal components.

The results indicate that the motivation under the unwilling to contribute showed by ER_most and CP_most is chiefly *greed*, meaning they did not

contribute even if (or just because) they believed that enough other participants might contribute (as in hypothesis **H3.2-PG**). The same idea is conveyed by the significant negative correlations of the motive *greed* with the *rejection* of CP (CP_least) and ER (ER_least) value systems. *Greed* is also positively related with the *rejection* of the BO value system (BO_least). This result reinforces hypothesis **H3.2-PG** as well, since BO_least correlates significantly positively with ER_most (Table 7.1). We did not find significant positive correlation between *greed* and the scores in egocentrism (Factor 1), but this motive correlates significantly negatively with the scores in sociocentrism (Factor 3), as in hypothesis **HF3.2-PG**.

Hypothesis **H3.2-PG** stated that when FS do not cooperate the motive out to be *fear* but not *greed*. However, we find that *fear* indeed correlates significantly *negatively* with FS_most. This an additional indication that FS is bent to support the provision of the public good, which is reinforced by the positive and significant correlations between FS_most and the motives sense of *duty* and *solidarity*, as posited in hypothesis **H3.1-PG**. As well, sense of *duty* is also significantly positively correlated with BO_most, as supposed. Still in accord with hypothesis **H3.1-PG**, though in the inverse way, sense of *duty* correlates significantly *positively* with *rejection* of ER (ER_least) and significantly *negatively* with it acceptance (ER_most). All these explain the significant and positive correlation of the motive *duty* with the factor scores in sociocentrism (Factor 3), as stated in hypothesis **HF3.1-PG**.

A similar story holds for the motive *solidarity*, where the positive correlation with the *rejection* of ER (ER_least) reinforces the correlation of this motive with the factor sociocentrism (hypothesis **HF3.1-PG**). Moreover, the motive *solidarity* correlates significantly positively with the rejection of CP (CP_least) and significantly negatively with the factor egocentrism (Factor 1). Finally, there is also a significant *negative* and theoretically sound (amenable to hypothesis **H3.1-PG**) correlation between the motive *solidarity* and the *rejection* of the DQ value system (DQ_least).

As we saw, the game of forces involving expectations, decision and motives resulted in the public good being supplied just up to the level of 40% of necessary contributions. Since those who contributed in the higher requirement conditions lost their offerings, as well as those who did not contributed at the lower levels had an extra profit, the participants' final payoff correlates inversely with the their disposition to support the provision of the public good. Thus, the sociocentrics, which appear also as the FS_most, ER_least and CP_least variables, received smaller payoffs while the ER_most, DQ_least and GT_most accrued more. Such indications are confirmed by the significant and strong correlations between the motives and the gains (not presented in the Table 7.26). The Pearson bivariate correlation (r) between the total (nine trails) payoff and the motives under the decision to contribute are r = - .884 for *solidarity* and r = - .554 for sense of *duty*. Conversely, the correlations between the gains and the motives under the decision of not contribute are .723 for *fear* and .297 for *greed*. All these correlation are statistically significant at $p < .0001$.

Notice the intricacy of the dilemma. On the one hand, and in the measure that such payoffs bestow some surviving value, it may appear that they proffer some indication of an adverse selection which could be interpreted as approximating Hardin's suggestion that "conscience is self-eliminating". Still, on the other hand, was such conscience, as sense of *duty* and *solidarity*, really eliminated, the public good would not be provided and the extra payoff accrued by the noncontributors would vanish as well. So, the short of "conscience", thus understood, could not go without the support of those who have more of it. From this view, it seems rather that "conscience" is not only self-preserving but also the source from with those who do not have it obtain their own surviving.

Table 7.26 – Pearson bivariate correlations: motives and payoff

Value systems levels of existence	n =126	Motive of decision				Total payoff
		"Fear"	"Greed"	"Duty"	"Solidarity"	
BO_most	Pearson Correlation	-0,009	-0,058	**0,125***	0,012	-0,047
	Sig. (1-tailed)	0,459	0,258	0,081	0,448	0,299
BO_least	Pearson Correlation	-0,079	**0,209*****	-0,051	-0,078	0,073
	Sig. (1-tailed)	0,190	0,010	0,285	0,192	0,209
CP_most	Pearson Correlation	-0,065	**0,139***	0,063	-0,078	0,023
	Sig. (1-tailed)	0,233	0,060	0,240	0,192	0,401
CP_least	Pearson Correlation	0,002	**-0,142***	-0,027	**0,139***	**-0,119***
	Sig. (1-tailed)	0,490	0,057	0,381	0,060	0,091
DQ_most	Pearson Correlation	0,094	-0,111	-0,015	-0,001	-0,011
	Sig. (1-tailed)	0,148	0,108	0,434	0,494	0,451
DQ_least	Pearson Correlation	0,065	0,103	-0,062	**-0,144***	**0,169****
	Sig. (1-tailed)	0,236	0,127	0,245	0,054	0,029
ER_most	Pearson Correlation	0,011	**0,176****	**-0,171****	-0,102	**0,163****
	Sig. (1-tailed)	0,452	0,024	0,027	0,128	0,034
ER_least	Pearson Correlation	-0,030	**-0,197****	**0,218*****	**0,124***	**-0,192****
	Sig. (1-tailed)	0,368	0,014	0,007	0,083	0,015
FS_most	Pearson Correlation	**-0,151****	-0,105	**0,169****	**0,202****	**-0,230*****
	Sig. (1-tailed)	0,046	0,120	0,029	0,012	0,005
FS_least	Pearson Correlation	-0,041	0,087	-0,023	-0,022	0,035
	Sig. (1-tailed)	0,323	0,165	0,399	0,402	0,348
GT_most	Pearson Correlation	0,112	0,006	-0,097	-0,095	**0,122***
	Sig. (1-tailed)	0,106	0,472	0,139	0,144	0,087
GT_least	Pearson Correlation	0,111	-0,053	-0,042	-0,063	0,076
	Sig. (1-tailed)	0,109	0,279	0,319	0,242	0,198
Egocentric	Pearson Correlation	0,027	0,094	0,034	**-0,132***	0,114
	Sig. (1-tailed)	0,382	0,148	0,351	0,070	0,103
Absolutistic	Pearson Correlation	0,030	-0,060	0,023	0,010	-0,025
	Sig. (1-tailed)	0,368	0,252	0,400	0,455	0,390
Sociocentric	Pearson Correlation	-0,062	**-0,173****	**0,205****	**0,145***	**-0,194****
	Sig. (1-tailed)	0,246	0,027	0,011	0,053	0,015

*** Correlation is significant at the 0.01 level (1-tailed).
** Correlation is significant at the 0.05 level (1-tailed).
* Correlation is significant at the 0.10 level (1-tailed).

7.6.3. Multivariate analysis

As we did as to the CPR experiment, let us check the strength of the relationships just sketched by examining them further using multivariate analysis. As before, we shall rely mostly on the principal components.

7.6.3.1. Sociocentrics are more optimist regarding the contribution of others for the low contribution level (hypotheses HF1-PG)

Concerning the expectations, given the few and small correlations with the principal components (Table 7.24) no linear regression model using the three factors as predictor is statistically significant. However, the model in which the factor sociocentrism (Factor 3) is the only predictor is significant (F = 3.295; p = .072) for the low (10-20) contribution level. The significant coefficient (p < .10) shows that the sociocentrics (Factor 3) expect that more participants would contribute for the provision of public good at this level of required contribution. This result strengths what we have already find through the bivariate analysis.

Table 7.27 – Linear regression dependent variable: participant expected that enough participants contributed at the low (10-20) contribution level

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,908	,022		41,410	,000
	Sociocentric	,039	,022	,161	1,815	,072

7.6.3.2. Sociocentrics are more supportive of the provision of public good (hypotheses HF2-PG and HF4-PG)

The multivariate regressions using the factor scores to predict the percentages of actual contributions also confirm the earlier bivariate indications. The models for the total of the nine trails (Table 7.28) and for the low (10-20) contribution requirement (Table 7.29) are both statistically significant, with F = 2.396 (p = .071) and F = 2.839 (p = .041), respectively. In each of these models the only statistically significant coefficient strengthens the conclusion that the sociocentrics are the main responsible for the provision of the public good.

Table 7.28 – Linear multivariate regression dependent variable: participant contributed for the common pool (total of the nine trails)

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,263	,020		13,225	,000
	Egocentric	-,029	,022	-,119	-1,349	,180
	Absolutistic	,008	,021	,033	,378	,706
	Sociocentric	,048	,020	,213	2,405	,018

Table 7.29 – Linear multivariate regression dependent variable: participant contributed for the common pool (low (10-20) contribution level)

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,280	,032		8,627	,000
	Egocentric	-,028	,036	-,069	-,791	,430
	Absolutistic	-,012	,033	-,031	-,352	,725
	Sociocentric	,091	,032	,247	2,805	,006

As for the very high contribution level (70-90) the evidence comes from the bivariate regression using only the factor sociocentrism as predictor (Table 7.30). The model (F = 2.783; p = .098) shows that the higher the scores in sociocentrism the higher the frequency of contribution (p < .10). As observed during the discussion of the bivariate correlations, this result is particularly meaningful, since it confirms that the sociocentrics contributed more at the very high (70-90) contribution level. We shall also verity that this behavior comes out of sense of *duty*.

Table 7.30 – Linear regression dependent variable: participant contributed for the common pool (very high (70-90) contribution level)

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,126	,023		5,487	,000
	Sociocentric	,038	,023	,148	1,668	,098

7.6.3.3. Sociocentrics are motivated by solidarity and sense of duty and not by fear or greed (hypotheses HF3.1-PG and HF3.2-PG)

In searching the relationships with the motivations, we first attempted multinomial logistic model using the three principal components as predictors of the four categories of motives. Unfortunately, we did not find statistically significant results using this preferable procedure. As well, no linear multiple regression model using the factor scores as predictors of the relative frequency in which each motive appear separately is statically significant.

We then changed the positions of the dependent and independent variables, so that the relative frequency of each motive was used as predictors of

the scores in each of the principal components, separately. Still, owing to the existence of multicollinearity, not all motives can enter as predictors at the same time. We then selected two models using only three out of the four motives each time. Again, only the models relating the motives with the factor sociocentrism (Factor 3) were statistically significant.

The first model ($F = 3.026$, $p < 0.05$), which considers all trails, confirms that both motives under the decision to contribute for the provision of the public good (i.e. *solidarity* and sense of *duty*) are positively and significantly associated with the factor scores in sociocentrism, as in hypothesis **HF3.1-PG** (Table 7.31). Otherwise, the second model ($F = 3.026$, $p < 0.05$) shows that both motives under the decision of not contributing for the common pool (i.e. *fear* and *greed*) are negatively and significantly associated with the factor scores in sociocentrism (Table 7.32). While the negative correlation with *fear* was not predicted, the negative correlation with *greed* reinforces the hypothesis **HF3.2-PG**.

Table 7.31 – Linear multivariate regression dependent variable: sociocentrism – factor 3 (all trails)

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	-,648	,361		-1,794	,075
	Solidarity	1,054	,566	,206	1,863	,065
	Sense of duty	2,946	1,182	,243	2,492	,014
	Fear	,733	,530	,167	1,384	,169

Table 7.32 – Linear multivariate regression dependent variable: sociocentrism – factor 3 (all trails)

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
2	(Constant)	2,298	1,073		2,142	,034
	Solidarity	-1,891	1,239	-,369	-1,526	,130
	Fear	-2,213	1,072	-,505	-2,065	,041
	Greed	-2,946	1,182	-,529	-2,492	,014

7.7. How different levels of existence play the ultimatum game

In order to take advantage of having our participants already reunited for the public good experiment, and aiming to generate additional behavioral observations, we asked them to play the well know Ultimatum Game after the main experiment was over.

We then employed the simplest version of game consisting of just one shot and anonymity. The participants were divided into two groups where one side played the role of Proposer and the other the role of Responder. The experimental task consisted of dividing a fixed sum of money (R$ 10.00). Each Proposer makes a proposal to an anonymous Responder, who can either accept or reject the proposal, similarly without knowing the identity of the Proposer. If the proposal is accepted, the participants earn the amounts specified in the proposal. If the proposal is rejected, each participant receives a zero payoff.

To remind, the theoretical prediction which follows from the "rational" actor model is for the Proposer to make the lowest possible offer, say R$ 1.00. The reasoning is that the Proposer presumably knows that the Responder is a "rational" player, and accordingly will realize the R$1.00 is better than nothing. So, the model predicts that the Responders will accept that proposal. The

equilibrium outcome is thus assumed to be R$9.00 for the Proposer and R$1.00 for the Responder.

Note that we needed an even number of participants to play this game. When we had an uneven number of participants in the class room we ought to leave one out. We did not randomize this choice, though, but just asked for a volunteer. We finally get to a sample of $\underline{n}$ = 124 but this means that we have only 62 observation for each position in the game.

Our general results replicate the common finding that players systematically reject proposals below a certain threshold, presumably out of a "sense of fairness". For our experiment, the threshold was clearly defined at R$ 3.00. There were 6 proposals of R$ 3.00 being that 3 of them were rejected and 3 were accepted. Figure 7.33 presents the distribution of frequencies for the various levels of proposals. The modal value is located at the even division of the fixed amount supplied by the experimenters, that is, at R$ 5.00. There were 25 (40.3%) proposals of this amount and other 6 (9.7%) proposals above the R$ 5.00 (two proposals of R$ 6.00 and one of R$ 7.00).

Accordingly, exactly 50% of the participants offered either R$ 5.00 or more than R$ 5.00 while the other 50% offered less then the even division. Among the proposal below the R$5.00, the modal value was R$ 4.00. Four individuals had the initiative of proposing non integer amounts, since we were not explicit concerning this issue. Three of them offered R$ 4.50 and one proposed just R$ 3.25. The five proposals below R$ 3.00 together with 3 proposals of R$ 3.00 were rejected. So, on the overall, 8 (12.9%) proposal were rejected while 54 (87.1%) were accepted.

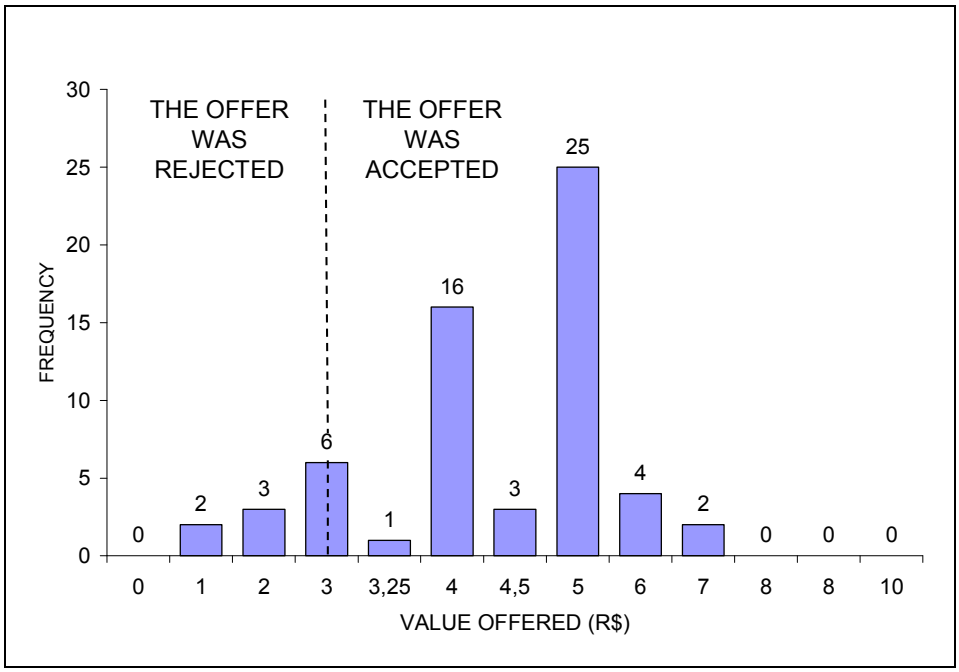


Figure 7.33 – Frequencies of values offered and the threshold of acceptance in the ultimatum game.

7.7.1. Reframing the working hypotheses

We set forth two hypotheses as to the relationship between the value systems and the behaviors responses in the Ultimatum game. Hypothesis **H1-UG** posits that CP and ER make lower offering while BO, DQ, FS and GT make higher offerings. In terms of the principal components this hypothesis implies:

- **HF1-UG**: Egocentrics make low offerings while absolutists and sociocentrics make higher offerings.

As for the Respoder, we posited that CP and DQ are more likely to refuse low offerings than the other value systems. According to the principal components we should expect:

- **HF2-UG**: Egocentrics and absolutists are more likely to refuse low offering than the sociocentrics.

We note, however, that both high and low sociocentrism is consistent with the acceptance of uneven offerings. On the one hand, FS accepts them out

of “spirit of conciliation”. On the other hand, ER accepts them in conformance with the conventional rationality of economic thinking.

As usual, we first examine the bivariate correlations and than turn to the regression analysis.

7.7.2. Bivariate analysis

The reduced number of observations turns more difficult bringing into light the statistical relationships between these behavioral responses and the values systems. Even so, the few statistically significant correlations are meaningful and consistent with the behavioral responses in the two previous experiments. Our results permit show that:

7.7.2.1. Egocentrics make lower offerings (evidence for hypothesis HF1-UG)

Table 7.33 shows that the higher the scores in egocentrism (Factor 1) the lower the players’ offering in the Ultimatum game, as posited by hypothesis **H1F-UG**. By disaggregating the principal components, we see that this result is mostly due to the fact that those who *rejected* the CP value system (CP_least) offered more while those who *rejected* the FS system (FS_least) offered less. This result is consistent with hypothesis **H1-UG**.

Also there is a significant negative correlation between the amount offered and the acceptance of the BO value system (BO_most). This result is contrary to our expectation (**H1-UG**). Still, our hypothesis was set forth only on the basis that BO is a sacrifice-self system. However, it is so “to the way of one’s elders, to the traditions of one’s ancestors”. We now realize that, clearly, BO_most does not fell at home in the experimental settings, so that its decision of keeping more money to him- or herself is not really contradictory with its characterization as a sacrifice-self system.

Table 7.33 – Ultimatum game: pearson bivariate correlations

Values system level of existence	n = 62	Value offered (R$)	The responder rejected the offering
BO_most	Pearson Correlation	**-0,265****	0,107
	Sig. (1-tailed)	0,019	0,203
BO_least	Pearson Correlation	0,070	-0,082
	Sig. (1-tailed)	0,295	0,264
CP_most	Pearson Correlation	-0,133	-0,125
	Sig. (1-tailed)	0,151	0,166
CP_least	Pearson Correlation	**0,173***	0,122
	Sig. (1-tailed)	0,089	0,173
DQ_most	Pearson Correlation	-0,047	**0,289****
	Sig. (1-tailed)	0,357	0,011
DQ_least	Pearson Correlation	-0,125	-0,094
	Sig. (1-tailed)	0,166	0,233
ER_most	Pearson Correlation	0,040	-0,072
	Sig. (1-tailed)	0,380	0,288
ER_least	Pearson Correlation	0,053	0,027
	Sig. (1-tailed)	0,340	0,418
FS_most	Pearson Correlation	0,075	0,048
	Sig. (1-tailed)	0,281	0,355
FS_least	Pearson Correlation	**-0,283****	0,036
	Sig. (1-tailed)	0,013	0,392
GT_most	Pearson Correlation	0,116	**-0,260****
	Sig. (1-tailed)	0,184	0,021
GT_least	Pearson Correlation	-0,059	0,000
	Sig. (1-tailed)	0,324	0,500
Egocentric	Pearson Correlation	**-0,233****	-0,104
	Sig. (1-tailed)	0,034	0,212
Absolutistic	Pearson Correlation	-0,146	**0,244****
	Sig. (1-tailed)	0,129	0,028
Sociocentric	Pearson Correlation	-0,023	0,016
	Sig. (1-tailed)	0,431	0,450

*** Correlation is significant at the 0.01 level (1-tailed).
** Correlation is significant at the 0.05 level (1-tailed).
* Correlation is significant at the 0.10 level (1-tailed).

7.7.2.2. Absolutists reject low offerings (evidence for hypothesis HF2-UG)

However few, the statistically significant correlations involving the behavioral observations for the Responders are rather suggestive. We initially point to the evidence concerning trivial observation that people differ in their judgment as to what should be the proper response to an uneven partition of the experimental endowment. This was indicated by the result that while half of the individuals accepted the R$ 3.00 offer, the other half rejected the same amount.

Graves's ECLET conception, nonetheless, provides a way to add new information and improve our understanding concerning which psychosocial trait is connected with such observation. Moreover, the results indicate that Graves's conception presents predictive power in this regard. Accordingly, in agreement with hypothesis **HF2-UG**, the rejections of the low offers are significantly positively related with the scores in absolutism (Factor 2). Given the variable-component correlations within Factor 2 we can presume that this response is positively related with DQ (blue) and negatively with GT (yellow). This is confirmed by the significant correlations with these value systems, as posited in hypothesis **H2-UG**.

In order to investigate these relations further, we plotted the individual profiles of the six participants who have received an R$ 3.00 offering. To the left of Figure 7.34 there are the principal components profiles of those who have rejected the offering. To right there are the profiles of those who accepted it. The pattern for the Factor 2, at once pro-DQ and anti-DT (Table 7.4) is clear, even if that which explains the different responses for the first individual in each group is possibly not the scores in absolutism (Factor 2), which are similar, but the differences concerning the levels of egocentrism, since both high and low sociocentrism are in principle consistent with the acceptance of the low offering. On the other hand, the third individual who has accepted the low offering present the highest score for egocentrism (red), among these six individuals. We note, however, that he or she also present the lowest scores for both absolutism (blue)

and sociocentrism (green). As pointed out, both traits are consistent with the choice to accept the uneven division in the Ultimatum game setting.

This analysis of the individual profiles highlights the intricacy involved is such evaluations. Notwithstanding, it also indicates that the finding of statistical significant correlations, however moderate, is a non trivial result to suggest the robustness of Graves's ECLET conception.

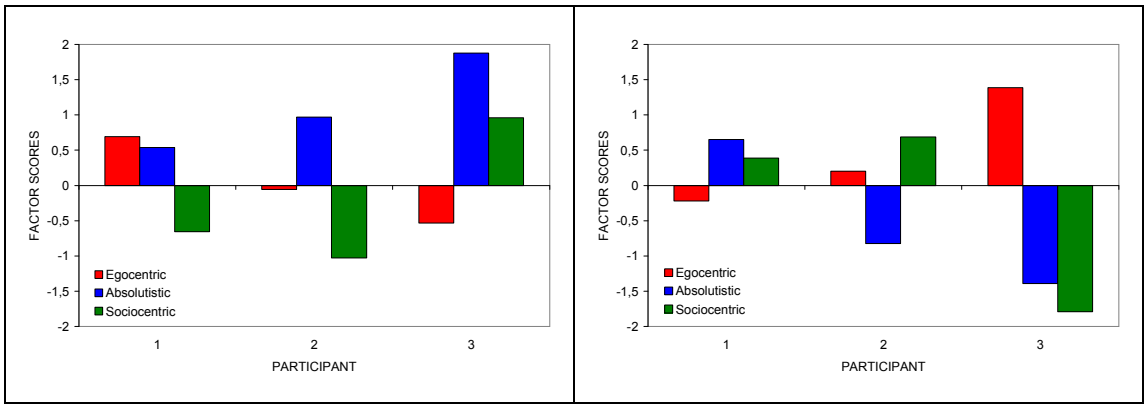


Figure 7.34 – Individual levels of existence profiles of the participants who either rejected (left) or accepted (right) the threshold offering of R$ 3.00.

