

## **Cognition and Norms: Toward a Developmental Theory Linking Trust, Reciprocity, and Willingness to Cooperate**

**Leandro F. F. Meyer**  
Socio-Environmental Institute  
Federal Rural University of the Amazon – Brazil  
[leandro.meyer@ufra.edu.br](mailto:leandro.meyer@ufra.edu.br)

### **Abstract**

I suggest that the constructivist developmental framework in psychology is of real significance for advancing our understanding on rational action and normative commitment in social action dilemmas. Yet, the recognition of the implications of the developmental perspective to deal with intersubjective conflicts of action has been hindered by often undisputed epistemological presuppositions which deny cognitive content to “value judgments,” “moral questions,” and the existential relevance of consciousness and culture. In this paper, I bring the epistemological issue to the fore in order to introduce a proposal for integrating of the developmental point of view into the Institutional Analysis and Development framework. I rely chiefly upon Jürgen Habermas’s *discourse theory of ethics* and his developmental account of the human capacity to coordinate interaction through *communicative action*. I then suggest how the implications of the resulting integration can be tested in subsequent experimental research. The interested reader is directed to preliminary experimental results reported elsewhere.

*Key words:* social dilemmas; IAD framework; developmental psychology; discourse ethics

### **Introduction**

The way we frame the inquiry in any subject has decisive implications. With reference to the co-governance of Common-Pool Resources (CPRs) and the investigation of other complex social-ecological dilemmas, the Institutional Analysis and Development (IAD) framework has been object of considerable thought and reflection by many scholars over the years. The IAD framework has influenced the analysis of a myriad of issues during recent decades, and