7.7.3. Multivariate analysis

7.7.3.1. More on egocentric greediness (evidence for hypothesis HF2-UG)

Once more, multiple regressions models including all three principal components did not show statistical significance. This is not to wonder in view of the results for the public good experiment and that our sample is now reduced to a half than before (62 observations for each behavioral response).

Even so, the strength of the negative relationship between the scores for egocentrism (Factor 1) and the values offered by the Proposer sustains the statistical significance for the regression model in which egocentrism appear as the only explanatory variable (F = 3.431; p = .069). Accordingly, the result

presented in Table 7.34 substantiates the relationship suggested by hypothesis **HF1-UG**, that is, the higher Proposer scores in egocentrism the lower it is his or her offer in the Ultimatum game.

Table 7.34 – Linear regression dependent variable: value of the proposers' offering

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	4,377	,151		29,019	,000
	Egocentric	-,300	,162	-,233	-1,852	,069

7.7.3.2. More on absolutists' taste for punishing (evidence for hypothesis HF2-UG)

As for the behavioral alternatives of the Responder, the binary logistic regression indicates that the higher the scores in absolutism (Factor 2) the higher the probability of the Responder reject an offer. As we saw, there were only 8 proposals which were rejected out of a total of 62 proposals. Given this, the logistic model which includes only the constant as explanatory variable is able to predict correctly 87.1% of the cases pertaining to each category. The explanatory contribution supplied by the factor absolutism is thus necessarily small. Thus, by settling the cutting point at .3, in order to take the mentioned imbalance into account, the percentage of correctly predicted categories increases to 88.7% in the model including the three factors as predictors. In spite of this, the effect due to the scores in absolutism high (.813) is significant at 10% of probability (Table 7.23).

We remind that, while the rejection of uneven offering connects with "sense of fairness", in the Ultimatum game it also reveals a "taste for punishing" those who do not conform to that norm, since the Responder has nothing to loose by accepting any positive amount, as asserted by the rational choice model. We claim that this result is consistent with both **H2-UG** and **HF2-UG** hypothesis.

Table 7.35 – Binary logistic regression dependent variable: the responder rejected the offering

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Egocentric	-,3632	,525	,478	1	,489	,696
	Absolutistic	,813	,456	3,174	1	,075	2,255
	Sociocentric	,026	,410	,004	1	,950	1,026
	Constant	-2,341	,530	19,472	1	,000	,096

[a] Variable(s) entered on step 1: Egocentric, Absolutistic, Sociocentric.

7.8. Discussion and limitations

Now that we have finished the general description and the most basic analysis of our results, I would like to emphasize the meaning of our findings, as it can be interpreted in light of both the ECLET conception and the Integral Theory. Such understanding is grounded on three interrelated aspects, namely: (1) the cross-cultural attribute implicated in the reliability of our findings, (2) the fundamental nature of Graves's ECLET conception, as a member of the broader field of developmental stage theoretical psychology, and (3) the rightness of interpreting Graves's constructs and our findings as *proxies* of a more general process of interior transformations, spiritual in nature, and in which the freedom from self-centeredness and the emancipatory power ensuing from increasing self-

control and capacity to act on behalf of the collective good are discernible traits associated with the higher levels of interior development.

All this is supported by Graves's ECLET conception itself. In this regard, the key contribution from the Integral Theory concerns the rescue of the profound meaning of subjective reality, from the modern decline as *mere* subjective, toward its restoration as the inseparable complement of the objective half-side of reality – truly and radically integrated to it. Such is the imperative epistemological integration, brought about by Wilber's Integral Theory, and which we regard as vital for the matter of approaching our contemporary dreadfully serious social dilemmas in more balanced ways, i.e., honoring the variety of forms of knowing and being-in-the-world, whereas avoiding sterile relativism by assembling the diversity in proper perspective.

In this broad context, I claim that our results bestow experimental evidence in support of Wilber's statement involving the centrality of the evolution of consciousness in connection with our present environmental crisis and most general social tragedy. Clearly, in view of the much greater complexity of the real-life situations compared with our experimental settings, the limitations of our findings must be pointed out. Ultimately, though, I claim just that we were able to discern the contributions due to the Gravesian Levels of Existence constructs for the explanation of the behavioral observations from those attributable to the different institutional settings and experimental situations. It should be noticed, as well, that we were able to do so not owing to any special feature of our experimental design – apart from the previous information concerning the psychosocial systems of our participants – but precisely because of that that different psychosocial systems should respond differently to the same action situation and institutional settings are but an essential feature of Graves's model itself. Likewise, as our findings are shored up by the understanding that human psychosocial development has its own internal rules of transformation, and is not simply a function of the external circumstances of action, following some presumable basic form of (instrumental) rationality, they also support the claim respecting the appropriateness of exploring new ways to deal with social

dilemmas, namely, not only through more complete analysis of the exterior-systemic circumstances of action – out of which the number of important influencing variables never stop increasing – and the exploitation of some version of the basic carrot-stick model for changing unsustainable behaviors, but also by fostering human natural potential (and right) for interior transformation and spiritual awakening.

Again, these general propositions ensue directly from Graves's model. The additions from the Integral Theory relates mostly with the way that Wilber's account of Graves's conception allows us to see it as a *proxy* of a much more general phenomenon, whose profound and essentially spiritual meaning cannot be grasped if not by acknowledging Integral Methodological Pluralism, in the root of Wilber's epistemological integration.

7.8.1. Our findings impart strengthened cross-cultural reliability to Graves's ECLET conception

We first remind the reader that Graves intended his ECLET conception to be "applicable to any adult human being, regardless his culture". Moreover, it would be applicable in the very sense of providing the "applied social scientist and social planners with a different means to the end of comprehending and approaching human problems than they have had in their disposal before" (GRAVES, 2005, p. 4). This endeavor has been undertaking in worldwide thriving applications of the ECLET conception to cope with both matters of organizational change and open social maladies such as the South African apartheid (BECK; COWAN, 1996).

In regard to our present contribution, we have already stressed out our initial concern respecting the translation of the assessment tool and the application of the Spiral Dynamics survey to our sample of Brazilian students. In this sense, the test instrument developed by Hurlbut (1979) during her Ph.D. was of course expected to be valid only for speakers of Standard American English and for persons with the equivalent of high school education. Yet, following the

standpoint of the contemporary evolutionary psychology, we have also argued that cultural differences are usually overrated, as both cultural and individual variability stands against a backdrop of human universals.

Beached on such understanding, we undertook the translation process and put into action our research strategy, convinced of its general promise. However, there was absolutely nothing guaranteed concerning neither the likelihood of extracting meaningful factors from the psychosocial systems covariance matrix obtained from our application nor, of course, of finding theoretical sound relationships between the scores of the separate variables or the factor scores and behavioral observations in the experimental settings. So, in the measure that we got hold of both, as alleged, we claim further that our findings indeed bestow renewed cross-cultural evidence in support of Grave's ECLET conception.

7.8.2. The experimental observations verified the cyclical turn of the agency *versus* communion capacities described by the ECLET conception

Apart from the special situation involving the GT level of existence – which is itself theoretically sound – the overall observations coming from all three experiments widely concur with the expectations associated with the cyclical dominance of either the agency or the communion capacity of the self at the subsistence levels, as proposed by the ECLET conception.

In this regard, we were able to show that the go-getting and self-concerned dispositions are consistently associated with the choices attributable to both express-self systems: CP and ER. Conversely, the self-restraining and cooperative disposition consistently correlates with the sacrifice-self systems: BO, DQ, and FS. The evidences in this respect ensued from both the affinity and the rejection of the corresponding psychosocial systems.

Accordingly, the affinity with CP (CP_most) correlated significantly positively with the frequency of defections in the CPR experiment (Table 7.9), negatively with the frequency of contributions (low contribution level) and

positively with the motive *greed* in the Public Good experiment (Tables 7.24-26). Complementarily, the *rejection* of CP (CP_least) correlated significantly negatively with the frequency of defections and with the overinvestment during the communication condition of the CPR experiment, negatively with the motive *greed* and positively with *solidarity* in the Public Good experiment. The rejection of CP egocentric stance also correlated significantly positively with the value offered by the Proposer in the Ultimatum game (Table 7.33).

Likewise, the affinity with the go-getting character of ER (ER_most) appeared as significant positive correlations with overinvestment, both in the baseline and communication conditions of the CPR experiment, as well as with the frequency of defections (Table 7.9). The ER stance also correlated significantly negatively with the expectations regarding the contribution of others (low contribution level) and with the frequency of the actual contributions in the overall to the nine trials of the Public Good experiment (Tables 7.24-26). By disaggregating the contribution levels we were able to expose the strategic thinking of ER. Through this, we saw that ER contributed less frequently especially at the low and the very high levels of required contributions. Therefore, the ER psychosocial system correlated significantly positively with the motive *greed* and significantly negatively with the motive sense of *duty*. All these observations were reinforced by consistently reversed correlations with the rejection of the ER stance (ER_least). Hence, ER_least correlates significantly negatively with overinvestment in the baseline setting and with both overinvestment and the frequency of defections in the communication condition of the CPR experiment, and correlates significantly positively with the expectations regarding the contributions of others (low contribution level) and with actual contributions (on the average of all trials and for the very high contribution level) in the Public Good experiment. Accordingly, the rejection of ER standpoint correlates significantly negatively with the motive *greed* and positively with the motives *solidarity* and sense of *duty*.

All these observations concur with the agency drive that superimposes the communion capacity of the self at the CP and ER stages of psychosocial

development, according to the ECLET conception. Therefore, coercive means are certainly necessary if the task is to uphold the collective outcome in groups where either the CP or the ER psychosocial systems are likely to have non negligible influence. We acknowledge, moreover, that this might be expected in most of the practical circumstances. Then, our results showed further that both CP and ER can be put into line through the use of the sanctioning mechanism (Table 7.9). But we also saw that defiant defections under the sanctioning conditions can likewise be attributed to CP and ER, and that the costs associated with the sanctioning mechanisms may force the economic performance down to levels below the ones occurring at the unregulated setting.

Conversely, we also showed that higher levels of cooperation and economic performance are connected with the choices of the sacrifice-self systems. The statistical significance of these findings is largely associated with participants' differences involving the FS psychosocial standpoint. Nevertheless, the fewer statistically significant associations involving the BO and DQ systems also agree with the theoretical expectations.

Accordingly, the subjective proximity to the BO level of existence (BO_most) correlated significantly positively with the frequency of contribution (medium contribution level) and with the motive sense of *duty*, in the Public Good experiment (Tables 7.24-26). This cooperative disposition is reinforced by reversed correlations with the rejection of BO system (BO_least). Thus, BO_least correlated significantly negatively with the decision to contribute (low contribution level) and significantly positively with the motive *greed*. The positive correlations between the rejection of BO and the expectations regarding the contribution of others (high and very high contribution levels) are less meaningful owing to the issue mentioned before (disperse correlations with the other value systems). Also, BO_most correlated significantly *negatively* with the amount offered by the Proposer in the Ultimatum game (Table 7.33). From the previous results, we know that this should not be understood as a sign of unwillingness to cooperate. Indeed, as in the one-shot Ultimatum game what is at sake is not cooperation itself but sense of equity and fairness, this occurrence

harmonizes with the system specific trait of BO systems. I suggest that these aspects interact more directly with the individual's needs and strategic thinking – for the evaluation of the risk of having a low offer rejected by the Responder. So, as Graves's theory associates BO with the most basic wants and low-grade strategic thinking, which if fully opened only at ER stage, we regard the aforementioned negative association between BO and Proposer's offerings perfectly understandable from the angle of the ECLET conception. Finally, in the CPR experiment, BO_most correlates significantly negatively with the payment of fines while BO_least correlates positively with both the payment of fines and the frequency of defections (Table 7.9).

As for the DQ system, the only significant correlation in the CPR experiment connects the rejection of the absolutistic standpoint (DQ_least) negatively with the level of investment during the sanctioning condition (Table 7.9). The inference in this regard is that the lower the absolutist sturdiness the greater the responsiveness as to the changing of the institutional setting. We found no significant correlations between DQ_most and the frequency of contributions in the in Public Good experiment, but we do found three significant negative correlations connecting the rejection of DQ with the frequency of those contributions (Table 7.25). Two other significant correlations involving the expectations regarding the contribution of others (Table 7.24) are less evocative, since both the acceptance and the rejection of DQ correlated negatively with them (at the medium contribution level, for DQ_most, and at the high contribution level, for DQ_least). From the factor loadings matrix (Table 7.4) we can infer that the last association may possibly be reflecting either CP or most probably GT expectations. Finally, we have already mentioned that the identification with the absolutistic stance helped to explain which trait was associated with the rejection of low offerings in the Ultimatum game (Table 7.33).

However few, the significant correlations linking the behavioral responses and the BO and DQ psychosocial levels of existence concur either with their communion bent or with some system specific trait that agrees with

Graves's characterizations. This general conformity is significantly strengthened by the correlations with the FS system. Accordingly, the cooperative disposition expected from FS appears as significant correlations connecting *negatively* the affinity with its social value orientation (FS_most) with the frequency of defections and the level of investment in the Market 2 during the communication condition of the CPR experiment (Table 7.9). Complementarily, the rejection of the FS system (FS_least) correlates significantly positively with both the frequency of defections and overinvestment during the communication condition. Since the responsiveness to the imposition of the sanctioning mechanism is common to the various value systems, the level of investment during this treatment correlates negatively with both the acceptance and the rejection of FS orientation. The system specific trait tied to FS appears in the application of sanctioning. In this regard, we have explained the FS is disinclined to use coercion. This trait showed up both in the significant negative correlation between the expenditures with fees and FS_most, and the significant positive correlation between those expenditures and the rejection of FS (FS_least). Also, the rejection of FS correlated positively with the total payment of fines.

In the Public Good experiment (Tables 7.24-26), the proximity to the FS level of existence connects significantly positively with the decision to contribute for the common pool (at the low and very high contribution levels as well as in the average of the nine trials), with the motives *solidarity* and sense of *duty*, and significantly negatively with the motive *fear*. The *rejection* of FS (FS_least) correlated positively with the expectation as to the contribution of others at the very high contribution level, in the Public Good experiment (Table 7.25) and negatively with the value of Proposer's offer in the Ultimatum game (Table 7.33).

In short, our experimental results are in great measure concurring with the theoretical expectations and thus supporting the ECLET conception regarding the cyclical emergence of agency and communion capacities of the self. Yet, that the Gravesian constructs are *not* personality *types* but *levels* of psychosocial development configures just the most essential character of Graves's theory, and

tied to our present reviewing of the commons dilemma. Moreover, we saw through Wilber's extensive research how Graves's constructs and general schema concur with a wide range of other developmental systems that focus on other developmental *lines*. I then suggest that the general theoretical agreement just exposed extends toward the fundamental stage developmental nature of the ECLET conception, in order to claim that the willingness to cooperate is also a matter of psychosocial development, so that it can be fostered and may be helpful to solve collective action dilemmas independently of the external circumstances of action.

7.8.3. Our experimental observations show that decreasing egocentrism and increasing willingness to cooperate underline the psychosocial development of adult human beings

To say that the developmental level of existence may assist the solution of collective action dilemmas independently of the external circumstances of action is not the same as saying that this is expected to be so in *all* circumstances – and even less that individuals become increasingly insensible to the contingencies of action. To be sure, the very nested structure of the psychosocial systems comprises increasing behavioral freedom as the individual moves up the hill of interior development. Of course, this confers *more* flexibility to adapt to the contingencies of action, so that, in general, those circumstances can be expected to play an *increasing* role in influencing behavior. This trait is most visible in the passage from the absolutistic stance, at DQ, to the strategic (and thus contingent) behavior at the ER stage. It was also particularly emphasized by Graves himself when describing the momentous jump from the *subsistence* group levels of existence (AN to FS) to the second spiral defined by the *being* levels (GT and HU).

But when it comes to the particular issue involved in the collective action dilemma, namely, the conflict between the short-term individual interest vis-à-vis the long-term collective one, the behavioral freedom expresses itself precisely through the ability to transcend the egocentric drive in order to be *capable* of

behaving on behalf of the collective good. Incidentally, Hardin properly remind us of this subtle facet of free will, when he quotes Hegel's saying that "Freedom is the recognition of necessity" (HARDIN, 1993, p. 18). Yet, as we saw, Hardin misinterpreted Hegel as if the philosopher intended to mean "recognition of necessity of being coerced", instead of be passing on the emancipatory realization of the desirability of freely exercise self-restraint.

So, resuming to our experimental results, we found that the earlier stage of adult psychosocial development, as represented by the egocentric wave (Factor 1), is significantly uncooperative, whereas the latter stage, i.e. the sociocentric wave (Factor 3), is significantly more cooperative. We don't need inflating the fact that the ER system is higher than both CP and DQ but is still basically uncooperative, because we know that this is a trait associated with the cyclical turn of the agency and the communion bents in the ECLET model. Moreover, even if weakly, the acceptance of the ER system is present in the structure of the egocentric wave, whereas its rejection secondarily appears in the absolutistic wave and is a primary characteristic of the sociocentric wave (Table 7.4). Likewise, we have already argued that the negative correlation involving the GT system and the frequency of decisions to contribute in the Public Good experiment is not in disagreement which GT's conditional disposition to cooperate, since in our Public Good experiment trustworthiness is not an issue, and, apart from that, GT is expect to express what self wants.

So, as pointed out, we can better grasp the developmental aspect of the willingness to cooperate by comparing the behavioral observations tied to both the egocentric (third subsistence) and the sociocentric (sixth subsistence) stages or levels of existence. Beginning with the egocentric level (Factor 1), we saw that it correlates significantly positively with the frequency of defections in the CPR dilemma. This was shown through both Person correlation (Table 7.9) and multivariate regression analysis involving all three principal components (Table 7.11). Results from binary logistic regression using all three components as explanatory variable exposed that the likelihood of taking the initiative to defect is higher for the egocentric wave (Table 7.13). Finally, the significant negative

correlation with total invested in the CPR during the sanctioning condition reinforces the idea that the egocentrics respond to coercion accordingly. Although no correlation involving the egocentric wave and the frequency of contribution in the Public Good experiment is statistically significant, all of them are negative, and the scores in egocentrism correlated significantly negatively with the motive *solidarity*. Finally, in the Ultimatum game, the egocentric wave correlated significantly negatively with value offered by the Proposer. All these are evidences that the third stage or level of psychosocial existence is self-centered and noncooperative.

On the other hand, when the adult human gets to the sixth psychosocial stage the behavioral observations tell a quite different story. Thus, the sociocentric wave of existence (Factor 3) correlates (Pearson) significantly negatively with the level of appropriation from the CPR in both the unregulated and the unbinding communication conditions (Table 7.9). The significance of this last relationship was sustained under multivariate regression analysis including all three principal components as explanatory variables. Still concerning the CPR experiment, the significant negative relationship between the sociocentric wave and the frequency of defections appears through the Pearson correlation (Table 7.9) as well as in the multiple regression analysis (Table 7.11). The binary logistic regression showed that the likelihood of being a cheater is significantly lower for the sociocentric wave (Table 7.12). Conversely, the likelihood of sustaining the joint strategy by sticking with the self-organized collective agreement – even when others are deviating from it – is significantly higher for the sociocentric wave (Tables 7.13 and 7.14). As it should be in this circumstance, the sociocentric wave relates significantly negatively with the payment of fines. This was shown both through Pearson correlations (Table 7.9) and regression analysis (Table 7.15). Finally, the system specific character of the sociocentric wave regarding its distaste for coercive means was confirmed by the negative association with the application of fines (variable total expenditure with fees). This was verified through Pearson correlation (Table 7.9) as well as using regression analysis (Table 7.16).

The enhanced willingness to cooperate attributable to the move toward the sixth level appears clearly also in the Public Good game. Accordingly, the significant higher disposition to cooperate for the common-pool was verified through both Pearson correlation (Table 7.25) and multivariate regression using all three components as predictors (Tables 7.28-30). This was shown for the both the low and the very high contribution levels as well as for the average of the nine trials. Accordingly, the sociocentric wave correlated significantly positively with the motives *solidarity* and sense of *duty* and negatively with the motive *greed* (Table 7.26). Once again, this relationship remains significant when tested through multiple regression analysis, where the motives appear together as predictors of the scores in sociocentrims (Tables 7.31 and 7.32).

Therefore, summing up our findings, and as far as the ECLET conception endows us with a good theory of psychosocial development, I believe we provided sufficient evidence that greed, self-centeredness and untrustworthiness decreases, whereas solidarity and willingness to cooperate for the collective good increases as the biologically mature human being moves upward the hill of interior evolution. I want now to expound more explicitly the elements supporting our claim about the relative independence of the manifestation of this trait from the circumstances of action.

7.8.4. The evidences support that psychosocial development is a separate factor in solving collective action problems

At this point of our discussion, the attempt to expand on the separate implication of the psychosocial development on the outcome of collective action dilemmas runs the risk of overdoing, for the groundwork of our approach is of course that all participants are submitted to the same experimental conditions. So, the diversity of the responses under the controlled experimental settings clearly ought to be attributed to individual variability. Moreover, we have already reiterated that such individual variability is not been subsumed within broad domains of personality traits but expresses stages of psychosocial development.

However, the current academic environment seems dominated by the belief that, in general, no systematic differences in the behavior of individuals can be attributed to stable traits of character, and that such traits are easily overwhelmed by situational factors. Thus, it is usually claimed that no matter what personality traits people may have, whether they stop to help another person turns out to be almost entirely determined by how much of a hurry they are in (Latané, and Darley 1970). Yet, such assertion overlooks the fact that what counts as something that should put someone in a hurry is itself determined by those personality traits. In particular, the ECLET conception puts forward a hierarchy of needs and concerns, underlining and structuring the formation of diverse value systems and worldviews that interact and weight the strength with which different situational factors are supposed to bring out behaviors in different individuals.

So, concerning our observations, we just showed that the importance of the situational factors (i.e. the communication opportunities and the sanctioning mechanism) for group performance in the CPR experiment depends on the psychosocial profile of the individuals in the group (e.g. Figure 7.4). It is true that the direct comparison of groups is rather unclear owing to the great variability within the groups themselves (Figure 7.3 and Figures 7.5 through 7.29). Notwithstanding, the statistical analyses confirmed the significance of the relationships advanced during our exploratory visual inspection. In particular, we showed that communication is not that effective if either the egocentric (CP) or the strategic go-getting (ER) drives determine behavior, just as the standard noncooperative game theory predicts, whereas the general positive effect of communication is tied to the incidence of the sociocentric wave (FS) (Tables 7.10 to 7.12). Besides, other than the simple exhibition of the effects of these different predispositions, we have provided a reputable theoretical account of why the effectiveness of communication interacts with the psychosocial systems. Even if only superficially expounded, we called attention to the connections with Habermas's theory of communicative action and Wilber's account of the two

(horizontal and vertical) dimensions involved in the hermeneutic circle of mutual understanding.

Likewise, we saw that despite the general improvement of the collective outcome due to the introduction of the sanctioning mechanism – when compared with the baseline condition (Tables 7.5 and 7.7) – its end result on the collective performance was quite variable depending on the psychosocial profile of the group. Indeed, the sanctioning mechanisms was an unwanted instrument when those profiles were marked by low egocentrism and dominated by either the absolutistic or the sociocentric waves of existence (Figures 7.5 through 7.10 and Tables 7.13 to 7.16).

The same decisiveness of the subjective stance, in opposition to that of the situational factors, was evaluated through the assessment of the willingness to cooperate for the public-pool as the requisite contribution level varied. Such effect is summarized in the associations between the psychosocial systems and the motives underlining individuals' decision (Table 7.26), since these motives already combine the actual decisions with the expectations involving the nine contribution levels. Again, the general observation shows that the contribution levels import (Tables 7.22-23 and Figures 7.31-32). The matter is why and how they are brought in, as they interact with different subjective dispositions. Therefore, the significant correlations show very clearly that the motives under the decisions whether or not to contribute for the common-pool correlate with the stable traits of character – as the stages of psychosocial equilibrium – instead of being primarily determined by the contribution level – the situational factor.

Despite the aforementioned intellectual atmosphere, the fact that social value orientation is an important behavioral-determining factor in social dilemmas is neither new nor our main point. As noticed, the issue has been shown elsewhere (MESSICK; McCLINTOCK, 1968; McCLINTOCK, 1978), and is generally understood. Our point, once again, is that the *developmental* approaches infuses the subjective side of the coin with its own internal rationality, structure and dynamics, which obviously do not evolve apart from the exterior domain but are similarly not determined by the situational factors alone.

Finally, to conclude our discussion, I would like to recapitulate the opening issue regarding the epistemological status of first-person knowledge in order to bring back the claimed link between interior development and the perennial spiritual wisdom, now reinforced by our experimental results.

7.8.5. Nothing new under the sun: the validity claim of mysticism

As stated since the introduction, this inquiry on the implications of the psychosocial development of adult human beings on collective action problems was motivated primarily by the connections one can ascertain between the moral issues drawn in both subjects, on the one hand, and the enduring ethical precepts ensuing from virtually all spiritual traditions, on the other. The suitability of establishing such a link is not just a personal conviction, but an essential trait of the Integral Theory itself. On the side of conventional science, we saw how Wilber puts together a significant number of empirically grounded developmental models that tell us pretty much the same story, where (line-specific) psychological development is described as a series of stage-like states of equilibrium comprising a regular process of *descentering* of the self. Out from this basic descentering process, the ego-serving stances potentially decrease whereas psychological autonomy increases together with the concern for a wider and wider circle of beings. In fact, the promise of exploring the positive relationship between adult psychosocial development and the strengthen of the willingness to cooperate – in order to assist policy making directed to unravel collective action dilemmas – constitute the very substance of the thesis defended in this study. Accordingly, I just claimed we have verified experimentally such a positive relationship, even if somewhat dimmed, due the cyclical unfurling of either the agency or the communion capacities of the self along the stage developmental sequence described by the ECLET conception.

I now want to reiterate our further claim connecting these observations with the moral content of the precepts entrenched into virtually all spiritual traditions – which likewise encourage moving away from selfishness and

deceptiveness, and toward trustworthiness, compassion and care for others – to ascertain that such conformity is not coincidental. As I have repeatedly suggested, the spiritual sages of all the times are not simply propounding virtue theories willy-nilly. In so far as morality is concerned, these precepts just pass on the quintessence of the spiritual awakening, which is an achievement of the farther reaches of human potential. Far from being an unwarranted inference, this connection is in the core of the Integral Theory, and coheres pretty much with the ECLET conception as well.

Nonetheless, aside the already enough intricate issue involving scholar prevalent relativism and moral skepticism; there is also the additional difficulty that the moral precepts entrenched in the spiritual traditions typically come together with a knowledge foundation which is promptly dismissed by the contemporary scientific thought as just unauthorized. But we have already expounded that this epistemological attitude comprises a subtle form of reductionism, which disregards the fact that the spiritual knowledge is grounded on the very same strands of any valid knowledge, namely: practical injunction, apprehension, and communal checking[60]. The terrible tricky point is that subtle reductionism consists in the failure of acknowledging precisely this, for it does not grant validity to the communal checking of whatever so-called record supposedly gathered through introspective inquiry. Still even trickier is that subtle reductionism is sustained by the fact that we are ordinarily clueless a propos of what sort of data such introspective inquiry is supposed to disclose in any case.

But now I regard as plainly clear that one shall never know this if one does not give it a try. As Wilber puts it, the contemplative practices such as meditation and many others (e.g. zazen, shikan-taza, vipassana, satsang, darshan, and so forth) are not things to think – they are things to do! These are practices which usually take years to master, pretty much like the theoretical education and

[60] As for shadow spirituality, blindly believing, or mere formal religiosity, the dogmatic character of their precepts and at times bizarre rituals – even if their contents and original principles might be upright – just parallel the certainties and ceremonies we easily find in the domain of science as well, often similarly sustained on the grounds of first-person authority and prestige.

practical training that must be carried out before one may be apt to properly conduct and take to mean the data produced through conventional experiments. Furthermore, if the problem *was* one of first-person testimony, the opportunity for engaging in introspective inquiry and checking out the data they disclose are in fact open to everyone, whereas those concerning the experiments of the conventional science, increasingly reliant on sophisticated and expensive means, is virtually out of reach for the vast majority of the people, including within the academic community.

If so, why such alleged valid records of the spiritual traditions are so much arcane then? But this is just what the ECLET conception helps us to understand. The issue is linked with the hierarchical structure of needs and concerns that underlines, first, the biological survival and then the further psychosocial development of adult humans. Moreover, we saw how this structuration coheres with the epistemological integration provided by Henry Bergson, when he discriminates between the function of the intellect, on the one hand, from that of intuition, on the other, as the opportunity for working out the intuitive competence, in order to carry out conscious and sustained introspective inquiry, does not come along until the intellect – as the tool of conscious action – can be put at rest. The same is put otherwise by Wilber's categorization of cognition into gross, subtle and causal (Figure 2.15).

Thus, reviewing Graves's schema, we saw that at whichever *subsistence* level of existence the locus of attention is turned outwardly (due the pressing of the needs of action, supported by the intellect), no matter if the individual experiences the unfurling of either the agency or the communion capacity, or if the chief concern rests upon the satisfaction of basic needs, material comfort, or social status and acceptance. According to Graves's model, the time for seriously *begin* the interior exploration (supported by intuition) only arrives at GT, the first *being* level of existence, passes through HU, and should potentially continue toward higher stages. So, a propos of the GT level, Graves says: "Thus far", "man [woman] has been just another animal, a pawn in the hand of the spirit world, a sacrificer of the self, an attacker of the world and other men, and a social

automaton; but *man has never been himself*. Here [at the GT level and not before] we step over the line which separates those needs that man has in common with other animals and those needs which are distinctly human" (GRAVES, 2005, p. 367, my italics). As for the eightieth (or second *being*) level of existence (HU), Graves says that the individuals moving toward this stage are "beginning to think in a way that intuition, subjectivism plays a great deal more in their behavior than in any of the other systems" (GRAVES, 2005, p. 396). Pertaining to the value systems and morals, Graves states the following a propos of the HU level:

> Since eighth-level man need not attend so much to the problems of his existence (for him they have been solved), he values those newer, deeper things in life which are there to be experienced. He values escaping '…from the barbed wire entanglement of his own ideas and his own mechanical devices…' He values the 'marvelous reach world of context and sheer fluid beauty and [fearless] face-to-face awareness of now-naked-life…' Perceiving the world as somewhat beyond his ken, there is a serious, stable cast to the values of eighth-level man. Cooperation and trust are most seriously valued to the extent that he will withdraw from relationship that cannot be based of such.
> Play, exhibition, receiving the plaudits of others, mean little if anything to man at this level. It is not that he cannot play, nor is it that he cannot or won't dominate. It is that he prefers serious endeavor and cares not to dominate. He does not value adjusting to the world as authority says it is; nor does he value the imposition of his self upon the world. What he values is adjusting to the world as he senses it to be (GRAVES, 2005, p. 398).

Of course, any individual development shall follow different and even radically dissimilar pathways. Even if the stage developmental theory sets forth a hierarchical structure of levels, the prerequisites for moving up through the stages are in essence minimalists. While individuals' potential certainly varies, the road is quite open to everyone. Accordingly, the ECLET conception provides us with an overall historical (phylogenetic) reconstruction of the adult human psychosocial development, which presents cross-cultural robustness, as we saw, without specifying the conditions underlining any individual developmental course. In this sense, if different life conditions somehow help one to detach from pointless worries and concerns, the person may more naturally and smoothly, say, travel from the first spiral toward the second spiral of existence. *This* is why every spiritual tradition *prescribes* such detachment. It is not to justify poverty and authorize melancholy. Following Bergson's theory, we can easily notice that

such detachment from whatever outwardly concern is essential in order to put the intellect at rest and thus unshackle the power of intuition, i.e., the possibility of apprehending concrete reality directly, without whatsoever conceptual representation. In any case, the ability to engage into this sort of introspective inquiry must as well be worked out. As the practitioner learns how the stay conscious while *not thinking*, in profound meditation, the amazing order ingrained in the very human body, which subsumes the entire evolutionary process, just start unfolding before the eyes of the intuitive mind. These sorts of visions are not merely subjective or private, for what they illuminate is our common ground. Accordingly, they are plainly reproducible and can be shared by everyone who performs the experiment.

As just put, the detachment from competitiveness and whatever mindset that turns the mind outwardly is a basic prerequisite for achieving truly deep meditative states, for there is no satisfactory technique-like substitute to make the mind calm (HONGZHI, 2003, p. 334-341). Nonetheless, the connection with the moral attitude toward life, in general, and the human struggles, in particular, ensues as well from the other way around, as the meditative inquiry eventually leads the practitioner to *experience* not only the essential unity of everything with everything else – the same unity system theory proclaims and describes *conceptually* – but also to realize the actual position of *the* human holon, amid the energetic, indivisible and ever-flowing, scaffolding of causation and creation. This position, as already expanded on, can be conceptually located in the very center of the AQAL map, at the level of noosphere. This means, on the one hand, that the human holon transcends-and-includes the entire physiosphere and biosphere, so that the skilled person in profound meditation, despite of eventually being completely uneducated, is able to *experience* the concrete reality which underlines the most sophisticated conceptual representation provided by system theories. On the other hand, this means that the human holon is defined neither by the individual (you or I) nor the collective (we); neither by its interior (human mind) nor by its exterior (human body) presentations. It can only be defined, paradoxically, in these four domains simultaneously.

So, the reason why these experiences have high potential to foster compassion and wipe narcissism and egocentrism completely out is that, at least, one realizes the pathetic character of our ordinary concerns and self-interest, when viewed from the perspective of the universal flowing. Yet, I am not suggesting that these experiences *necessarily* have the effect of transforming the individual into a "saint", even if someone will hardly become one without crossing this line. Peak experiences and occasional alternated states of consciousness usually do not suffice to stabilize the interior transformations that characterize the higher stages of authentic spirituality. Again, this is so precisely because, in the ordinary life, the person continues *attached* to her outwardly concerns. Moreover, the mystical experience has to be interpreted so as to make sense within the ordinary worries and cognitive complexity that defines the psychosocial center of gravity of the individual, and may thus reinforce either superficial spirituality, false notions, or even fundamentalism and intolerance. However, when achieved *as* a higher and stable level of existence, and even if there is the risk of being trapped by pathological expressions as well, the healthy expression of these levels of existence is definitely marked by deep concern for life in general and compassion for others.

7.8.6. Limitations

The attempt to carry out the inferences from out experimental results this far is certainly vulnerable to a number of limitations. These limitations involve both issues which are inherent to the assessment of the interior domain to the subjects and issues related with our research conditions and procedures.

Concerning the first, the study is limited by the fact that attitudes, beliefs, and thought patterns are not accessible for direct measurement and therefore can only be measured through self-report. As such, the accuracy of the personality test is proportional both to the test taker's awareness and understanding of his own personality and to his truthful disclosure of that personality (HURLBUT, 1979). Moreover, the fluid and flowing nature of the interior phenomenon under

investigation, together with the factors just mentioned, result into a pattern of responses marked by high dispersion, and this makes the search for significant relationships considerably challenging. We saw that, in terms of magnitude, all the correlations between the psychosocial systems and the behavioral responses are low to moderate. Likewise, in our application of the multivariate regression analysis no intention of establishing quantitative relations was pursued, since the goodness of fit of the linear models was generally very poor. Also, the additional information imparted by the principal components in the logistic regressions is quite modest. Notwithstanding, we did obtain pretty meaningful components from the Factor Analysis and found that basically all statistically significant relationships concur with both key traits attached to our research question and hypotheses. In any event, those statistical relations cannot be attributed to chance, and it seems to me that we have no better explanation to them than the one provided by the ECLET conception itself.

Also, we don not need to go back all through the complexities ensuing from the fact the human interiority involves about two dozen of relatively independent developmental lines, out of which Graves's ECLET conception supplies but a partial view. Integral Theory already instructed us a propos to in which measure Graves's approach can be viewed as *representative* of a more general process of interior transformations. The point is that, respecting the evolution of the psychosocial value systems, Graves's conception laid down a foundation grounded on the six subsistence levels plus two being levels of existence which connects pretty harmoniously with the findings of the contemplative practices respecting the farther reaches of human potential.

Pertaining to our research conditions and procedures, the chief concern relates with the impossibility of assuring no post-experiment communication among acquainted participants as the experimental sections of the CPR experiment were succeeding. As pointed out, this situation led us to employ a design based on full information concerning the number of rounds and the nature of the treatments. We regard that this caution solved the problem of asymmetric information by setting up common knowledge in the beginning of each

experimental section. However, we cannot assure that some participants had not come to the experimental section bringing some preconceived idea about the optimal joint strategy. Still, we regard as safe enough to assume that this eventual occurrence have been just peripheral, in terms of frequency. Moreover, whether some few individuals' ideas a propos of the optimal strategy was formed outside the experimental control or only during the communication opportunity, as expected, this does not change the chief problem as to how to gain the adherence of others, and it is *then* that the gaps in the psychosocial levels of existence was posited to make difference.

A second issue relates with the possible influence of the experimenter. Even if we have strictly followed the experimental protocol and imparted the same instructions all through the experimental sections, the questions raised by the participants varied and some of them have indeed surprised us. We refer basically to the questions concerning the use of the sanctioning mechanism in the CPR experiment, which showed up in the Groups 21 and 22. Our possible interference is of course related with our inexperience in conducting experimental research. Nonetheless, we believe we gave the proper answer to those questions, and that the unusual conducts of those participants – who have decided to pay in order to punish of all the others – reflected authentic behavioral responses tied to their psychosocial profiles (Figure 7.30). In fact, we believe that this sort of behavior would have appeared more frequently had such sort of questioning being raised in other groups. Still, such utilization of the sanctioning mechanism in these two groups was considered inconsistent with the formal instructions, and this is just a flaw in our procedure.

We have, then, checked over the influence of these occasions by comparing the results from the entire data set which those excluding these groups. As pointed out, we found that these occurrences had a significant impact upon the net yield accrued from the investments in the Market 2, in the comparison between the communication and the sanctioning conditions, while not upon the gross profit (Tables 7.7 and 7.8). We thus decided for keeping both groups into the sample while excluding the abnormal defections and expenditure

during the sanctioning conditions, so that they would not impact the subsequent statistical analysis involving the psychosocial systems.

In summary, despite the clear room for methodological perfection, our results are certainly not dependable on the said procedural burden. It is then time to bring our study to a close by put forward the main implications of our results on policy addressing collective action problems.

8. CONCLUSIONS AND IMPLICATIONS

I opened this study by pointing to the importance of the implications of diverse frameworks into the inquiry on collective action problems. In particular, we attempted to show that, as a theoretical prediction, the ill-fated end resulting from Hardin's analysis of the *commons dilemma* is tied to a particular model of individual behavior, and that such a model may possibly be suitable, for analytical purposes, only within very restrictive conditions. We saw that these frameworks are in general useful in connection with the *original* assumptions and scope of neoclassical economics, although they by no means support a general theory of social relations. So, based on the condition that actors' interactions are mediated through markets, only, the model of "rational egoists" does apply, and may eventually provide a good enough approximation of the actual intentional states and rationales of the individuals in those circumstances. Yet, as discussed, in such a high level of abstraction it does not make much sense to see the actions of the economic agents as implying egoism, for the theoretical model simply does address alternatives of action.

Despite this, I also argued that the *symmetry thesis*, entrenched in the hypothetic-deductive model of science, has been either overtly or unnoticing supportive of a number of attempts to justify the factual validity of that particular model of "rational egoists". In this regard, some social version of the Darwinian

theory of evolution has been the most appealing candidate. To be sure, it is not by chance that the economic theory of perfect competition transplants to the market place the evolutionary principle of the *survival of the fittest*. Nor is it by a linguistic accident that the economic system co-enacted in concert with such principle is popularly refereed to as *savage* capitalism. Relating to our subject matter, we saw how Hardin drew into the Darwinian mindset to emerge with the odd proposition that "conscience is self eliminating", and to misread Nietzsche's account of guilt, bad conscience, and similar things. In fact, as discussed, by using the Darwinian evolutionary theory to interpret Nietzsche, Hardin just turned the author's position on its head, for Nietzsche indeed defined the human being as "the animal that is able to make promises". Moreover, Nietzsche saw this capacity as the summit of the *human* evolution thus far. To him, the capacity to keep promises and being trustworthy is a human *potency*, an evolutionary/developmental sign that the individual achieved sovereignty upon his or her ego-serving will and no longer needs to rely on the social straitjacket in order to be reliable and responsible (NIETZSCHE [1887], 1998, p. 47-50). First and foremost Nietzsche's attitude, just like that of Kant, was a rebellion against a philosophical tradition that subordinates the concrete individuals to powers and principles which are extrinsic to them.

I then attempted to show that, despite the many noteworthy developments in the human behavioral sciences and their cognitive foundations, enhance the tractability of different forms of cooperative behavior and interdependent choice – that the research agenda relating to collective action problems remains, in the main, confined within the instrumental conception of rationality. To be sure, we saw how Gintis (2003) forwarded the model of instrumental action and strategic interactions as the very foundation for the unification of the behavioral sciences. However, even though – provided the *assumptions* of consistent and self-interested preferences are abandoned – the conception of instrumental rationality and strategic relations does indeed support the development of significantly enhanced theoretical models, the exclusion of whatever other form of rational action builds in two extra and usually unstated

presuppositions, namely: that value judgment and moral issues are rationally undecidable (moral skepticism), and that communication capabilities can be taken for granted. As a result, despite the incorporation of an enriched set of intentional states in the so-called "second generation" models of the human actor, the most recent approaches to collective action problems only looks outside the individual – at the circumstances of action – mapping an ever increasing number of situational variables that describe an extraordinary complex socioecological system. These systems then must be carefully scrutinized in the attempt to *predict* the probabilities that alternative institutional arrangements may have in improving the collective outcome ensuing from strategic interactions, as well as to advance "design principles" aiming to foster the likelihood of having these arrangements emerge from the self-organization of the people involved. No doubt about the importance of such undertakings.

But now, as far as normative commitment is concerned, we get to the outermost frames of accepted epistemology, for the aforesaid approach tacitly assumes that human values and ideas of morality do not pertain to the scope of scientific justification, and thus cannot comprise a form of rational action. This is precisely why the approach which separates, on the one hand, instrumental rationality as the choice criterion, and preference functions with no prescribed content, on the other, is so appealing. The difficulty is that this makes the assumption of instrumental rationality unfalsifiable, as it will always be some preference function which can be made consistent which it in some circumstance of action. As pointed out by Gintis (2003), the significance of empirical tests using controlled experiments ensues precisely from this fact.

Concerning the present undertaking, we saw how the strategy of asking the participants about their expectations, in the Public Good experiment, allowed us to infer the motives under their actual choices whether or not to contribute for the common-pool. Besides, a major quality of our offer is that Graves's theory provides not only a way to get inside the individual preferences, but to predict how such preferences should evolve following the patterns of the psychosocial development of adult humans. Yet, all these additional information refer only to

the preferences of the individuals and are perfectly consistent with the instrumental model action, for I am surely not arguing that people are not attempting to realize their preferences when making choices.

The whole point is that whatever the preferences of the individuals might be they cannot impose restrictions to test the generability of the instrumental action model, for at the very moment of choice the preferences of the individuals will for ever remain private. It is only in the domain of public commitments, as during face-to-face communication in the Common-Pool Resource experiment, that some other form of rationality which comprises normative constraints may take an effect. It is *then* that the model of strategic action cannot explain the effect communication. For while providing agents with the information with which they determine maximizing strategies, no matter how egoistic or considerate they might be, this sort of language use does not exercise any constraint on the range of action alternatives open to the agents (HEATH, 2001). Incidentally, it is precisely this lack of restriction that justify both the explicit exclusion of linguistic communication from noncooperative game theory, and Hardin's vilification of it as essentially covering a "double binding" quality whenever the collective and the individual interests collide.

Notwithstanding, the adequacy of departing from such assumptions on the matters of rationality and theory of justification has been aptly challenged by the so-called *linguistic turn*, as it puts into question the axiomatic credit of the ordinary "reflection paradigm" and the corresponding *monological* model of justification. As discussed, the serious consideration of the justificatory practices expose that its inborn quality is *dialogical*, so that justification *to others* should be taken as the primary phenomenon (the key importance of empirical evidence ensues just from this fact, as it imposes the requirement of reproducibility, whereas saying nothing about the essential nature of the sensory-mediated phenomena, as Kant had showed us not too long ago). In particular, this position is developed in Habermas's theory of communicative action. Moreover, as pointed out, the sort of case that Habermas had in mind when establishing his key distinction between the strategic and the communicative kind of interaction goes

right in the core of the *commons dilemma*, that is, the case when agents adopt a rule to govern their interactions relating to a certain allocation problem. To do so, agents put themselves in need of justifying alternative rules *to others* – a process which enacts not only descriptions of "objective" facts (i.e. judgments grounded in same level perspectives), but also the involvement of value judgments and moral issues (e.g. Why should I care about the collective interest whatsoever?). This last involvement will certainly call for diverse and possibly even incommensurable perspectives, but perspectives which, nonetheless, present structural dependability, as posited and verified in this study, in connection with agents' level of psychosocial development. Accordingly, such diverse mindsets are not merely alternative possibilities that follow from the environmental determinations, with different probabilities. They do not impart mere different or equally applicable opinions and worldviews.

However, given the difficulties of challenging moral skepticism and incorporating others conceptions of rationality, scholars' exertions that explain normative commitment are driven further toward the exterior. Thus, the now enlarged set of intentional states – other than exclusive self-interest – presumably must not only cohere with the proximate circumstances of action but truly *ensue* from the evolutionary principle of survival of the fittest. As we saw, the result is that the creative inner activity of human beings is assumed to be beyond explanation and is treated as secondary relative to environmental conditions, no matter if biological or sociocultural. We saw that, when in the form of biological fundamentalism, such a standpoint ends up denouncing the first-person acquaintance of ones' subjective disposition, such as altruism, as mere "image and illusion" (see CRABTREE, 2006), as if the authenticity (not the sincerity) of what I allege to *experience* be bestowed to me by means of some intellectual interpretation of how it might possibly come into being. But even if such crude reductionism is passed up, the subtle form of it still grants that a valid explanation for human interior experiences and the corresponding actions ought to be found in the surroundings. As we saw, this kind of objectivistic approach has brought about sophisticated behavioral and evolutionary theories to explain

the so-called "emergence" of cooperation, altruism, and similar mindsets. Once again, we have acknowledged the significance of these endeavors, whereas pointing out that they remain limited by an unquestioned identification of rationality, in general, with its instrumental form – a viewpoint which is inseparable from the often unstated assumption that normative commitments must either be just a shadow of strategic interactions (reciprocity or reputation) or have some kind of nonrational source, such as conditioning or habit.

But once we acknowledge that rationality is more than cleverness, and that the content of our minds are more than imperfect reflections of a preset reality, then we draw out our conceptual framework so that it graciously opens itself toward a more balanced inquiry on the origins and evolution of the contents of individuals' mindsets and worldviews. In this vein, the suggested linkage relating the so-called "memetic" transmission with the stage developmental theory is provided by both Habermas and Wilber's Integral Theory. On the one hand, Habermas sets forth an isomorphic correspondence between the phylogenesis of the communicative rationality – built in the work of Emile Durkheim and Herbert Mead – and the ontogenesis of the sociocognitive and moral reasoning presented by Lawrence Kohlberg. On the other hand, Ken Wilber unites this correspondence with a number of other developmental models – including the ECLET conception – and expands on the vertical (developmental) dimension involved in the likelihood of getting to mutual understanding. As Integral Theory goes farther to bring about the epistemological integration between all these developments and the offerings of the spiritual wisdom and perennial philosophy, we needed merely to envision an empirical test to look into, first, whether participants' ability to get to whatever agreement and keep their promises respecting their collective action situations has or has not something to do with their stages of psychosocial development, and, second, in which measure our behavioral observations concur with the theoretical predictions ensuing from the ECLET conception, as we vary the institutional settings and experimental situations.

We have already discussed the strengths and weakness of our undertaking. On this ground, I just claimed that the ECLET conception added a significant contribution to our understanding about what exactly is going on within individuals uneven responses to the challenge imposed upon them by the experimental situations, namely: participants' psychosocial levels of existence. Specifically, as a primary evident pattern, the ECLET model showed us that, on the one hand, the express-self systems either cooperate less or just raze collective agreement to the ground, and, on the other hand, that the sacrifice-self systems not only cooperate more but – as far as the FS system is concerned – are in the main responsible for sustaining the collective outcome in the face of cheating and the greed of others. Concurring with the underlining trait ensuing from the process of decentering of the self along the developmental path, our observations supported the proposition that the capacity to keep promises and stand within collective action dilemmas, with an attitude bent toward the collective good is a trait associated with the higher levels of psychosocial development. Accordingly, even if the experimental situations certainly do not intend to stand for complex concrete action arenas, these trends suggest, nonetheless, that the sponsorship of favorable conditions supporting the health associated with human interior growth does present a qualified opportunity to enhance the outcomes of human interactions, in general, and those involving collective action dilemmas, in particular. Moreover, the most basic implication of the ECLET conception is that no incentive structure should be expected to work straightforwardly, regardless of the stage of psychosocial development of either the individual or the community of individuals. Instead, in order to produce the expected responses, any incentive arrangement ought to be congruent with those levels of development, honoring their healthy expression and fostering upward movement. Again, we have verified this trait experimentally, as the varied behavioral responses to both the different institutional settings in the Common-Pool Resource experiment and the different contribution levels in the Public Good experiment produced a systematic pattern which is dependent on the psychosocial levels of existence.

So, as posited, the *prime directive* of a genuine integral politics, which intends not to impose preset "solutions" but empower the local people, so that they may more likely find and sustain their own answer for their social dilemmas, would neither assume any epitome conception of rationality – and particularly not that of the "rational egoists" – nor try to get everybody to a particular high level of consciousness. As Wilber puts it, the prime directive should be "to *ensue the health of the entire spiral of development* at all of its levels and waves" (WILBER, 2001, p. 88, emphasis in the original). The reason being just that every infant born in every society still has to start at the level of sensorimotor instincts and perceptions – such as AN in Graves's schema – and then grow and evolve all through the stages of consciousness awakening, intermingling the numerous aspects or lines of the subjective life with the colors of his or her particular life conditions, to eventually transcend-and-include all temporal-spatial boundaries and get into a cosmic level of awareness – which is more often than not accompanied by a detached compassion for all sentient beings.

We have already discussed enough – as it is a central attribute of Integral Theory – *a propos* of the theoretical unassailability which relates the findings of the stage developmental theory, particularly concerning value systems and morals, with those coming from the spiritual traditions and perennial philosophy. Yet, regardless whether one thinks that such ties are fitting or not, the results of the prescribed science alone, including this one, are sufficient to establish a *positive* relationship linking the level of psychosocial development with the capacity to be trustworthy and altruistic. Then, our claim is that this relationship provides a foundation for the most common recommendations as to how to promote cooperation.

As just discussed, no straightforward incentive arrangement shall disentangle the payoff structure that characterizes the collective action dilemma, for the individuals' payoffs are weighted according to different subjective perceptions of the situation, including diverse reactions to the kind and form of presentation of the incentive mechanism itself. To be sure, the fundamental

proposition underlining this study is just that the fundamental tension between the individual and the collective interests, in the root of the collective action dilemmas, is consistently wakened as the individuals move upward along the path of consciousness awakening. As we saw, even if the FS system is characterized as presenting a generalized short-term rewarding perspective, it is not so for the sake of putting self-interest ahead of the collective good. As a result, as far as the weight of the FS willingness to cooperate is enough in the experimental group, the opportunity to share this subjective evaluation basically "straightened out" the payoff "puzzle" in the appropriation problem, without the need of coercion.

Likewise, the importance of the "shadow of the future", to use Axelrod's (1984) expression, is not determined by any technical variable, such as the interest rate. Much on the contrary, the interest rate is perhaps the economic variable most clearly affected by a whole complex mess of beliefs and intentional states, involving the most diverse perceptions and evaluations of the future relative to the present. Incidentally, and laughable as tragic, the contemporary archetype of sagacity underpinning the dictum "time is money" – which exposes the most profound angst and futile aspirations of our times – is what has been demolishing our hope of sustaining human civilization in this planet for the long run. To be sure, in terms of a temporal perspective, research on endogenous growth, for instance, commonly regards a time series of two hundred years as a quite long period of time. On the other hand, the Buddha School measures time in *kalpa*s. In Hinduism, one kalpa is equal to 4,320 million years. This corresponds to a "day of Brahma" or one thousand "Yugas", measuring the duration of the world. A "month of Brahma" is supposed to contain thirty such kalpas, or 129.6 billion years. According to the Mahabharata, 12 months of Brahma constitute his year, and 100 such years the life cycle of the universe. In Buddhism, there are four different lengths of kalpas. A *regular* kalpa is approximately 16 million years long, and a *small* kalpa is 1,000 regular kalpas, or 16 billion years. Further, a *medium* kalpa is 320 billion years, the equivalent of 20 small kalpas. A *great* kalpa is 4 medium kalpas, or 1.28 trillion years. Of

course, the issue is not about units of measurement, but instead, about the time span connected with different mindsets. While economics are concerned with the convergence the nations' Gross Domestic Product, as a measure of material well-being, the disciples of the Buddha School are concerned with the Enlightenment and salvation of their main consciousness, as called. Well, unless one simply disallows such knowledge traditions, out of the contemporary epistemological fundamentalism, it seems that no institutional device shall be able to extend the "shadow of future" far enough.

As well, the improvement of the recognition abilities, as a tool to foster conditional cooperation by bookkeeping the outcome of past interactions, would be useless if all that was necessary was to recognize that "everybody is a 'rational egoist', and that everybody knows that everybody knows that". So, we are back again to the problem of taking seriously the understanding of the origins of the content of agents' beliefs and intentional states. And this makes Hardin's appeal to education seem quite reified, for once he had formerly disqualified whatever plea to "conscience", he concludes his paper on the commons dilemma by suggesting that it would be the role of education "to *reveal to all* the necessity of abandoning the freedom to breed" (HARDIN, 1993, p. 19, my italics).

But the whole problem is that there is nothing to be revealed, for everything can only be *realized*. Understanding is an active occasion. Even a conformist learning depends on agents' ability to take the sense of whatever is intended to be taught in the first place. The readiness to do so will vary, as we have been discussing, precisely because psychosocial development presents its own law of transformation and cannot simply be internalized through the professed education. Still, to keep ones' way of thinking confined within the sphere of public acceptance, the individual has to continuously disarm the interior pressure to awake further, by means of all sorts of distractions, and thus wrap the intuitive suspicion that it might by something profoundly wrong with the contemporary and educated acceptation of human ignorance of what life is all about. And this is so, furthermore, just because the fundamental difficulty is precisely that no one is able to pursue the path of consciousness awakening

alone, for *communal checking* is but the quintessence of every single piece of the so-called knowledge. So, the recommendation to "educate" people to be more cooperative when dealing with their social dilemmas incurs pretty much the same difficulty involved in communicatively justifying to others the desirability of following a rule to govern their interactions. As we saw, communication opportunities are neither a panacea nor a waste of time. Their effectiveness depends on a quite concrete encounter of perspectives in the event-horizon of consciousness, that is, in the phenomenological **we-space** of mutual understanding. Clearly, the same issue applies when devising any institutional arrangements or incentive structure, as we have attempted to show.

In closing, I would like reiterate the tremendous importance of bringing about the epistemological integration put forward by the Integral Theory attempted in this re-view of the commons dilemma. Honoring local beliefs and worldviews is a basic requirement for carrying out successful participatory initiatives aiming to empower local people to solve their social dilemmas. To be able to connect authentically with local worldviews, and to foster local ownership of development work, development practices must involve the subjective and intersubjective dimensions of change. The ECLET conception, applied here, is but one possible way of working with such dimensions. Other models may possibly be better suited to address different sorts of questions. Yet, even though many Northern development practitioners have more secular leaning worldviews, to be effective in the subjective and intersubjective domains, the inclusion of teleologies and belief systems is crucial (HOCHACHKA, 2006). Of course, one does not need to share the same spirituality or religious beliefs of local people to eventually succeed in collaborative inquiry and to assist them in the process of creating their own solutions to their collective action problems. Yet, many communities are guided by a spiritual or religious understanding of reality, and, when working in such social systems, one cannot hold secularism as a definitive cosmology. Instead, as Baum puts it (HARPER, 2000, p. 81), the practitioner should be open to the possibility that the religious symbols that define the identity of a human community have a transcendent referent; and that

practitioner's own secular presuppositions is a Western cultural product rather than a universal truth.

Likewise, as Kapur (2000) says *a propos* of the methodological emphasis on 'people centeredness', I too am left to wonder how this rehabilitated emphasis on altruism and cooperation will manifest without a more explicit focus on spiritual principles. For are not the challenges of tapping into creativity, distributing benefits justly, and providing equal access to opportunities, in the final analysis asking what motivates people to act, and, more important, to act compassionately? And does that question not require us to ask what is at the center of ourselves?

My hope is that this contribution has credibly suggested in what precise sense such an explicit focus on spiritual principles may contribute to the inquiry on collective action dilemmas.

REFERENCES

ABBINK, K.; BOLTON, G.E.; SADRIEH, A.; TANG, F.F. **Adaptive learning versus punishment in ultimatum bargaining**. Bonn, Germany: Rheinishe Friedrich-Wilhelms-Universität Bonn, 1996. (Discussion Paper, B-381).

AGRAWAL, A. **World Development**, v. 29, p. 1649-1672, 2001.

AJZEN, I. The theory of planned behavior. **Organizational Behavior and Human Decision Processes**, v. 50, p. 179-211, 1991.

AJZEN, I.; MADDEN, T.J. Prediction of goal-directed behaviour: attitudes, intentions, and perceived behavioral control. **Journal of Experimental Social Psychology**, v. 22, p. 453-474, 1985.

AKERLOF, G. A theory of social custom, of which unemployment may be one consequence. **Quarterly Journal of Economics**, v. 94, p. 749-775, 1980.

AKHTAR, S. The syndrome of identity diffusion. **American Journal Psychiatry**, v. 141, p. 1381-1385, 1984.

ALBIN, P.S.; FOLEY, D.K. The co-evolution of cooperation and complexity in a multi-player local-interaction prisoners dilemma. **Complexity**, v. 6, n. 3, p. 54-63, Jan./Feb. 2001.

ALCHIAN, A.A. Uncertainty, evolution, and economic theory. **Journal of Political Economy**, v. 58, n. 3, p. 211-221, 1950.

ALEXANDER, C.; LANGER, E. **Higher stages of human development**. Oxford: Oxford University Press, 1990.

ANDREONI, J. Cooperation in public-goods experiments: kindness or confusion? **The American Economic Review**, v. 85, n. 4, p. 891, Sep. 1995.

ARIAS, E.; EDEN, H.; FISCHER, G.; GORMAN, A.; SCHARFF, E. Transcending the individual human mind – creating shared understanding through collaborative design. **Transactions on Computer-Human Interaction**, v. 7, n. 1, p. 84-113, 2000.

ARNOLD, J.E.M.; CAMPBELL, J.G. Collective management of hill forests in Nepal: the community forestry development project. In: CONFERENCE ON COMMON PROPERTY RESOURCE MANAGEMENT, Washington, 1986. **Proceedings…** Washington: National Academy Press, 1986. p. 425-454.

ARNOLD, J.E.M.; STEWART, W.C. **Common property management in India**. Oxford: Oxford Forest Institute, Oxford University, 1991. (Tropical Forestry Papers, 24).

ARTHUR, W.B. The end of certainty in economics. In: AERTS, D.; BROEKAERT, J.; MATHIJS, E. (Eds.). **Einstein meets Magritte**. Holland: Kluwer Academic Publishers, 1999.

ARTHUR, W.B.; DURLAUF, S.N.; LANE D. Introduction. In: ____. **The economy as an evolving complex system II**. New York: Westview Press, 1997. (Proceedings volume in the Santa Fe Institute in the Sciences of Complexity).

AXELROD, R. An evolutionary approach to norms. **American Political Science Review**, v. 80, n. 4, p. 1095-1111, Dec. 1996.

AXELROD, R. Advancing the art of simulation. In: CONTE, R.; HEGSELMANN, R.; TERNA, P. (Eds.). **Simulation social phenomena**. Berlin: Springer, 1997a. p. 21-47.

AXELROD, R. Promoting norms. In: AXELROD, R. **The complexity of cooperation**: agent-based models of competition and collaboration. Princeton, New Jersey: Princeton University Press, 1997b. p. 40-68.

AXELROD, R.; HAMILTON, W.D. The evolution of cooperation. **Science, v.** 211, n. 4489, p. 1390-1396, 1984.

BACON, F. **Advancement of learning**. London: Printed for William Washington, 1629. 335 p.

BALAND, J.M.; PLATTEAU, J.P. **Halting degradation of natural resources**: is there a role for rural communities? Oxford: Clarendon Press, 1996.

BARBOUR, I. **Religion and science** – historical and contemporary issues. London: SCM Press, 1998. (Gifford Lectures Series).

BARDHAN, P. **Water community**: an empirical analysis of cooperation on irrigation is South India. Berkley: University of California, Department of Economics, 1999. (Working Paper).

BARON, R.A. Personality and organizational convict: effects of the type A behavior pattern and self-monitoring. **Organizational Behavior and Human Decision Processes**, v. 44, p. 281-296, 1989.

BARRY, B.; HARDIN, R. **Rational man and irrational society?**: an introduction and sourcebook. Beverly Hills, CA: Sage, 1982.

BATESON, G. **Steps to an ecology of mind**. New York: Ballantine, 1972.

BATESON, G. **Mind and nature**: a necessary unity. New York. Dutton, 1979.

BECHARA, A.; DAMASIO, H.; DAMASIO, A. Emotions, decision making, and the orbitofrontal cortex. **Cerebral Cortex,** v. 10, n. 295-307, 2000.

BECHARA, A.; DAMASIO, H.; DAMASIO, A.; ANDERSON, S. Insensitivity to future consequences following damage to human prefrontal cortex. **Cognition**, v. 50, p. 7-15, 1994.

BECHARA, A.; DAMASIO, H.; TRANEL, D.; DAMASIO, A. Deciding advantageously before knowing the advantageous strategy. **Science**, v. 275, p. 1293-1295, 1997.

BECK, E.D. **Stages of social development**: the cultural dynamic that spark violence, spread prosperity, and shape globalization. Disponível em: <http://www.humanemergence.org/essays/stages_of_social_development.htm>. Acesso em: Oct. 2000a.

BECK, E.D. **Sustainable cultures, sustainable planet**: a values system perspective on consctructive dialogue and cooperative action. Disponível em: <http://www.humanemergence.org/essays/substantial.htm>. Acesso em: 2000b.

BECK, E.D.; COWAN, C.C. **Spiral dynamics**: mastering values, leadership and change – exploring the new science of memetics. London: Blackwell Publishers, 1996.

BECK, E.D.; LINSCOTT, G. **The crucible**: forging South Africa's future. Johannesburg, South Africa: The New Paradigm Press, 1991.

BECKER, G. **Treatise on the family**. Cambridge, MA: Harvard University Press, 1991.

BECKOFF, M. Science, religion, cooperation and social morality. **BioScience**, v. 51, n. 3, p. 171, 2001. (Editorial).

BENABOU, R. Equity and efficiency in human capital investment: the local connection. **Review of Economic Studies**, v. 63, p. 237-264, 1996a.

BENABOU, R. Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance. **American Economic Review**, v. 86, p. 584-609, 1996b.

BERGSON, H. **Creative evolution**. New York: Dover Publications, 1998.

BERGSON, H. **The creative mind**: an introduction to metaphysics. New York: Kensington Publishing, 2002.

BLAUG, M. **A metodologia de economia, ou, como os economistas explicam**. 2.ed.rev. São Paulo: Editora da Universidade de São Paulo, 1993.

BLOOM, P. **How children learn the meaning of words**. Cambrigde, MA: MIT Press, 2000.

BLOOR, D. **Wittgenstein, rules and institutions**. London and New York: Routledge, 2002.

BLUME, L.; DURLAUF, S. **Equilibrium concepts for social interactions models**. 1998. [http://www.santafe.edu/~leb/durlauf09.pdf].

BOHEM, C. Emergency decisions, cultural-selection mechanics, and group selection. **Curr. Anthro.**, v. 37, p. 763-793, 1996.

BOHM, D. **Wholeness and the implicate order**. London: Routledge, 1980.

BOHM, D. **On dialogue**. London and New York: Routledge, 2004.

BOHM, D.; HILEY, B.J. **The undivided universe**. London: Routledge, 1993.

BOLAND, L.A. On the futility of criticizing the neoclassical maximization hypothesis. **American Economic Review**, v. 71, p. 1031-1036, 1981.

BOLTON, G.E. A comparative model of bargaining: theory and evidence. **American Economic Review**, v. 81, p. 1096-1136, 1991.

BOOKSTABER, R.; LANGSAM, J. Predictable behavior: commest. **The American Economic Review**, v. 75, p. 571-575, 1985.

BOONE, C.; BRABANDER, B.; WITTELOOSTUIJN, A. The impact of personality on behavior in five prisoner's dilemma games. **Journal of Economic Psychology**, v. 20, p. 343-377, 1999.

BORNSTEIN, G.; RAPOPORT, A. Intergroup competition for the provision of step-level public goods: effects of preplay communication. **European Journal of Social Psychology**, v. 18, p. 125-142, 1988.

BORNSTEIN, G.; RAPOPORT, A.; KERPEL, L.; KATZ, T. Within – and between – group communication in intergroup competition for public goods. **Journal of Experimental Social Psychology**, v. 25, p. 422-436, 1989.

BOSMAN, R.; VAN WINDEN, F. Emotional Hazard in a power-to-take experiment. **Economic Journal**, v. 112, p. 147-169, 2002.

BOULDING, K.E. Towards a pure theory of threat systems. **American Economic Review**, v. 53, p. 424-434, 1963.

BOWLES, S. Endogenous preferences: the cultural consequences of markets and other economic institutions. **Journal of Economic Literature**, v. 36, p. 75-111, Mar. 1998.

BOWLES, S.; GINTIS, H. The moral economy of community: structured populations and the evolution of prosocial norms. **Evolution & Human Behavior**, v. 19, n. 1, p. 3-25, Jan. 1998.

BOWLES, S.; GINTIS, H. **Social capital and community governance**. Amherst, Massachusetts: University of Massachusetts and Santa Fe Institute, 2001a.

BOWLES, S.; GINTIS, H. **The evolution of strong reciprocity**. University of Massachusetts at Amherst, 2001b. (Discussion Paper).

BOWLES, S.; GINTIS, H. **Prosocial emotions**. Santa Fe Institute, 2003. (Working Paper, #02-07-028).

BOWLES, S.; BOYD, R.; FEHR, E.; GINTIS, H. **Homo reciprocans**: a research initiative on the origins, dimensions, and policy implications of reciprocal fairness research initiative on reciprocal fairness. 1997.

BOWLES, S.; CARPTENTER, J.; GINTIS, H. **Mutual monitoring in teams**: the importance of shame and punishment. University of Massachusetts, 2001.

BOYD, R.; RICHERSON, P.J. **Culture and evolutionary process**. Chicago: University of Chicago Press, 1985.

BOYD, R.; RICHERSON, P.J. The evolution of indirect reciprocity. **Social Network**, v. 11, p. 213-236, 1989.

BOYD, R.; RICHERSON, P.J. Group selection among alternative evolutionary stable strategies. **Journal of Theoretical Biology**, p. 145, n. 3, p. 331-342, 1990.

BOYD, R.; RICHERSON, P.J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. **Ethology and Sociobiology**, v. 13, p. 171-195, 1992.

BOYD, R.; RICHERSON, P.J. The evolution of reciprocity in sizable groups. **Journal of Theoretical Biology**, v. 132, p. 337-356, 1998.

BRANDOM, R. **Making it explicit**. Cambridge, MA: Harvard University Press, 1994.

BRAVER, S.L.; WILSON, L.A. **A laboratory study for social contracts as a solution for public good problems**: surviving on the lifeboat. San Diego: Western Social Science Association, 1984.

BRAVER, S.L.; WILSON, L.A. Choices in social dilemmas: effect of communication within subgroups. **Journal of Conflict Resolution**, v. 20, n. 1, p. 51-62, 1986.

BRENNAN, G.; BUCHANAN, J. **The reason of rules**. Cambridge: Cambridge University Press, 1985.

BRESNEN, M.; MARSHALL, N. The engineering or evolution of co-operation? A tale of two partnering projects. **International Journal of Project Management**, v. 20, p. 497-505, 2002.

BRENNER, T.; VRIEND, N.J. On the behavior of proposers in ultimatum games. **Journal of Economic Behaviour and Organization**, In Press, Corrected Proof, 2006.

BROSIG, J. Identifying cooperative behavior: some experimental results in a prisoner's dilemma game. **Journal of Economic Behavior & Organization**, v. 47, p. 275-290, 2002.

BUENO, N.P. Lógica da ação coletiva, instituições e crescimento econômico: uma resenha temática sobre Nova Economia Institucional. **Economia Brasileira**, v. 5, n. 2, p. 343-402, 2004.

CALDWELL, M.D. Communication and sex effect in a five-person prisoner's dilemma game. **Journal of Personality and Social Psychology,** v. 33, n. 3, p. 273-280, 1976.

CAMERER, C.F. **Behavioral game theory**: experiments in strategic interaction. Princeton: Princeton University Press, 2003.

CAMERON, L.A. Raising the stakes in the ultimatum game: experimental evidence from Indonesia. **Economic Inquiry,** v. 37, n. 1, p. 47-59, 1999.

CÁRDENAS, J.C. Thinking globally and getting others to act locally: polycentricity and the conservation of biodiversity. In: WORKSHOP IN POLITICAL THEORY AND POLICY ANALYSIS, 2000, Bloomington. **Paper presented...** Bloomington: Indiana University, 2000.

CASARI, M.; PLOTT, C.R. Decentralized management of common property resources: experiments with a centuries-old institution. **Journal of Economic Behavior & Organization**, v. 51, p. 217-247, 2003.

CASON, T.N.; SAIJO, T.; YAMATO, T.; YOKOTANI, K. Non-excludable public good experiments. **Games and Economic Behavior**, v. 49, p. 81-102, 2004.

CAVALLI-SFORZA, L.L.; FELDMAN, M. **Cultural transmission and evolution**. Princeton, NJ: Princeton University Press, 1981.

CHARDIN, P.T. **The phenomenon of man**. Harper Perennial, 1976.

CHAUDHURI, A.; SOPHER, B.; STRAND, P. Cooperation in social dilemmas, trust and reciprocity. **Journal of Economic Psychology**, v. 23, p. 231-249, 2002.

CILLIERS, P. Boundaries, hierarchies and networks in complex systems. **International Journal of Innovation Management**, v. 5, n. 2, p. 135-147, 2001.

CILLIERS, P. Introduction to Simon's. **The Architecture of Complexity**, v. 7, n. 3-4, p. 138-154, 2005.

CLARK, A.; KARMILOFF-SMITH, A. The cognizer's innards: a psychological and philosophical perspective on the development of thought. **Mind and Language**, v. 8, n. 4, p. 487-519, 1991.

CLUTTON-BROCK, T.H.; PARKER, G.A. Punishment in animal societies. **Nature**, v. 373, n. 6511, p. 209-216, 1995.

COASE, R. The problem of social cost. **Journal of Law and Economics**, v. 3, p. 1-44, 1960.

COASE, R. The nature of the firm. In: WILLIAMSON, O.; WINTER, S. (Ed.). **The nature of the firm**: origins, evolution and development. New York: Oxford University Press, 1991. v. 4.

COGNITION, A.W.B. **The black box of economics**. Northampton, Mass.: Edward Elgar Publishing, 2000.

COHEN, J.D. Reward and decision. **Neuron**, v. 36, p. 193-198, Oct. 2002.

COLE, H.; MAILATH, G.; POSTLEWAITE, A. Social norms, savings behavior, and growth. **Journal of Political Economy**, v. 100, p. 1092-1125, 1992.

COMMONS, M.L. A comparison and synthesis of Kohlberg's cognitive-developmental and Gewirtz's learning-developmental attachment theories. In: GEWIRTZ, J.L.; KURTINES, W.M. (Eds.). **Intersections with attachment**. Hillsdale, NJ: Erlbaum, 1991. p. 257-291.

COMMONS, M.L. **Hierarchical complexity of contingencies of reinforcement and of stimulus discriminations**. Washington, D.C., 2000. (Presented at the Society for Quantitative Analysis of Behavior).

COMMONS, M.L.; RICHARDS, F.A. A general model of stage theory. In: COMMONS, M.L.; RICHARDS, F.A.; ARMON, C. (Eds.). **Beyond formal operations**: late adolescent and adult cognitive development. New York: Praeger, 1984. v. 1, p. 120-140.

COMMONS, M.L.; RICHARDS, F.A.; ARMON, C. **Beyond formal operations**: late adolescent and adult cognitive development. New York: Praeger, 1984.

COMMONS, M.L.; MILLER, P.M.; GOODHEART, E.A.; DANAHER-GILPIN, D. **Hierarchical complexity scoring system (HCSS)**: how to score anything. Cambridge: Unpublished Scoring Manual Available from Dare Institute, 2005. <http://www.dareassociation.org/>.

CONCEIÇÃO, O.A.C. O conceito de instituição nas modernas abordagens institucionalistas. **R. Econ. Contemp.**, Rio de Janeiro, v. 6, n. 2, p. 119-146, 2002.

COOK, H.; CHI, C. Cooperative behavior and locus of control among American and Chinese-American boys. **Journal of Psychology**, v. 118, p. 169-177, 1984.

COOK-GREUTER, S.R. Maps for living: ego-development stages from symbiosis to conscious universal embeddedness. In: COMMONS, M.L.; ARMON, C.; KOHLBERG, L.; RICHARDS, F.A.; GROTZER, T.A.; SINNOTT, J.D. (Eds.). **Adult development**: models and methods in the study of adolescent and adult thought. New York: Praeger, 1990. v. 2, p. 79-104.

CORNES, R.; SANDLER, T. On commons and tragedies. **American Economic Review**, v. 73, n. 4, Sep. 1983.

CORNES, R.; MASON, A.; SANDLER, T. The commons and the optimal number of firms. **Quarterly Journal of Economics**, v. 100, p. 641-646, 1986.

CORNES, R.; MASON, A.; SANDLER, T. **The theory of externalities, public goods and club goods**. 2.ed. Cambridge: Cambridge University Press, 1996.

COSMIDES, L.; TOOBY, J. Evolutionary psychology and the generation of culture: case study – a computational theory of social exchange. **Ethology and Sociobiology**, v. 10, n. 1-3, p. 51-97, 1989.

COSMIDES, L.; TOOBY, J. Cognitive adaptations for social exchange. In: BARKOW, J.; COSMIDES, L.; TOOBY, J. (Eds.). **The adapted mind**. New York: Oxford University Press, 1992.

COSMIDES, L.; TOOBY, J. Better than rational: evolutionary psychology and the invisible hand. **American Economic Review**, v. 84, n. 2, p. 327-332, 1994.

CRABTREE, V. **Altruism**: is it real? Disponível em: <http://www.humantruth.info/altruism.html>. Acesso em: 2006.

CRANSTON, M. **The noble savage**. Chicago: University of Chicago Press, 1991.

CRANSTON, M. **The solitary self**. Chicago: University of Chicago Press, 1997.

CUMMINS, R.; POIRIER, P.; ROTH, M. Epistemological strata and the rules of right reason. **Synthese**, v. 141, p. 287-331, 2004.

CURRY, O. Is patience a virtue? In testing theories of the evolution of collective action: reciprocity, punishment, guilt and patience. In: WORKSHOP IN POLITICAL THEORY AND POLICY ANALYSIS, 2005, Bloomington. **Colloquium Presentations...** Bloomington: Indiana University, 2005.

DAMASIO, A.R. **Descartes' error**: emotion, reason, and the human brain. New York: Avon Books, 1994.

DAMASIO, A.R. **The feeling of what happens**: body and emotion in the making of consciousness. New York: Harcourt Brace, 1999.

DAMASIO, A.R. **Em busca de Espinoza**: prazer e dor na ciência dos sentimentos. São Paulo: Companhia da Letras, 2004.

DAWES, R.M.; McTAVISH, M.J.; SHAKLEE, H. Behavior, communication and assumptions about other people's behavior in a commons dilemma situation. **Journal of Personality and Social Psychology**, v. 35, n. 1, p. 1-11, 1977.

DAWES, R.M.; ORBELL, J.M.; VAN DE KRAGT, A. **Normative constraint and incentive compatible design**. University of Oregon, Department of Psychology, Typescript, 1984.

DAWKINS, R. **The selfish gene**. New York: Oxford University Press, 1976.

DAWKINS, R. **The god delusion**. Bantam Books, 2006.

DEMSETZ, H. Toward a theory of property rights. **American Economic Review**, v. 57, p. 347-359, 1967

DENZAU, A.T.; NORTH, D.C. **Shared mental models**: ideologies and institutions. 1994.

DENZAU, A.T.; NORTH, D.C. Shared mental models: ideologies and institutions. In: LUPIA, A.; McCUBBINS, M.D.; POPKIN, S.L. (Eds.). **Elements of reason** – cognition, choice, and the bounds of rationality. 3.ed. Readings: Cambridge University Press, 2000. p. 23-46.

DIRVER, C.F. On the origin of predictable behavior: a comment. **American Economic Review**, v. 82, n. 1, p. 36-38, 1992.

DUBRAY, C.A. **Teleology**. Disponível em: <http://www.newadvent.org/cathen/14474a.htm>. Acesso em: 2007.

DUGATKIN, L.A. **Cooperation among animals**: an evolutionary perspective. New York: Oxford University Press, 1997.

ECKEL, C.C.; WILSON, R.K. **In trust and reciprocity**: interdisciplinary lessons from experimental research. In: OSTROM, E.; WALKER, J. (Eds.). New York: Russell Sage Foundation, 2003. p. 245-274.

EDNEY, J.J.; HARPER, C.S. The commons dilemma: a review of contributions from psychology. **Environmental Management**, v. 2, n. 6, p. 494-507, 1978.

EIGEN, M.E.; WINKLER, R. **Laws of the game**: how the principles of nature govern chance. New York: Harper & Row, 1981.

EKMAN, P. **Emotions in human face**: guide lines for research and an integration of findings. New York. Pergamon Press, 1972. (Pergamon General Psychology Series, 11).

EKMAN, P. **Emotions in human face**. 2.ed. Cambridge: Cambridge University Press, 1982. (Studies in Emotion and Social Interaction).

ENCYCLOPEDIA OF WORLD PROBLEMS AND HUMAN POTENTIAL. **Patterns of alternation**: cycles of dissonance and resonance. Global strategies Project – notes and commentaries. <http://www.uia.be/strategies/stratcom_bodies.php?kap=66>.

ERIKSON, E.H. **Childhood and society**. New York: Norton, 1950.

FEHR, E.; GÄCHTER, S. How effective are trust – and reciprocity-based incentives? In: BEN-NER, A.; PUTTERMAN, L. (Eds.). **Economics, values and organizations**. Cambridge: Cambridge University Press, 1988. p. 337-363.

FEHR, E.; GÄCHTER, S. Reciprocity and economics: the economic implications of *Homo reciprocans*. **Eur. Econ. Rev.**, v. 42, p. 845-859, 1998.

FEHR, E.; GÄCHTER, S. Cooperation and punishment in public goods experiments. **The American Economic Review**, v. 90, p. 980-994, 2000.

FEHR, E.; FISCHBACHER, U.; GÄCHTER, S. Strong reciprocity, human cooperation, and the enforcement of social norms. **Human Nature**, v. 13, p. 1-25, 2002.

FESSLER, D.M.T.; HALEY, K.J. The strategy of affect: emotions in human cooperation. In: HAMMERSTEIN, P. (Ed.). **Genetic and cultural evolution of cooperation**. Cambridge, Massachusetts; London, England: The MIT Press in cooperation with Dahlem University Press, 2003.

FEYERABEND, P.K. **Against method, outline of an anarchist theory of knowledge**. London: NLB, 1975.

FISCHBACHER, U.; GÄCHTER, S.; FEHR, E. Are people conditionally cooperative? Evidence from a public goods experiment. **Economics Letters**, v. 71, p. 397-404, 2001.

FISHER, D.; TORBERT, W.R. **Personal and organizational transformation**: the true challenge of continual quality improvement. London: McGraw-Hill, 1995.

FLETCHER, J. **Situation ethics**. Westminster, Philadelphia, 1966.

FOSTER, D.P.; YOUNG, H.P. **On the impossibility of predicting the behavior of rational agents**. Disponível em: <http://www.santafe.edu/research/publications/wpabstract/200108039>. Acesso em: 2001.

FRANKL, V.E. **Man's search for meaning**. New York: Pocket Books, 1984.

FREIRE, P. **Pedagogy of Freedom**: ethics, democracy, and civic courage. Rowman & Littlefield, 1998.

FREY, B.S.; EICHENBERGER, R. Should social scientis care about choice anomalies? **Rationality and Society**, v. 1, n. 1, p. 101-122, 1989.

FRIEDMAN, M. **The methodology of positive economics**: essays in positive economics. 1953.

FRIEDMAN, M.; ROSENMAN, R.H. **Type a behavior and your heart**. New York: Knopf, 1974.

FROHLICH, N.; OPPENHEIMER, J.A. I get by with a little help from my friends. **World Politics**, v. 23, p. 104-120, 1970.

FROHLICH, N.; OPPENHEIMER, J.A. **Choosing**: a cognitive model of economic and political choice. Winnipeg: University of Manitoba, Faculty of Management, 2001. (Working Paper).

FROMM, E. **Escape from freedom**. New York: Farrar, Straus & Giroux, 1941.

FROMM, E. **Análise do homem**. Rio de Janeiro: Zahar, 1966.

FROMM, E. **To have or to be**. London: Abacus, 1979.

GÄCHTER, S.; FEHR, E. Collective action as a social exchange. **Journal of Economic Behavior & Organization**, v. 39, p. 341-369, 1999.

GARDNER, H. **Frames of mind**: the theory of multiple intelligences. New York: Basic, 1983.

GARDNER, H. **Multiple intelligences**: the theory in practice. New York: Basic, 1993.

GARRISON, R.W. Predictable behavior: commest. **The American Economic Review**, v. 75, n. 3, p. 571-578, 1985.

GILLIS, M. Indonesia: public policies, resource management, and the tropical forestry. In: REPETTO, R.; GILLIS, M. (Eds.). **Public policies and the misuse of forest resources**. New York: Cambridge University Press, 1988. p. 43-113.

GINTIS, H. Beyond homo economicus: evidence from experimental economics. **Ecological Economics**, v. 35, [special issue], p. 311-322, 2000a.

GINTIS, H. Strong reciprocity and human sociality. **Journal of Theoretical Biology**, v. 206, p. 169-179, 2000b.

GINTIS, H. **Game theory evolving**. Princeton, NJ: Princeton University Press, 2000c.

GINTIS, H.; SMITH, E.; BOWLES, S. **Costly signaling and cooperation**. 2001. (unpublished manuscript).

GLAESER, E.; LAIBSON, D.; SCHEINKMAN, J.; SOUTTER, C. **What is social capital?** The determinants of trust and trustworthiness. National Bureau of Economic Research, 1999. (Working Paper, 7216).

GODE, D.K.; SUNDER, S. What makes markets allocationally efficient? **Quarterly Journal of Economics**, v. 112, n. 2, p. 603-630, 1997.

GÖDEL, K. **Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme**. I. Monatshefte für Mathematik und Physik. From Frege to Gödel: a source book on mathematical logic. Harvard: Harvard University Press, 1931. p. 596-616.

GORDON, H.S. The economic theory of a common-property resource: the fishery. **The Journal of Political Economy**, Apr. 1954.

GOSWAMI, A. **O universo autoconsciente**: como a consciência cria o mundo material. Rio de Janeiro: Rosa dos Tempos, 2003.

GOSWAMI, A. **A física da alma**. São Paulo: Aleph, 2005.

GRAVES, C.W. **An emergent theory of ethical behavior based upon an epigenetic model**. Schenectady, NY, 1959. (Unpublished paper).

GRAVES, C.W. On the theory of ethical behavior. In: FIRST UNITARIAN SOCIETY OF SCHENECTADY, NY, 1961. **Paper presented…** New York, 1961.

GRAVES, C.W. The deterioration of work standards. **Harvard Business Review**, v. 44, n. 5, p. 120, 1966.

GRAVES, C.W. Levels of existence: an open system theory of values. **Journal of Humanistic Psychology**, v. 10, n. 2, p. 131-155, 1970.

GRAVES, C.W. The levels of existence related to learning systems. In: ANNUAL CONFERENCE OF THE NATIONAL SOCIETY FOR PROGRAMMING INSTRUCTIONS, 9, 1971a. Rochester, N.Y., 1971a. **Paper presented…** Rochester, N.Y., 1971a.

GRAVES, C.W. **A systems conception of personality**: remarks by Clare W. Graves on his levels of existence theory. Santa Barbara, CA: ECLET Publishing, 1971b. (Presented at the Washington School of Psychiatry, Washington, DC).

GRAVES, C.W. Human nature prepares for a momentous leap. **The Futurist**, v. 8, p. 72-87, 1974.

GRAVES, C.W. **Summary statement**: the emergent cyclical, double-helix model of the adult human biopychosocial systems. Boston, 1981.

GRAVES, C.W. **The never ending quest**: a treatise on an emergent cyclical conception of adult behavioral systems and their development. Santa Barbara, CA: ECLET Publishing, 2005.

GRAY, J. **John Stuart Mill on liberty and other essays**. Oxford University Press, 1998.

GREENE, J.D. An FMRI investigation of emotional engagement in moral Judgment. **Science**, v. 293, p. 2105-2108, Sep. 2001.

GRUBER, H.; VONECHE, J. **The essential Piaget**. New York: Basic Books, 1977.

GÜTH, W. An experimental analysis of ultimatum bargaining. **Journal of Economic Behavior and Organization**, v. 3, n. 4, p. 367, 1982.

GÜTH, W. An evolutionary approach to explain cooperative behavior by reciprocal incentives. **International Journal of Game Theory**, v. 24, n. 4, p. 323-344, 1995.

HABERMAS, J. **The theory of communicative action**: reason and rationalization of society. Boston, MA: Beacon Press, 1984. v. 1.

HABERMAS, J. **The theory of communicative action**: the critique of functionalist reason. Boston, MA: Beacon Press, 1987. v. 2.

HABERMAS, J. **Moral consciousness and communicative action**. Cambridge, MA: MIT Press, 1990.

HACKETT, S.; SCHLAGER, E.; WALKER, J.M. The role of communication in resolving commons dilemmas: experimental evidence with heterogeneous appropriators. **Journal of Environmental Economics and Management**, v. 27, n. 2, p. 99-126, 1994.

HAKSTIAN, A.R.; ROGERS, W.D.; CATTELL, R.B. The behavior of numbers factors rules with simulated data. **Multivariate Behavioral Research**, v. 17, p. 193-219, 1982.

HALAS, E. **The humanistic approach of Florian Znaniecki**. University of Munich, 2001.

HALTIWANGER, J.; WALDMAN, M. The role of altruism in economic interaction. **Journal of Economic Behavior and Organization**, v. 21, n. 1-15, 1993.

HAMILTON, W.D. The genetical evolution of social behavior. **J. Theor. Biol.**, v. 7, p. 1-52, 1964.

HAMMERSTEIN, P. Understanding cooperation: an interdisciplinary challenge. In: HAMMERSTEIN, P. **Genetic and cultural evolution of cooperation**. Cambridge, Massachusetts; London, England: The MIT Press in cooperation with Dahlem University Press, 2003. p. 1-6.

HARDIN, G. The tragedy of the commons. In: DORFMAN, R.; DORFMAN, N.S. (Eds.). **Economics of the environment**: selected readings. 3.ed. New York/London, W.W. Norton & Company, 1972. p. 5-19.

HARDIN, R. **Collective action**. Baltimore, MD: Johns Hopkins University Press, 1982.

HARSANYI, J.C. Value judgments. **The New Palgrave: A Dictionary of Economics**, v. 4, p. 792-793, 1987.

HARSANYI, J.C.; SELTEN, R. **A general theory of equilibrium selection in games**. Cambridge, Mass.: MIT Press, 1988.

HAUSMAN, D.M.; McPHERSON, M.S. **Economic analysis and moral philosophy**. Cambridge: Cambridge University Press, 1996.

HAWKINS, D.R. **I**: reality and subjectivity. W. Sedona, Arizona: Veritas Publishing, 2003.

HAWKINS, D.R. **Power and force**: the hidden determinants of human behavior. Sedona, Arizona: Veritas Publishing, 1995.

HAWKINS, D.R. **The eye of the I**: from which nothing is hidden. 2.ed. W. Sedona, Arizona: Veritas Publishing, 2001.

HAYEK, F.A. The use of knowledge in society. **American Economic Review**, v. 35, p. 519-530, 1945.

HEATH, J. **Communicative action and rational choice**. Cambridge, MA: MIT Press, 2001.

HEATH, J. **Deontic constraints**. Disponível em: <http://www.chass.utoronto.ca/~jheath/drafts.htm>. (forthcoming) Acesso em novembro: 2006.

HEBERT R.; LEHMANN, D. Theta burst and EEG pattern in normal subjects practicing the transcendental meditation technique. **Electronceph. Clin. Neurophysiol.**, v. 42, p. 397-405, 1977.

HEINER, R.A. The origin of predictable behavior. **The American Economic Review**, v. 73, n. 4, p. 560-595, 1983.

HEINER, R.A. Predictable behavior: reply. **The American Economic Review**, v. 74, n. 3, p. 579-585, 1985.

HENRICH, J. Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga. **The American Economic Review**, v. 90, p. 973-980, 2000.

HENRICH, J.; BOYD, R. Why people punish defectors: weak conformist transmissions can stabilize costly enforcement of norms in cooperative in cooperative dilemmas. **Journal of Theoretical Biology**, v. 208, p. 79-89, 2001.

HENRICH, J.; GIL-WHITE, F. The evolution of prestige. Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. **Evol. Hum. Behav.**, v. 22, p. 165-196, 2001.

HENRICH, J.; BOYD, R.; BOWLES, S.; CAMERER, C.; FEHR, E.; GINTIS, H.; McELREATH, R. **Cooperation, reciprocity and punishment in fifteen small-scale societies**. 2001. (Unpublished manuscript).

HENRICH, J.; BOYD, R.; HOPFENSIT, A.; RICHERSON, P.; SIGMUND, K.; SMITH, E.; WEISSING, F.; YOUNG, P. Group report: the cultural and genetic evolution of human cooperation. In: HAMMERSTEIN, P. (Ed.). **Genetic and cultural evolution of cooperation**. Cambridge, Massachusetts; London, England: The MIT Press in cooperation with Dahlem University Press, 2003. p. 445-468.

HENRICH, J.; BOYD, R.; BOWLES, S.; CAMERER, C.; FEHR, E.; GINTIS, H. Foundations of human sociality: economic experiments and ethnographic evidence from fifteen small-scale societies. In: OOSTERBEEK, H.; SLOOF, R.; VAN DE KUILEN, G. Differences in ultimatum game experiments: evidence from a meta-analysis. **Experimental Economics**, v. 7, p. 171-188, 2004.

HIRSHLEIFER, J. Economics from a biological viewpoint. **Journal of Law and Economics**, v. 20, p. 1-54, 1977.

HIRSHLEIFER, R.F.; COMMONS, M. Postformal cognitive-developmental theory and research: a review of its currents status. In: ALEXANDER, C.; LANGER, E. (Eds.). **Higher stages of human development**. p. 139-161, 1990.

HOBBES, T. **Leviathan**. Cambridge: Cambridge University Press, 1991.

HODGSON, G.M. The approach of institutional economics. **Journal of Economic Literature**, v. 36, p. 166-192, 1998.

HOFFMAN, E.; McCABE, K.A.; SMITH, V.L. On expectations and monetary stakes in ultimatum games. **International Journal of Game Theory**, v. 25, n. 3, p. 289-301, 1996.

HOFSTADTER, D.; GÖDEL, E. **Bach**: an eternal golden braid. New York: Basic Books, 1980.

HONGZHI, L. **Falun gong**. Taiwan: Yih Chyun Corp, 1999.

HONGZHI, L. **Falun Buddha Fa**: lectures in the United States. Taiwan: Yih Chyun Corp, 2001.

HONGZHI, L. **Falun Dafa**: essentials for further advancement II. Taiwan: Yih Chyun Corp, 2002.

HONGZHI, L. **Zhuan falun** [turning the law wheel]. Taiwan: Yih Chyun Corp, 2003.

HOWARD, P. **Reading summary for network organizations, theories of collective action**. 2002.

HUME, D. **A treatise of hunan nature**. London: Penguin Books, 1985.

HUME, D. **Dialogues concerning natural religion**. London: Penguin Books, 1990.

HURLBUT, M.A. **Clare Graves' levels of psychological existence**: a test design. 1979. Dissertation (Ph.D.) – North Texas State University, Denton, Texas.

HURWICZ, L. **But who will guard the guardians?** (Working Paper). Disponível em: <www.econ.umn.edu/workingpapers/hurwicz_guardians.pdf>. Acesso em: 1998.

IMMERGUT, E.M. The theoretical core of the new institutionalism. **Politics & Society**, v. 26, n. 1, p. 5-34, 1998.

ISAAC, R.M.; WALKER, J. Communication and free-riding behavior: the voluntary contribution mechanism. **Economic Inquiry**, v. 26, n. 4, p. 585-608, 1988.

ISAAC, R.M.; WALKER, J.M. Costly communication: an experiment in a nested public good problem. In: PALFREY, T.R. (Ed.). **Laboratory research in political economy**. Ann Arbor: University of Michigan Press, 1991. p. 269-286.

ISAAC, R.M.; MARK, J.W.; WILLIAMS, A.W. Group size and the voluntary provision of public goods: experimental evidence utilizing large groups. **Journal of Public Economics**, v. 54, n. 1, p. 1-36, 1994.

JAGER, W. **Modelling consumer behaviour**. 2000. Thesis (Ph.D.) – University of Groningen, Berlin, Germany.

JAGER, W.; JANSSEN, M.A.; DE VRIES, H.J.M.; DE GREEF, J.; VLEK, C.A.J. Behavior in commons dilemmas: homo economicus and homo psychologicus in an ecological-economic model. **Ecological Economics**, v. 35, special issue, p. 357-379, 2000.

JAMES, W. **Pragmatism**: a new name for some old ways of thinking. 1995.

JAMES, W. **The variety of religious experience**: a study of human nature. New York: The Modern Library, 2002.

JAMES JUNIOR, H.S. The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. **Journal of Economic Behavior & Organization**, v. 47, p. 291-307, 2002.

JANIS, I.L.; MANN, L. **Decision making**: a psychological study of conflict, choice and commitment. New York: The Free Press, 1977.

JANTSCH, E. **The self-organizing universe**: scientific and human implications of the emerging paradigm of evolution. Oxford and New York: Pergamon Press, 1980.

JELLAL, M.; WOLFF, F.C. Cultural evolutionary altruism: theory and evidence. **European Journal of Political Economy**, v. 18, p. 241-262, 2002.

JERDEE, T.; ROSEN, B. Effect of opportunity to communicate and visibility of individuals decisions on behavior on the common interest. **Journal of Applied Psychology**, v. 59, p. 712-716, 1974.

JOHMSON, R.N.; LIBECAP, G.D. Contracting problems and regulation: the case of fishery. **American Economic Review**, v. 27, n. 5, p. 1005-1023, 1992.

JONES, S. **The economics of conformism**. Oxford: Basil Blackwell, 1984.

JONES, J.; WILSON, W. **An incomplete education**. 3.ed. Ballantine Books, 1987.

JUNG, C.G. **Memories, dreams, reflections**. New York: Vintage Books, 1989.

KAHN, P.H. **The human relationship with nature**: development and culture. Cambridge, Massachusetts: MIT Press, 1999.

KAISER, H.F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, v. 20, p. 141-151, 1960.

KANDORI, M. Social norms and community enforcement. **Review of Economic Studies**, v. 59, p. 63-80, 1992.

KANT, I. **Crítica da razão pura**. São Paulo: Martin Claret, 2003.

KEGAN, R. **In over our heads**: the mental demands of modern life. Boston, Mass.: Harvard University Press, 1994.

KIRZNER, I. **The economic point of view**: an essay in the history of economic thought. Princeton, N.J.: Van Nostrand, 1960.

KOESTLER, A. **The act of creation**. New York: Dell, 1964.

KOESTLER, A. **The ghost in the machine**. New York: Random House. 1973a.

KOESTLER, A. The tree and the candle. In: GRAY, W.; RIZZO, N.D. **Unity through diversity**. New York: Gordon and Breach Science Publishers, 1973b. p. 287-314.

KOHLBERG, L. The child as a moral philosopher. **Psychology Today**, v. 2, n. 4, p. 25-30, 1968.

KOHLBERG L. Moral stages and moralization: the cognitive developmental approach. In: LICKONA, T. (Ed.). **Moral development behavior**. New York: Holt, Rinehart and Winston, 1976.

KOHLBERG, L.; ARMON, C. Three types of stage models used in the study of adult development. In: COMMONS, M.L.; RICHARDS, F.A.; ARMON, C. (Eds.). **Beyond formal operations**: late adolescent and adult cognitive development. New York: Praeger, 1984. v. 1, p. 357-381.

KOMORITA, S.; PARKS, C.D. Interpersonal relations: mixed-motive interaction. **Ann. Rev. Psychol.**, v. 46, p. 183-207, 1995.

KRAMER, R.M.; BREWER, M.M. Social group identity and the emergence of cooperation in resource conservation dilemmas. In: HENK, A.; WILKE, D.M.; RUTTE, C.G. **Experimental social dilemmas**. Frankfurt am Main: Lang. 1986. p. 205-234.

KREPS, D.M. Corporate culture and economic theory. In: ALT, J.E.; SHEPSLE, K.A. (Eds.). **Perspectives on positive political economy**. New York: Cambridge University Press, 1990.

KUHN, T.S. **The structure of scientific revolutions**. Chicago: University of Chicago Press, 1996.

KURZBAN, R. Biological foundations of reciprocity. In: OSTROM, E.; WALKER, J. (Eds.). **Trust & reciprocity**: interdisciplinary lessons from experimental research. New York, NY: Russell Sage Foundation, 2003. p. 105-127.

LA VIOLETTE, P.A. **Genesis of the cosmos**: the ancient science of continuous creation. Rochester, Vermont: Bear&Company, 2004.

LAIBSON, D. A cue theory of consumption. **Quarterly Journal of Economics**, v. 116, n. 1, p. 81-119, 2001.

LAKATOS, A. **Proofs and refutations**. Cambridge: Cambridge University Press, 1976.

LAKATOS, A. **The methodology of scientific research programmes**: Cambridge: Cambridge University Press, 1977.

LANGE, O. **On the economic theory of socialism**. University of Minnesota Press, 1938.

LATANÉ, B.; DARLEY, J.M. **The unresponsive bystander**. New Jersey: Prentice Hall, 1970.

LAZEAR, E. **Economic imperialism**. National Bureau of Economic Research, 1999a. (Working Paper, 7300).

LEE, W.L. **A reliability and validity study of the selected levels of psychological existence scale**. 1982. Dissertation (Ph.D.) – The University of North Carolina, Chapel Hill.

LEHMANN, D.R. The impact of altruism and envy on competitive behavior and satisfaction. **Intern. J. of Research in Marketing**, v. 18, p. 5-17, 2001.

LEIMAR, O.; HAMMERSTEIN, P. Evolution of cooperation through indirect reciprocity. **Proc. Roy. Soc. Lond. B**, v. 268, p. 745-753, 2001.

LENSKI, G. **Human societies**. New York: McGraw-Hill, 1970.

LESNOFF, M. **The structure of social science**: a philosophical introduction. London: Allen & Unwin, 1974.

LESTER, D. Cooperative/competitive strategies and locus of control. **Psychological Reports**, v. 71, p. 594, 1992.

LEVI, M. A model, a method, and a map: rational choice in comparative and historical analysis. In: LICHBACH, M.I.; ZUCKERMAN, A.S. (Eds.). **Comparative politics**: rationality, culture, and structure. Cambridge: Cambridge University Press, 1997. p. 19-42.

LINDGREN, K. Evolutionary dynamics in game-theoretic models. In: DURLAYF, A.; LANE, A. (Eds.). **The economy as an evolving complex system II**. New York: Westview Press, 1997. (Proceedings volume in the Santa Fe Institute in the Sciences of Complexity).

LINSCOTT, G. **Sustainable development strategy in a country of partial development**: viewed from the values perspective of spiral dynamics. Disponível em: <http://www.globalvaluesnetwork.com/>. Acesso em: 2006.

LOEVINGER, J. Measuring personality patterns of women. **Genetic Psychology Monographs**, v. 65, 1962.

LOEVINGER, J. **Ego development**. San Francisco, CA: Jossey-Bass, 1976.

LOEWENSTEIN, G.F. Out of control: visceral influences on behavior. **Organizational Behavior and Human Decision Processes**, v. 65, p. 272-292, 1996.

LOVEJOY, A. **The great chain of being**. Cambridge: Harvard University Press, 1964.

LUCAS, R. On the mechanics of economic development. **Journal of Monetary Economics**, v. 22, p. 3-42, 1988.

LYOTARD, J.F. The postmodern condition: a report on knowledge. In: GODZICH, W.; SCHULTE-SASSE, J. **Theory and history of literature**. Manchester; UK: Manchester University Press, 1984.

MANSKI, C.F. **Economic analysis of social interactions**. Cambridge, MA: National Bureau of Economic Research, 2000. (NBER Working Paper Series, 7580).

MANTZAVINOS, C.; NORTH, D.C.; SHARIQ, S. Learning, institutions, and economic performance. **Perspectives on Politics**, v. 2, n. 1, 2004.

MARCH, J.G. Bounded rationality, ambiguity and the engineering of choice. **Bell Journal of Economics**, v. 9, p. 587-608, 1978.

MARCHAND, H. Some reflections on postFormal thought. **The Genetic Epistemologist**, v. 29, n. 3, 2005.

MARTIN, E.; YODER, R. The Chherlung Thulo: a case study of a farmer-managed irrigation system. In: SEMINAR ON WATER MANAGEMENT ISSUES, 1983, Kathmandu, Nepal. **Proceedings…** Kathmandu, Nepal: Ministry of Agriculture, 1983a. p. 203-217.

MARTIN, E.; YODER, R. Water allocation and resource mobilization for irrigation: a comparison of two systems in Nepal. ANNUAL MEETING OF THE NEPAL STUDIES ASSOCIATION, 1983, Madison. **Paper presented…** Madison: University of Wisconsin, 1983b.

MARWELL, G.; OLIVER, P. **The critical mass in collective action**: a micro-social theory. New York: Cambridge University Press, 1993.

MASLOW, A.H. Toward a psychology of being: a theory of human motivation. **Psychological Review**, v. 50, p. 370-396, 1943.

MASLOW, A.H. **Motivation and personality**. New York: Harper & Row, 1954.

MASLOW, A.H. **New knowledge in human values**. New York: Harper & Bros., 1959.

MASLOW, A.H. **Toward a psychology of being**. Princeton, N.J: Van Nostrand, 1962.

MASLOW, A.H. **The farther reaches of human nature**. New York: The Vicking Press, 1971.

McADAMS, D.P.; AUBIN, E. A theory of generativity and its assessment through self-report, behavioral acts, and narrative themes in autobiography. **Journal of Personality and Social Psychology**, v. 62, n. 6, p. 1003-1015, 1992.

McADAMS, D.P.; AUBIN, E. Stories of commitment: the psychosocial construction of generative lives. **Journal of Personality and Social Psychology**, v. 72, n. 36, p. 678-694, 1997.

McCABE, K.A. A cognitive theory of reciprocal exchange. In: OSTROM, E.; WALKER, J. (Eds.). **Trust and reciprocity**: interdisciplinary lessons from experimental research. New York: Russell Sage Foundation, 2003. p. 275-301.

McCABE, K.A.; SMITH, V. Strategic analysis in games: what information do players use? In: OSTROM, E.; WALKER, J. (Eds.). **Trust and reciprocity**: interdisciplinary lessons from experimental research. New York: Russell Sage Foundation, 2003. p. 275-301.

McELREATH, R.; CLUTTON-BROCK, T.H.; FEHR, E.; FESSLER, D.M.T.; HAGEN, E.H.; HAMMERSTEIN, P.; KOSFELD, M.; MILINSKI, M.; SILK, J.B.; TOOBY, J.; WILSON, M.I. In genetic and cultural evolution of cooperation. Cambridge, Massachusetts; London, England: The MIT Press in cooperation with Dahlem University Press, 2003. p. 125-152.

McKENZIE, A.J. **Evolutionary game theory**. The Stanford Encyclopedia of Philosophy, 2003. Disponível em: <http://plato.stanford.edu/>.

McKENZIE, A.J. The evolutionary foundations of human altruism. **Analyse & Kritik**, v. 27, p. 106-113, 2005.

McTAGGART, L. **The field**: the quest for the force of the universe. New York, NY: HarperCollins Publishers, 2002.

MELZER, A. **The natural goodness of man**: on the system of Rousseau's thought. Chicago: University of Chicago Press, 1990.

MICHOD, R.; SANDERSON, M. Behavioral structure and the evolution of cooperation. In: GREENWOOD, P.J.; HARVEY, P.H.; SLATKIN, M. **Evolution**: essays in honor of John Maynard Smith. Cambridge: Cambridge University Press, 1985.

MILL, J.S. **A system of logic, ratiocinative and inductive**. Toronto: University of Toronto Press, 1843. v. 8.

MILL, J.S. Collected works, a system of logic ratiocinative and inductive. In: ROBSON, J.M. (Ed.). **Introdução por H.F. McRae**. Londres: Routledge & Kegan Paul, 1973. v. 7 e 8.

MILLER, G. **Managerial dilemmas**: the political economy of hierarchy. New York: Cambridge University Press, 1992.

MOHTASHEMI, M.; MUI, L. Evolution in indirect reciprocity by social information: the role of trust and reputation in evolution of altruism. **Journal of Theoretical Biology**, v. 223, p. 523-531, 2003.

MOIR, R. **The effects of costly monitoring and sanctioning upon common property resource appropriation**. Saint John: University of New Brunswick, 1995.

MOLL, J. The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. **Journal of Neuroscience**, v. 22, n. 7, p. 2730-2736, 2002.

MONGIN, P. **Is there progress in normative economics**? 2002.

MONTEROSSO, J.; AINSLIE, G.; MULLEN, P.T.; GAULT, B. The fragility of cooperation: a false feedback study of a sequential iterated prisoner's dilemma. **Journal of Economic Psychology**, v. 23, p. 437-448, 2002.

MORRISON, D.F. **Multivariate statistical methods**. Englewood Cliffs, NJ: Prentice-Hall, 1976.

MURPHY, J.J.; CARDENAS, J.C. An experiment on enforcement strategies for managing a local environmental resource. **Journal of Economic Education**, v. 35, n. 1, p. 47-61, 2004.

MYERSON, R.B. **Game theory**: analysis of conflict. London: Harvard University Press, 1991. 568 p.

MYERSON, R. Nash equilibrium and the history of economic theory. **Journal of Economic Literature**, v. 37, p. 1067-1082, 1999.

MYERSON, R.B. Fundamental theory of institutions: a lecture in honor of Leo Hurwicz. In: NORTH AMERICAN MEETINGS OF THE ECONOMETRIC SOCIETY, 2006. **Proceedings…** University of Minnesota, 2006.

NATIONAL RESEARCH COUNCIL – NRC. **Proceedings of the conference on Common Property Resource Management**. Washington, DC: National Academy Press, 1986.

NEISSER, U. **Cognitive psychology**. Englewood Cliffs, NJ: Prentice-Hall, 1967.

NELSON, R.; WINTER, S. Neoclassical versus evolutionary theories of economic growth. **Economic Journal**, v. 84, p. 886-905, 1974.

NIETZSCHE, F. **Genealogia da moral**: uma polêmica. São Paulo, SP: Companhia das Letras, 2002. <http://www.mala.bc.ca/~johnstoi/Nietzsche/genealogy2.htm>.

NORGAARD, R.B. **Development betrayed**: the end of progress and a coevolutionary revisioning of the future. London; New York: Routledge, 1994.

NORGAARD, R.B. **Development betrayed**: the end of progress and a coevolutionary revisioning of the future. New York: Routledge, 1995.

NORTH, D. Institutions. **Journal of Economic Perspectives**, v. 5, n. 1, p. 97-112, 1991.

NORTH, D. Economic performance through time. **American Economic Review**, v. 84, n. 3, p. 359-638, 1994.

NORTH, D.C. **Understanding the process of economic change**. Princeton, NJ: Princeton University Press, 2005.

NOWAK, M.A.; SIGMUND, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. **Nature**, v. 364, n. 6432, p. 56-58, 1992.

NOWAK, M.A.; KAREN, M.; PAGE, K.S. **Science**, v. 289, n. 8, Sep. 2000. <www.sciencemag.org>.

OCHS, J.; ROTH, A.E. An experimental study of sequential bargaining. **American Economic Review**, v. 79, p. 355-384, 1989.

OECHSSLER, J. Cooperation as a result of learning with aspiration levels. **Journal of Economic Behavior & Organization**, v. 49, p. 405-409, 2002.

OLSON, M. **The logic of collective action, public goods and the theory of groups**. Cambridge, MA: Harvard University Press, 1965.

ONCKEN, A. The consistency of Adam Smith. **The Economic Journal**, v. 7, n. 27, p. 444, 1897.

OPHULS, W. Leviathan or oblivion. In: DALY, H.E. (ed.). **Toward a steady state economy**. Oakland, CA: Freeman, 1973.

ORBELL, J.M.; DAWES, R.M.; VAN DE KRAGT, A. The limits of multilateral promising. **Ethics**, v. 100, n. 4, p. 616-627, 1990.

ORBEL, J.M.; VAN DE KRAGT, A.; DAWES, R.M. Covenant without the sword: the role of promises in social dilemmas circumstances. In: KOFFORD, K.J.; MILLER, J.B. (Ed.). **Social sorms and economic institutions**. An Arbor: University of Michigan Press, 1991. p. 117-134.

OSTROM, E. **Crafting institutions for self-governing irrigation systems**. Oakland, CA: ICS Press, 1992a.

OSTROM, E. The rudiments of a theory of the origins, survival, and performance of common-property institutions. In: BROMLEY, D.W. et al. (Ed.). **Making the commons work**: theory, practice, and policy. Oakland, CA: ICS Press, 1992b. p. 41-59.

OSTROM, E. **Governing the commons**. Cambridge: Cambridge University Press, 1997.

OSTROM, E. **Self-governance and forest resources**. Sindangbarang, Indonesia: CIFOR, 1999. (Occasional paper, 20).

OSTROM, E. Collective action and the evolution of social norms. **Journal of Economic Perspectives**, v. 14, n. 3, p. 137-158, 2000.

OSTROM, E. **Toward a behavioral theory linking trust, reciprocity, and reputation**. In: OSTROM, E.; WALKER, J. (Eds.). **Trust & reciprocity**: interdisciplinary lessons from experimental research. 2003. p. 19-79.

OSTROM, E. **Understanding institutional diversity**. Princeton, NJ: Princeton University Press, 2005.

OSTROM, E. The challenge of crafting rules to change open access resources into managed resources. In: INTERNATIONAL ECONOMIC ASSOCIATION ROUNDTABLE ON THE SUSTAINABILITY OF ECONOMIC GROWTH, 2007, Beijing, China. **Paper Presented…** Beijing, China, 2007a.

OSTROM, E. **Under review**: a diagnostic approach for going beyond panaceas. National Academy of Sciences, 2007b.

OSTROM, E.; WALKER, J.M. Communication in the commons cooperation without external enforcement. In: PALFREY, T.R. (Ed.). **Laboratory research in political economy**. Ann Arbor: University of Michigan Press, 1991. p. 287-322.

OSTROM, E.; BURGER, J.; FIELD, C.B.; NORGAARD, R.B.; POLIKANSKY, D. Revisiting the commons: local lessons, global challenges. **Science**, v. 284, p. 278-282, 1999.

OSTROM, E.; GARDNER, R.; WALKER, J.M. **Rules, games, and common-pool resources**. Ann Arbor: The University of Michigan Press, 2002.

PALFREY, T.R.; ROSENTHAL, H. Private incentives in social dilemmas. **Journal of Public Economics**, v. 35, p. 309-332, 1988.

PANCHANATHAN, K.; BOYD, R. **A tale of two defectors**: the importance standing for the evolution of indirect reciprocity. Disponível em: <http:/www.bol.ucla.edu/ ~buddha/TakeOfTwoDefectors.pdf>. Acesso em: 2002.

PAUCHANT, T.C. **Integral leadership: a research proposal**. Disponível em: <http://web.hec.ca/leadergraphies/dropdown/preprint.html>. Acesso em: 2005.

PAUCHANT, T.C.; BENNETT, J.; BENEFIELD, M.; DELBECQ, A.; FITZGIBBONS, D.; GODDARD, K. Integrating spirituality at work: an interview with Ken Wilber. **Management, Spirituality and Religion**, v. 1, n. 1, p. 113-131, 2004.

PERSKY, J. The ethology of homo economicus. **Journal of Economic Perspectives**, v. 9, n. 2, p. 221-231, 1995.

PETERSON, B.E.; STEWART, A.J. Antecedents and contexts of generativity motivation at midlife. **Psychology and Aging**, v. 11, n. 1, p. 21-33, 1996.

PHARES, E.J. **Locus of control in personality**. Morristown, NJ: General Learning Press, 1976.

PIAGET, J. **The moral judgment of the child**. New York: Free Press, 1932.

PIAGET, J. Intellectual evolution from adolescence to adulthood. **Human Development**, v. 15, p. 1-12, 1970/1972.

PINKER, S. **The language instinct**. New York: W. Morrow, 1994.

PINKER, S. **Robert Wright interview at Cambridge, Massachusetts**. Disponível em: <http://video.google.com/videoplay?docid=3554279466299738997>. Acesso em: 1 jul. 2002.

PINKER, S. **The blanc slate**: the modern denial of human nature. London: Penguin Books, 2003.

POPPE, M. The speci.city of social dilemma situations. **Journal of Economic Psychology**, v. 26, p. 431-441, 2005.

POPPE, M.; ZWIKKER, M. The effect of threshold level on greed, fear, and cooperation in step-level give-some and take-some games. In: LIEBRAND, W.B.G.; MESSICK, D.M. (Eds.). **Frontiers in social dilemma research**. Berlin: Springer, 1996. p. 185-204.

POPPER, K.R. **The poverty of historicism**. London: Routledge, [1957]1954.

POPPER, K. **Em busca de um mundo melhor**. 3.ed. Lisboa: Fragmentos, 1992.

POPPER, K. **The myth of the framework**: in defence of science and rationality. London: Routledge, 1994.

POSNER, R. Economic analysis of law. In: SIMMONS, R.T.; SMITH JUNIOR, F.L.; GEORGIA, P. **The tragedy of the commons revisited**: politics versus private property. Washington: The Center for Private Conservation, 1996.

PRICE, J. Guilt-prone people contribute more. In: WORKSHOP IN POLITICAL THEORY AND POLICY ANALYSIS, 2005, Bloomington. **Colloquium Presentations...** Bloomington: Indiana University, 2005.

PRUITT, D.; KIMMEL, M. Twenty years of experimental gaming: critique, synthesis, and suggestions for the future. **Annual Review of Psychology**, v. 28, p. 363-392, 1977.

QUINE, W.V.O. **World and object**. Cambridge, MA: MIT Press, 1960.

RABIN, M. Incorporating fairness in game theory and economics. **American Economic Review**, v. 83, n. 5, p. 1281-1302, 1993.

RAIFFA, H. **The art and science of negotiation**. Cambridge, MA: Harvard University Press, 1982.

REICH, W. **Psicologia de massa do facismo**. Porto: Publicações Escorpião, 1974.

REICH, W. **A função do orgasmo**: problemas econômico-sexuais da energia biológica. 19.ed. São Paulo: Brasiliense, 1995a.

REICH, W. **Análise do caráter**. São Paulo: Martins Fontes, 1995b.

REICH, W. **O assassinato de Cristo**. São Paulo: Martins Fontes, 1999.

RICHARDS, F.; COMMONS, M. Postformal cognitive-developmental theory and research: A review of its currents status. In: ALEXANDER, C.; LANGER, E. (Eds.). **Higher stages of human development**. New York: Oxford University Press, 1990. p. 139-161.

RICHERSON, P.J.; BOYD, RT. The evolution of ultrasociality. In: EIBL-EIBESFELDT, I.; SALTER, F.K. (Eds.). **Indoctrinability, ideology and welfare**. New York: Berghahn, 1998. p. 71-96.

RICHERSON, P.J.; BOYD, RT. Complex societies: the evolutionary dynamics of crude superorganism. **Human Nature**, v. 10, p. 253-289, 2000.

RICHERSON, P.J.; BOYD, R. Culture is part of human biology. In: GOODMAN, M.; MOFFAT, A.S. (Eds.). **Probing human origins**. Cambridge, MA: American Academy of Arts and Sciences, 2002. p. 59-86.

RICHERSON, P.J.; BOYD, R.T.; HENRICH, J. Cultural evolution of human cooperation. In: HAMMERSTEIN, P. (Ed.). **Genetic and cultural evolution of cooperation**. Cambridge, Massachusetts; London, England: The MIT Press in cooperation with Dahlem University Press, 2003. p. 357-388.

RILLING, J.K. A neural basis for social cooperation. **Neuron**, v. 35, p. 395-405, 2002.

RODEN, H. **Assim dizia o mestre**. 5.ed. São Paulo: Martin Claret, 2002.

ROMER, P. Endogenous technological change. **Journal of Political Economy**, v. 98, p. S71-S102, 1991.

ROOK, D.; TORBERT, W.R. Organizational transformation as a function of CEOs developmental stages. **Organizational Development Journal**, v. 16, n. 1, p. 11-28, 1998.

ROTH, A.E. Game theory as a part of empirical economics. **Economic Journal**, v. 101, p. 107-114, 1991.

ROTTER, J.B. Generalized expectancies for internal versus external control of reinforcement. **Psychological Monographs: General and Applied**, v. 80, n. 609, 1966.

ROUSSEAU, J.J. **Emile, or on education**. New York: Basic Books, 1979.

ROZIN, P.; MILLMAN, L.; NEMEROFF, C. Operations of the laws of sympathetic magic in disgust and other domains. **J. Pers. Soc. Psych.**, v. 50, p. 703-712, 1986.

RUBENSTEIN, A. Perfect equilibrium in a bargaining model. **Econometrica**, v. 50, p. 97-109, 1982.

RUDNER, R.S. Philosophy and social science. **Philosophy of Science**, v. 21, n. 2, p. 164-168, 1954.

RUDNER, R.S. The scientist qua scientist makes value judgments. **Philosophy of Science**, v. 20, p. 1-6, 1966.

RUSSELL, P.O. **Buraco branco**: nossa evolução futura e o significado do agora. São Paulo: Aquariana, 1992.

RYAN, A. **The philosophy of the social sciences**. London: MacMillan, 1970.

RYLE, G. **The concept of mind**. London: Penguin, 1949.

SAMUELSON, P.A. The pure theory of public expenditure. **Review of Economics and Statistics**, v. 36, p. 387-389, 1954.

SANFEY, A.G.; RILLING, J.K.; JESSICA, A.; ARONSON, L.; NYSTROM, E.; COHEN, J.D. The neural basis of economic decision-making in the ultimatum game. **Science**, v. 300, p. 13, 2003.

SCHIEMANN, J.W. **History and emotions, beliefs and mental models**: toward a hermeneutics of rational choice. Boston, MA: American Political Science Association Meeting, 2002.

SCHLAGER, E. **Bounding unboundable resources**: an empirical analysis of property rights and rules in coastal fisheries. Indiana University, Workshop in Political Theory and Policy Analysis, 1989. (Working Paper).

SCHLAGER, E. **Model specification and policy analysis**: the governance of coastal fisheries. 1990. Thesis (Ph.D.) – Indiana University.

SCHLAGER, E. Fishers' institutional responses to common-pool resources dilemmas. In: OSTROM, E.; GARDNER, R.; WALKER, J. (Eds.). **Rules, games, and common-pool resources**. Ann Arbor: University of Michigan Press, 1994. p. 274-265.

SCHUMPETER, J.A. **History of economic analysis**. New York: Oxford University Press, 1954.

SCOTT, E. The science and religion movement. **Skeptical Inquirer**, Jul./Aug. 1999

SCOTT, W.R. **Institutions and organizations**. Thousand Oaks: Sage Publications, 1995.

SELTEN, R. A simple model of imperfect competition where four are few and six are many. **International Journal of Game Theory**, v. 2, p. 141-201, 1971.

SELTEN, R. Bounded rationality. **Journal of Institutions and Theoretical Economic Behavior**, v. 146, n. 4, p. 649-658, 1990.

SELTEN, R. Evolution, learning and economic behavior. **Games and Economic Behavior,** v. 3, n. 1, p. 3-24, 1991.

SEN, A.K. Collective choice and social welfare. **5.3 Basic and Nonbasic Judgments & 5.4 Facts and Values**, p. 59-64, 1970.

SENGE, P. Preface to the routledge classics edition. In: ____. **Bohm**: on dialogue. New York, NY: Routledge, 2004.

SILK, J.B. Cooperating without counting: the puzzle of friendship. In: HAMMERSTEIN, P. (Ed.). **Genetic and cultural evolution of cooperation**. Cambridge, Massachusetts; London, England: The MIT Press in cooperation with Dahlem University Press, 2003. p. 37-54.

SIMON, H. A behavioral model of rational choice. **The Quarterly Journal of Economics**, v. 69, p. 99-118, 1955.

SIMON, H.A. **Theories of decision-making in economics and behavioral science**. AER, 1959.

SIMON, H.A. The architecture of complexity: hierarchic systems. **Proceedings of the American Philosophical Society**, v. 106, p. 467-482, 1962.

SIMON, H.A. Information storage as a problem in organizational design. In: GOLDBERG, A. (Ed.). **Behavioral approaches to modern management**. 1970.

SIMON, H.A **Rationality as a process and ad as product of thought**. AER, 1978.

SIMON, H. **Decision making and problem solving**. Washington, DC: National Academy Press, 1986.

SIMON, H. A mechanism for social selection and successful altruism. **Science**, v. 250, p. 1665-1668, 1990.

SIMON. H.A. A behavioral model of rational choice. **The Quarterly Journal of Economics**, v. 69, n. 1, p. 99-118, 1995.

SINNOTT, J.D. Limits to problem solving; emotion, intention, goal, clarity, health, and others factors in postformal thought. In: SINNOTT, J.D.; CAVANAUGH, J.C. (Eds.). **Bridging paradigms**: positive development in adulthood and cognitive aging. New York: Praeger, 1991.

SLOVIC, P.; FISCHHOFF, B.; LICHTENSTEIN, S. 1977. Behavioral decision theory. **Ann. Rev. Psychol.**, v. 28, p. 1-39, 1977.

SMITH, J.W. Senses of subjectivism in value theory. **Journal of Philosophy**, v. 40, n. 15, p. 393-405, 1948.

SMITH, V.L. Microeconomic systems as an experimental science. **American Economic Review,** v. 72, p. 923-955, 1982.

SMITH, A. **The theory of moral sentiments**. Indianapolis: Liberty Found, 1984.

SNYDER, M. The self-monitoring of expressive behavior. **Journal of Personality and Social Psychology**, v. 30, p. 526-537, 1974.

SNYDER, M.; GANGESTAD, S. On the nature of self-monitoring: matters of assessment, matters of validity. **Journal of Personality and Social Psychology**, v. 51, p. 125-139, 1986.

SPINOZA, B. **Ética**. São Paulo: Editora Martin Claret, 2003.

STEVENS, J. **Applied multivariate statistics for the social sciences**. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers, 1996.

STIGSBY, B.; RODENBERG, J.C.; MOTH, H.B. Electroencephalographic findings during mantrameditation (transcendental meditation): a controlled, quantitative study of experienced meditators. **Electroencephal. Clin. Neurophysiol.**, v. 51, p. 434-442, 1981.

TANG, S.Y. Institutional arrangements and the management of common-pool resources. **Public Administration Review**, v. 51, n. 1, p. 42-51, 1991.

TANG , S.Y. **Institutions and collective action**: self-governance in irrigation. San Francisco: Institute for Contemporary Studies, 1992.

TARNAS, R. **A epopéia do pensamento ocidental**. Rio de Janeiro: Bertrand Brasil, 1999.

TAYLOR, C. Interpretations and the sciences of man. In: ____. **Philosophy and the human sciences**: philosophical papers. New York: Cambridge University Press, 1985.

TEBECIS, A.K. A controlled study of the EEG during transcendental meditation: comparison with hypnosis. **Psychiatry and Clinical Neurosciences**, v. 29, n. 4, p. 305-313, 1975.

TRAWICK, P. **The moral economy of water**: 'comedy' and 'tragedy' in the andean commons. Lexington: University of Kentucky, Department of Anthropology, 1999. (Working paper).

TRIVERS, R.L. The evolution of reciprocal altruism. **Qtly. Rev. Biol.**, v. 46, p. 35-57, 1971.

VAN DE KRAGT, A.; DAWES, R.; ORBELL, J.M.; BRAVER, S.R.; WILSON, L.A. Doing well and doing good as ways of resolving social dilemmas. In: HENK, A.M.; WILK, D.M.; RUTTE, C. (Eds.). **Experimental social dilemmas**. Frankfurt am Main: Verlag Peter Lang, 1986. p. 177-204.

VERNADSKY, V. **The biosphere**. Oracle, AZ: Synergetic Press, 1986. 86 p.

VON MISES, L. Economic calculation in the socialist commonwealth. In: HAYEK, F.A. (ed.). **Collectivist economic planning**. London: Routledge, 1935.

WADE, J. **Changes of mind**. Albany: State University of New York Press, 1996. (Suny Series in the Philosophy of Psychology).

WALLACE, R.K. Physiological effects of transcendental meditation. **Science**, v. 167, n. 3926, p. 1751-1754, Mar. 1970.

WALSH, R.; VAUGHAN, F. The riddle of consciousness. In: WALSH, R.; VAUGHAN, F. (Eds.). **Paths beyond ego**: the transpersonal vision. New York: Penguin Putnam, 1993. p. 13-17.

WEBER, M. **The protestant ethic and the "spirit" of capitalism**. England: Penguin Books, 2002.

WERKMEISTER, W.H. Prolegomena to value theory. **Philosophy and Phenomenological Research**, v. 14, n. 3, p. 293-308, 1954.

WHITEHEAD, A.N. **The function of reason**. Princeton: Princeton University Press, 1929.

WIKIPEDIA. **Humanitic sociology**. Disponível em: <http://en.wikipedia.org/wiki/Humanistic_Sociology>. Acesso em: 2006.

WILBER, K. **The spectrum of consciousness**. New York : Quest Books, 1977.

WILBER, K. **Sex, ecology, spirituality**: the spirit of evolution. Boston, Mass.: Shambhala, 2000a.

WILBER, K. **Integral psychology**: consciousness, spirit, psychology, therapy. Boston , Mass.: Shambhala, 2000b.

WILBER, K. **A theory of everything**. Boston , Mass.: Shambhala, 2001a.

WILBER, K. **The eye of spirit**: an integral vision for a world gone slightly mad. Boston: Shambhala Publications, 2001b.

WILBER, K. **Boomeritis**: a novel that will set you free. Boston, Mass.: Shambhala, 2003.

WILBER, K. **Excerpt B**: the many ways we touch. [forthcoming]. <http://wilber.shambhala.com/html/books/kosmos/excerptB/intro.cfM/>. 2006a.

WILBER, K. **Excerpt D**: the look of a feeling: the importance of post/structuralism. [forthcoming]. <http://wilber.shambhala.com/html/books/kosmos/excerptD/part1.cfm/>. 2006b.

WILBER, K. **Excerpt G toward a comprehensive theory of subtle energies**. [forthcoming].<http://wilber.shambhala.com/html/books/kosmos/excerptG/part1.cfm/>. 2006c.

WILBER, K. **Conditioned dharmas**: incomplete or uncertain (IOU). Emptyuniverse. Disponível em: <http://emptyun-i-verse.blogspot.com/2006/11/conditioned-dharmas-incomplete-or.html>. Acesso em: 2006d.

WILBER, K. **Introduction to excerpts**. Disponível em: <http://wilber.shambhala.com/html/books/kosmos/index.cfm>. Acesso em: 2006e.

WILLIAMSON, O. Transaction-cost economics: the governance of contractual relations. **Journal of Law and Economics**, v. 22, 1979.

WINTER, S.G. Economic 'natural selection' and the theory of the firm. **Yale Economic Essays**, v. 4, p. 225-272, 1964.

WINTER, S.G. Satisficing, selection, and the inovative remnant. **Quartely Journal of Economics**, v. 85, p. 237-262, 1971.

WITTGESTEIN, L. **Tractatus logico-philosophicus**. London: Routledge, 1961.

WITTGENSTEIN, L. **Philosophical investigations**. Oxford: Blackwell, 1967.

WITTGENSTEIN, L. Culture and value. In: VON WRIGHT, G.H.; NYMAN, H. (Eds.). Oxford: Blackwell, 1980.

WONG, S. Positive economics. **The New Palgrave: A Dictionary of Economics**, v. 3, p. 920-921, 1987.

WU, J.; AXELROD, R. How to cope with noise in the interacted prisoner's dilemma. **Journal of Conflict Resolution**, v. 39, n. 1, p. 183-189, 1995.

WU, J.; SUN, Y. Cooperation in multi-agent bidding. **Decision Support Systems**, v. 33, p. 335-347, 2002.

YOUNG, A.M. **The reflexive universe**: the evolution of consciousness. Novato, CA: Robert Briggs Associates, 1976a.

YOUNG, A.M. **The geometry of meaning**. Novato, CA: Robert Briggs Associates, 1976b.

YOUNG, A.M. **Which way out?** Novato, CA: Robert Briggs Associates, 1980.

YOUNG, A.M. **What is God?** <http://www.arthuryoung.com/whatisgod.html>. 1985.

YOUNG , H.P. **Individual strategy and social structure**: an evolutionary theory of institutions. Princeton, NJ: Princeton University Press, 1989.

YOUNG, A.M. **Thoughts on science of life**. <http://www.arthuryoung.com/deathwish.html>. 1992.

YOUNG, A.M. **Comments on science (journal extract)**. <http://www.arthuryoung.com/comments.html>. 2005a.

YOUNG, A.M. **Has ever been a paradigm shift?** <http://www.arthuryoung.com/paradigm.html>. 2005b.

YOUNG, A.M. **Science, spirit and soul**. <http://www.arthuryoung.com/sci-spir.html>. 2005c.

YOUNG, A.M. **Extending the vocabulary of science**. <http://www.arthuryoung.com/extend.html>. 2005d.

YOUNG, J.E. A spectrum of consciousness for CEOs: a business application of Ken Wilber's spectrum of consciousness. **The International Journal of Organizational Analysis**, v. 10, n. 1, p. 30-54, 2002.

YOUNG, P. The economics of convention. **Journal of Economic Perspectives**, v. 10, p. 105-122, 1996.

APPENDIXES

APPENDIX A

Table 1A – Graves's constructs compared with other theories (includes entering, nodal, and exiting phases of change)

Graves levels	Kohlberg	Perry	Calhoun	Loevinger	Fromm	Erikison	Peck	Schein
AN (nodal)				Autistic				
AN (exiting)						< Trust vs. Mistrut>		
BO (entering)								
BO (exiting)			1 Sapient Revolution	Symbiotic	Symbiosis			
BO (nodal)								
CP (entering)								
CP (nodal)			2 Living Agricultural Revolution	Impulsive		<Autonomy vs. Shame & Doubt>	Amoral	
CP (exiting)	1 Punishment & Obedience			Impulsive Self-protective				
DQ (entering)	2. Naive instrumental hedonism			Conformist Malignant Fixated				
DQ (nodal)	3. Good boy morality		3 Authoritarian Religious Revolution	Conformist	Conformity			
DQ (exiting)	4. Law and order morality	1 Duality	4 Holistic Artistic Revolution	Conformist conscientious		<Initiative vs. Guilt>		
ER (entering)		2 Multiplistic prelegitimate					Irrational Conscientious	
ER (nodal)	5. Morality of democratic contract	3 Multiplistic	5 Scientific Exploitive Revolution					Rational Economic
ER (exiting)		4 Multiplistic		Individualistic		<Industry vs. Inferiority>	Rational Altruistic	
FS (entering)		5 Relativism competing						
FS (nodal)	6. Morality of individual principles	6 Relativistic	6. Communication Electronic Revolution			<Identity vs. Role Difference>		Social
FS (exiting)		7 Initial commitment						
GT (entering)		8 Implications of commitment						
GT (nodal)		9 commitment	7 Compassionate Systems Revolution	Autonomous		<Intimacy vs. Isolation>		Self-Actualizing
GT (exiting)						<Generality vs. Self-absorption>		
HU (entering)						< Integrity vs. despair> [not included by Graves]		
HU (nodal)								

Source: Graves (2005).

APPENDIX B

GRAVES'S PSYCHOMETRIC DATA ARRANGED ACCORDING TO SIMILARITY OF MEAN SCORE RELATIONSHIP ACROSS THE FOUR ORIGINAL SUB-TYPES TO SHOW TYPE OF VARIATION

The number code aside the asterisk indicates the sub-types the mean scores of the sub-type in each colon is statistically different from.

SNFR	(1) =	sacrifice now for reward later (later DQ)
ESC	(2) =	express self calculatedly (later ER)
SNFA	(3) =	sacrifice now for approval now (later FS)
ESWCFO	(4) =	express self with concern for others (later GT)

Table 1B – Conceptions of healthy personality sub-types

Dimension measured	SNFRL	ESC	SNFAN	ESWCFO	Characteristic of dimension
1. Cognitive complexity	1	2	3	4	Quamtitative
2. Dogmatism	4	3	2	1	Quantitative
3. Rigidity	*2-3-4 4	*1-4 3	*1-4 2	*1-2-3 1	Quantitative
4. Arguing against belief public-private	*4 1	*4 2	*4 3	*1-2-3 4	Quantitative
5. Creative novelty	*4 1	*4 2	*4 3	*1-2-3 4	Quantitative
6. Relevancy of questions	*4 1	*4 2	*4 3	*1-2-3 4	Quantitave
7. Change of set	*4 2	*4 2	*4 2	*1-2-3 4	Quantitative
8.Cue utilization	*4 1.7	*4 1.9	*4 2.4	*1-2-3 4	System specific
9. Agressiveness	*2 2	*1-3-4 4	*2 2	*2 2	System specific
10. Self causality	*2-4 2	*1-3-4 1	*2 3	*1-2 4	System specific
11. Self concept	*4 2	*4 1	*4 3	*1-2-3 4	System specific
12. Self control	*2-3-4 4	*1 1.8	*1 2.1	*1 2.1	System specific
13. “Nettler” Anomie	*2-3-4 4	*1-3-4 1	*1-2 3	*1-2 2	System specific
14. Creativity desire to be different	*2 1	*1-3 4	*2 2	 3	System specific
15. Machiavelliarism	*2 2	*1-3-4 4	*2 2	*2 2	System specific
16. Maintenance of belief input-deviant	*4 3	*4 2	*4 1	*1-2-3 4	System specific
17. Attaining new concept speed	*2-4 1	*1 3	 3	*1 3	System specific
18. Arguing against belief in public	2	1	3	4	System specific
19. Independence	*2 1.5	*1-3 4	*2 1.5	 2	System specific
20. Integrating contradiction	*4 1.7	*4 1.9	*4 2.4	*1-2-3 4	System specific
21. Opinionation – right	4	1	3	2	Cyclical
22. Loyaly	*2-3-4 4	*1-3-4 1	*1 3	*1-2 2	Cyclical
23. Religiousness	*2-3-4 4	*1-3 1	*1 3	*1-3 2	Cyclical
24. Honesty	*2-3 4	*1-3 1	*1-2 3	2	Cyclical
25. Deference	*2-4 1	*1 1.5	3	*1 1.5	Cyclical
26. Redundancy	*2-4 4	*1 1.5	3	*1 1.5	Cyclical
27. Kindness	*2 3.5	*1-3 1	*2 3.5	2	Cyclical
28. Autonomy	*2-4 1	*1 3.5	2	*1 3.5	Cyclical
29. Affiliation	3	*1-3-4 1	*1-2-4 4	2	Cyclical
30. Doodlebug time	*2-4 4	*1-4 2	*4 3	*1-2-3 1	Cyclical
31. Influence on belief input – deviant	*4 3	*4 2	*4 4	*1-2-3 1	Cyclical trend
32. Arguing against belief in private	2	2	4	1	Cyclical trend
33. Relevancy of questions	*4 1	*4 3	*4 2	*1-2-3 4	Cyclical trend
34. Authoritarianism	4	2	3	1	Cyclical trend
35. Opinionation left	*2 1	*1 1.5	2	1.5	Cyclical trend
36. Doodlebug help sought	*4 4	*4 2	*4 3	*1-2-3 1	Cyclical trend

APPENDIX C

INVITATION FLY

QUER PARTICIPAR?

O **DEPARTAMENTO DE ECONOMIA RURAL** CONVIDA A COMUNIDADE VIÇOSENSE PARA PARTICIPAR EM UMA **PESQUISA SOBRE DECISÃO DE INVESTIMENTO**.

OS PARTICIPANTES NÃO PRECISAM POSSUIR QUALQUER CONHECIMENTO ESPECÍFICO. QUALQUER PESSOA COM MAIS DE 18 ANOS, QUE SAIBA LER, ESCREVER, E REALIZAR AS QUATRO OPERAÇÕES MATEMÁTICAS PODE PARTICIPAR (SERÃO FORNECIDAS CALCULADORAS).

OS INTERESSADOS DEVEM PODER DEDICAR CERCA DE **3 HORAS** À REALIZAÇÃO DE UMA ATIVIDADE EXPERIMENTAL EM QUE SERÃO CONVIDADOS A TOMAR DECISÕES SOBRE INVESTIMENTO EM MERCADOS ALTERNATIVOS. ALÉM DISSO, TODOS OS PARTICIPANTES DEVERÃO RESPONDER A UM QUESTIONÁRIO QUE SERÁ FORNECIDO PELOS PESQUISADORES, ANTES DA FASE EXPERIMENTAL. APENAS AS PESSOAS QUE PREENCHEREM PREVIAMENTE O QUESTIONÁRIO ESTARÃO HABILITADAS A PARTICIPAR.

OS EXPERIMENTOS SERÃO REALIZADOS NO DEPARTAMENTO DE ECONOMIA RURAL, DURANTE O MÊS DE **FEVEREIRO DE 2006**. AS DATAS E LOCAIS ESPECÍFICOS PARA REALIZAÇÃO DA ATIVIDADE EXPERIMENTAL SERÃO ANUNCIADOS OPORTUNAMENTE.

TODOS **OS PARTICIPANTES RECEBERÃO UMA GRATIFICAÇÃO** DE R$ 5,00 PELO PREENCHIMENTO DO QUESTIONÁRIO, **MAIS** UM VALOR CORRESPONDENTE AO RESULTADO DAS SUAS DECISÕES DE INVESTIMENTO. ESTE ÚLTIMO VALOR É, PORTANTO, VARIÁVEL, E DEPENDERÁ DAS SUAS ESCOLHAS DURANTE O EXPERIMENTO. O VALOR TOTAL DOS SEUS GANHOS SERÁ PAGO, EM DINHEIRO, AO FINAL DA PESQUISA.

SE VOCÊ QUER PARTICIPAR, ESCREVA UM E-MAIL, **ATÉ 27 DE JANEIRO**, PARA: lffmeyer@tdnet.com.br COLOCANDO NO "ASSUNTO" A FRASE: "QUERO PARTICIPAR". INFORME SEU NOME, IDADE E PROFISSÃO/OCUPAÇÃO PRINCIPAL. VOCÊ RECEBERÁ UMA MENSAGEM DE RETORNO COM AS INSTRUÇÕES PARA ACESSAR O QUESTIONÁRIO ON-LINE E, OPORTUNAMENTE, SERÁ INFORMADO SOBRE A DATA E LOCAL PARA PARTICIPAR DA ATIVIDADE. SE VOCÊ NÃO TEM ACESSO À INTERNET, LIGUE PARA 3899-1685 (CAMPIC) PARA OBTER MAIORES INFORMAÇÕES E TER ACESSO AO QUESTIONÁRIO.

O DEPARTAMENTO DE ECONOMIA RURAL DA UFV AGRADECE.

APPENDIX D

CPR EXPERIMENT: INSTRUCTIONS AND MATERIAL SAMPLES: GENERAL INSTRUCTIONS AND EXAMPLES

INSTRUÇÕES GERAIS

1. Informações gerais

Os participantes são convidados a tomar decisões sobre como distribuir certa dotação fixa de um dado recurso econômico entre duas alternativas de investimento que denominaremos genericamente por: *Mercado 1* e *Mercado 2*. A dotação fixa do recurso a ser investido *equivale* a quinze reais (R$15,00).

O experimento consiste em 30 rodadas divididas em três grupos de 10 rodadas. A cada rodada do experimento, é *como se* cada participante tivesse novamente a dotação de R$15,00 para aplicar nos dois mercados. A cada rodada toda a dotação deve ser distribuída entre as alternativas de investimento. Ou seja, aquilo que não for investido numa alternativa é automaticamente aplicado na outra. Por isso, a cada rodada, os participantes são convidados a informar apenas as suas decisões de investimento no *Mercado 2*.

Os ganhos financeiros ao longo do experimento referem-se aos rendimentos possíveis a partir das aplicações nos mercados experimentais, e não envolvem a dotação fictícia de quinze reais. Ao final da seção experimental, os participantes serão pagos o correspondente à 1/3 dos seus ganhos financeiros ao longo do experimento, mais os cinco reais (R$5,00) de gratificação pelo preenchimento dos questionários e colaboração em nossa pesquisa. Os pagamentos serão feitos de forma individual e sigilosa. A menos que você revele, espontaneamente, nenhum outro participante conhecerá os seus ganhos financeiros no experimento.

A cada rodada do experimento, cada participante deverá tomar sua decisão de investimento individualmente. Além disso, as decisões são confidenciais. As decisões individuais serão publicadas no quadro segundo os *números de identificação e preservação do anonimato*, distribuídos aleatoriamente no início do experimento.

Durante as dez primeiras rodadas, não será permitida comunicação entre os participantes. Entre a décima e a décima primeira rodadas, os participantes terão 10 minutos para discutir a situação experimental. Durante este tempo, se acharem conveniente, vocês

poderão combinar uma regra para regular a distribuição dos investimentos entre os dois mercados. Mas não será permitida a combinação de pagamentos paralelos como, por exemplo, a divisão dos ganhos totais entre os participantes, ao final do experimento. Após os 10 minutos de comunicação, os participantes retornam aos seus lugares e as rodadas acontecem normalmente até décima quarta quinta rodada, quando os participantes poderão optar por mais 05 minutos de discussão, antes de prossegui até a vigésima rodada. Neste momento (entre a 20ª e a 21ª rodadas), será fornecido um instrumento de fiscalização do cumprimento da regra de investimentos, caso o grupo tenha proposto uma. Neste caso, nós explicaremos oportunamente como funciona o instrumento de fiscalização. Caso contrário, as rodadas procedem como na fase inicial, sem novas oportunidades de comunicação, até a trigésima e última rodada.

2. Das alternativas de investimento

2.1. Mercado 1 – O *Mercado* 1 representa uma alternativa de investimento (ou qualquer situação que gere ganhos econômicos) em que os seus ganhos individuais são independentes do quanto os demais participantes decidem investir nesta alternativa. Assim, cada real investido no *Mercado 1* gera uma (1) unidade fixa de certo produto físico qualquer (por exemplo: boi gordo). E cada unidade de produto físico do *Mercado 1* vale R$0,05 (cinco centavos). Portanto, o retorno financeiro ou **rendimento** dos investimentos no *Mercado 1* é **fixo** e igual a 5 centavos por cada real investido, independentemente das aplicações dos demais participantes, neste mercado.

2.2. Mercado 2 – O *Mercado 2* representa uma situação (ou alternativa de investimento) em que os seus ganhos individuais dependem do investimento total do grupo, nesta alternativa. Cada unidade de produto físico do *Mercado 2* vale R$0,01 (1 centavo), mas a produtividade física é variável (exemplo: os bois podem engordar mais ou menos rapidamente, dependendo do número de animais numa dada área de pastagem). Além disso, os seus ganhos individuais, no *Mercado 2*, são proporcionais ao seu investimento em relação ao investimento total realizado pelo grupo, nesta alternativa. Portanto, os ganhos financeiros ou **rendimentos** provenientes de investimentos no *Mercado 2* são **variáveis** e dependentes das decisões dos demais participantes.

Os retornos financeiros de cada mercado, bem como os ganhos totais, para os diferentes níveis de investimento, podem ser consultados nas respectivas *tabelas de ganhos*. Todos os participantes receberão tabelas de ganhos idênticas.

3. Dos procedimentos e informações a cada rodada

Além das *tabelas de ganhos*, os participantes receberão também uma *ficha de registro* e vários *cartões de decisão*. A cada rodada, sua decisão de quanto investir no *Mercado 2* deve ser anotada tanto na ficha de registro (**Coluna A**) como no cartão de decisão. Em cada rodada, os cartões de decisão serão recolhidos pelo assistente de pesquisa para publicação, no quadro, das decisões individuas e cômputo do investimento total do grupo no *Mercado 2*.

As decisões individuais serão publicadas segundo os *números de identificação*, de modo a preservar a identidade dos participantes. O valor total investido pelo grupo, no *Mercado 2*, deve ser anotado na **Coluna B** da *ficha de registro*.

4. Como calcular e registrar os ganhos a cada rodada.

Para calcular o retorno do seu investimento no *Mercado 2* você deverá, primeiro, subtrair o valor do seu investimento no *Mercado 2* do valor total investido pelo grupo, conforme anunciado e publicado no quadro pelo assistente de pesquisa. O resultado indica o valor investido pelos demais participantes, no *Mercado 2*, exceto você. Este valor deve ser anotado na **Coluna C** da *ficha de registro* e será usado para encontrar o retorno do seu investimento na tabela de ganhos do *Mercado 2*. A cada rodada, o valor dos seus ganhos, no *Mercado 2*, deve ser anotado na **Coluna D** da ficha de registro. Nas **Colunas E** e **F** são registrados os ganhos dos investimentos no *Mercado 1*. O seu investimento no *Mercado 1* é dado pela diferença entre os 15 reais de dotação inicial e o valor do seu investimento no *Mercado 2*. Os retornos dos investimentos no *Mercado 1*, que devem ser consultados na tabela de ganhos do *Mercado 1*, são anotados na **Coluna F** da ficha de registro. Para obter o seu ganho total nesta rodada basta

somar os valores anotados nas colunas D e F. O resultado deve-se anotar na **Coluna G**. Este valor deve ser igual ao encontrado na tabela de ganhos totais. Veja o exemplo.

EXEMPLO 01

Suponha que, numa dada rodada, o seu investimento no Mercado 2 tenha sido de 8 reais e que o assistente de pesquisa anuncie que o investimento total do grupo foi de 38 reais. Anote estes valores nas Colunas A e B da ficha de registro.

Para achar o seu ganho no Mercado 2, primeiro calcule o investimento dos demais participantes, fazendo: 38 – 8 = 30. Anote o valor na Coluna C. Veja o retorno do seu investimento na tabela de ganhos do Mercado 2, no encontro da coluna 8 (seu investimento) com a linha 30 (investimento dos outros participantes). Anote o valor (1,08) na Coluna D.

Neste exemplo, o seu investimento no Mercado 1 é de 7 reais (15 – 8 = 7). Anote o valor na Coluna E. Na tabela de ganhos do Mercado 1 vê-se que o retorno para o investimento de 7 reais é de R$0,35 (35 centavos). Anote o valor na Coluna F. O seu ganho total nesta rodada é portando 1,43 reais (1,08 do Mercado 2 + 0,35 do Mercado 1). Anote o resultado na Coluna G. Para conferir, veja que o mesmo valor (1,43) encontra-se na tabela de ganhos totais (Mercado 1 + Mercado 2), no encontro da coluna 8 com a linha 30. Proceda da mesma forma a cada rodada do experimento.

FICHA DE REGISTRO							
Nome: _____ Número de identificação _____							
Rodada	MERCADO 2				MERCADO 1		
	Coluna A	Coluna B	Coluna C	Coluna D	Coluna E	Coluna F	Coluna G
	Meu investimento	Investimento total	Investimento dos outros	Meu ganho	Meu investimento	Meu ganho	Ganho total
1	08	38	30	1,08	07	0,35	1,43
2							
3							

EXEMPLO 02

Seu investimento no *mercado 2* = 6 reais
Investimento total do grupo no *mercado 2* (anunciado pelo assistente de pesquisa) = 48 reais
Investimento dos outros participantes no *mercado 2* => 48 – 6 = 42 reais
Seu ganho no *mercado 2* = 0,66 reais
(veja o encontro da coluna 6 com a linha 42, na TABELA DE GANHOS: MERCADO 2)
Seu investimento no *mercado 1* = 9 reais (15 reais menos 6 reais investidos no *mercado 2*)
Seu ganho no *mercado 1* = 0,45 real (TABELA DE GANHOS: MERCADO 1)
Seu ganho total (mercado 1 + mercado 2) = 0,45 + 0,66 = 1,11
[confira o resultado na TABELA DE GANHOS TOTAL, no encontro da coluna 6 (seu investimento do mercado 2) com a linha 42 (investimento dos demais participantes no mercado 2)]

FICHA DE REGISTRO							
Nome: _____ Número de identificação _____							
Rodada	MERCADO 2				MERCADO 1		
	Coluna A	Coluna B	Coluna C	Coluna D	Coluna E	Coluna F	Coluna G
	Meu investimento	Investimento total	Investimento dos outros	Meu ganho	Meu investimento	Meu ganho	Ganho total
Teste 1							
Teste 2							
Teste 3							

---corte aqui----------

CARTÃO DE DECISÃO	
Número do jogador:	
Número da rodada:	TESTE 01
Meus investimento Mercado 2:	

CARTÃO DE DECISÃO	
Número do jogador:	
Número da rodada:	TESTE 02
Meus investimento Mercado 2:	

Instructions communication

INSTRUÇÕES PARA AS RODADAS COM COMUNICAÇÃO

Antes da próxima rodada, vocês terão dez (10) minutos para discutir entre si o que tenham observado sobre o comportamento dos ganhos obtidos dos investimentos nos *Mercados 1* e *2* até agora. Durante este tempo, vocês têm a oportunidade de estabelecer qualquer tipo de acordo para disciplinar a distribuição dos investimentos entre os dois mercados, exceto a proposição de pagamentos paralelos entre os participantes.

O estabelecimento ou não do acordo é uma decisão do grupo, não sendo obrigatório. Além disso, durante as próximas 10 rodadas (da 11ª à 20ª), não haverá nenhum mecanismo formal de fiscalização do cumprimento do acordo, caso o grupo tenha proposto algum. Neste caso, um instrumento de fiscalização será fornecido e se tornará efetivo durante as 10 últimas rodadas do experimento (da 21ª à 30ª rodada).

Após o período de dez minutos de comunicação, as rodadas acontecem normalmente até a rodada número 15. Neste ponto, os participantes deverão decidir se querem ou não utilizar mais cinco (5) minutos de comunicação. Caso julguem desnecessários os cinco minutos, as rodada ocorrerão sem interrupção até a rodada número 20. Não há mecanismo formal para a decisão sobre usar ou não os cinco minutos de comunicação.

Durante os períodos de comunicação, os pesquisadores e assistentes deixarão a sala para que os participantes discutam a situação experimental mais à vontade. Por isso, os períodos de comunicação estarão sendo gravados, a fim de manter este registro na pesquisa.

Para manter o anonimato referente às comunicações, serão sorteados crachás com as letras de A até H. Qualquer referência que eventualmente venha a ser feita a respeito das comunicações, por ocasião da discussão dos resultados da pesquisa, será realizada de acordo com as letras nos crachás de identificação.

Notem que as distribuições tanto dos crachás quanto dos números de identificação são feitas de forma aleatória. Portanto, não há qualquer relação entre as letras nos crachás e os números de identificação. Sendo assim, o anonimato das decisões de investimento continua sendo preservado.

Instructions sanctioning

INSTRUÇÕES PARA RODADAS COM FISCALIZAÇÃO

A partir deste momento, estamos fornecendo um instrumento para permitir a fiscalização da norma ou regra para distribuição dos investimentos entre os *Mercados 1* e *2*, caso os participantes tenham instituído uma. O instrumento de fiscalização consiste na possibilidade de aplicação de uma *multa* sobre os participantes que descumprirem a norma estabelecida pelo grupo durante os períodos de comunicação. Caso o grupo não tenha

estabelecido nenhuma norma ou regra para os investimentos, as rodadas acontecerão normalmente, sem a possibilidade de aplicação de multas ou comunicação.

A decisão sobre aplicar a multa é individual e exige o pagamento de uma taxa correspondente, por parte do participante que decida aplicá-la. A tabela abaixo mostra os valores das multas existentes e das respectivas taxas que serão cobradas para a sua aplicação.

Multa (R$)	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80
Taxa (R$)	0,05	0,08	0,10	0,12	0,14	0,16	0,18	0,20

Como a decisão sobre aplicação e valor da multa é individual, o participante que houver descumprido a norma, em cada rodada, pode ou não ser multado. Além disso, ele pode ser multado por mais de um participante, numa mesma rodada.

Os participantes que desejarem aplicar multas podem fazê-lo sobre um ou mais participantes que tenham violado a regra estabelecida pelo grupo. Participantes que respeitarem a norma não estão sujeitos à multa. Além disso, as multas não se aplicam sobre eventuais quebras da norma em rodadas anteriores ao estabelecimento deste instrumento de fiscalização.

Nas rodadas em que houver infração, as suas decisões sobre aplicação de multas, com a identificação (pelo número) dos participantes a serem multados, devem ser registradas na *ficha de fiscalização*. Nestes casos, as fichas de fiscalização deverão ser completadas e em seguida serão recolhidas para entrada dos dados referentes a multas e taxas. Se o participante decide pela não aplicação da multa, basta entregar a ficha de fiscalização em branco ou com o valor zero, abaixo do número de identificação do(s) infrator(es).

Após o cômputo do valor total das multas eventualmente aplicadas, e antes que se tomem novas decisões de investimento, os *autos de infração* serão distribuídos, individualmente, para todos os participantes. Dessa forma, o valor das multas (se houver) não será de conhecimento público. Apenas o infrator conhecerá o valor da multa recebida. Além disso, os números de identificação dos participantes que decidirem aplicar multas também não serão conhecidos pelos participantes eventualmente multados.

ATENÇÃO: O valor das multas e taxas será descontado dos ganhos dos participantes, em cada rodada, podendo ocorrer valores negativos.

APPENDIX E

TEST OF THE FACTORS' STABILITY

In order to facilitate comparison, we show only the factor loadings above |.30| (which are the statistically significant ones in the sample used in the main text) and highlight (in bold) the strongest variable-component correlations. In both cases, the order of the factors extraction changes, in relation with the results in the main text (Table 6.4), though their explanatory power (Table 6.3) remains pretty much the same. The egocentric wave now appears as component 2 while the absolutistic wave appear as the component 1. The sociocentric wave remains as the component 3.

In the extraction using only the participants in the CPR experiment (n = 196 due to four missing cases) the egocentric wave (component 2) now explains 21.336 percent of the total variance and the absolutistic wave (component 1) explains 23.157 percent. The sociocentric wave (component 3) explains 26.426 percent of the total variance. The three components explain 70.918 percent of the total.

In the extraction using only the participants in the PG and the UG experiments (n = 126, one missing case) the egocentric wave (component 2) explains 20.543 percent of the total variance and the absolutistic wave

(component 1) explains 22.860 percent. The sociocentric wave (component 3) explains 26.862 percent of the total variance. The three components together explain 70.264 percent of the total.

Notice also that the strongest component correlation with the variable BO_most changed position in the extraction using the data set of the participants in the CPR experiment, in relation to both other data set (using the entire sample and only the information of the participants in the PG and UG experiments). This fact does not affect the interpretation of the factors, though, since the variable BO_most is somewhat evenly correlated with the there components (Table 6.4). Also, as explained in the text, we are not relying only in the strongest variable-component correlations for the interpretation of the factors, since we this is not a purely exploratory application of the Factor Analysis and Graves's constructs provide us the theoretical background for taking to mean the factor loading matrix.

Only the participants in the CPR experiment

Rotated Component Matrix[a]

	Raw			Rescaled		
	Component			Component		
	1	2	3	1	2	3
BO_most	.960			**.479**		
BO_least	-1.665		-2.120	-.458		**-.583**
CP_most		1.487			**.715**	
CP_least		-3.893			**-.929**	
DQ_most	2.967	-1.340		**.789**	-.357	
DQ_least	-1.358	2.296		-.428	**.723**	
ER_most			-3.406			**-.846**
ER_least			1.965			**.698**
FS_most			3.632			**.831**
FS_least		1.054	-.644		**.527**	-.322
GT_most	-3.569			**-.884**		
GT_least	1.422			**.694**		

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Only the participants in the PG and UG experiments (n = 126)

Rotated Component Matrix[a]

	Raw			Rescaled		
	Component			Component		
	1	2	3	1	2	3
BO_most	.849	.927		.448	**.489**	
BO_least	-1.489		1.817	-.423		**.516**
CP_most		1.317			**.696**	
CP_least		-3.620			**-.905**	
DQ_most	3.122			**.832**		
DQ_least	-1.533	2.272		-.477	**.707**	
ER_most			3.495			**.869**
ER_least	.927		-2.013	.327		**-.711**
FS_most		-1.629	-3.618		-.367	**-.815**
FS_least		.886	.654		**.462**	.341
GT_most	-3.201			**-.857**		
GT_least	1.403			**.688**		

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

APPENDIX F

PERSON CORRELATIONS (GROUPS 21 AND 22 EXCLUDED)

Table 1F – Pearson correlations (groups 21 and 22 excluded)

	n = 180	Individual total investment per treatment (tokens)			Individual totals		
		Baseline	Communication	Sanctioning	Defection (frequency)	Fine (R$)	Fee (R$)
Egocentric (Factor 1)	Pearson Correlation	-0,055	-0,001	-0,200**	0,092	0,046	0,096*
	Sig. (1-tailed)	0,233	0,497	0,004	0,109	0,270	0,099
Absolutistic (Factor 2)	Pearson Correlation	0,029	-0,020	-0,006	0,090	-0,059	-0,051
	Sig. (1-tailed)	0,349	0,393	0,470	0,115	0,216	0,247
Sociocentric (Factor 3)	Pearson Correlation	-0,031	-0,026	0,139*	-0,231**	-0,041	0,006
	Sig. (1-tailed)	0,340	0,367	0,032	0,001	0,294	0,467
BO_most	Pearson Correlation	0,016	0,009	-0,039	0,020	0,031	0,123**
	Sig. (1-tailed)	0,416	0,451	0,303	0,395	0,341	0,050
BO_least	Pearson Correlation	0,024	0,046	-0,119*	0,119*	0,119*	0,049
	Sig. (1-tailed)	0,376	0,272	0,056	0,055	0,056	0,257
CP_most	Pearson Correlation	-0,016	0,065	-0,152**	0,171**	-0,023	-0,001
	Sig. (1-tailed)	0,414	0,194	0,021	0,011	0,381	0,496
CP_least	Pearson Correlation	0,047	-0,005	0,198**	-0,089	-0,094	-0,092
	Sig. (1-tailed)	0,267	0,471	0,004	0,119	0,105	0,110
DQ_most	Pearson Correlation	0,019	-0,061	-0,027	0,043	-0,028	-0,054
	Sig. (1-tailed)	0,401	0,208	0,357	0,285	0,352	0,237
DQ_least	Pearson Correlation	-0,083	-0,049	-0,223**	0,049	0,018	0,092
	Sig. (1-tailed)	0,133	0,259	0,001	0,257	0,406	0,111
ER_most	Pearson Correlation	0,007	0,018	-0,051	0,156**	-0,007	-0,048
	Sig. (1-tailed)	0,461	0,405	0,249	0,018	0,461	0,260
ER_least	Pearson Correlation	-0,053	-0,021	0,227***	-0,269***	-0,032	-0,097*
	Sig. (1-tailed)	0,240	0,392	0,001	0,000	0,337	0,097
FS_most	Pearson Correlation	0,006	-0,013	0,218**	-0,234***	-0,051	-0,032
	Sig. (1-tailed)	0,468	0,429	0,002	0,001	0,247	0,333
FS_least	Pearson Correlation	-0,003	0,023	-0,102*	0,147**	0,041	0,134**
	Sig. (1-tailed)	0,483	0,377	0,087	0,024	0,295	0,037
GT_most	Pearson Correlation	-0,030	0,013	-0,055	-0,044	0,082	0,067
	Sig. (1-tailed)	0,346	0,432	0,231	0,278	0,137	0,187
GT_least	Pearson Correlation	0,065	0,011	-0,040	0,102*	-0,045	-0,041
	Sig. (1-tailed)	0,193	0,444	0,298	0,087	0,276	0,290

*** Correlation is significant at the 0.01 level (1-tailed).

** Correlation is significant at the 0.05 level (1-tailed).

* Correlation is significant at the 0.15 level (1-tailed).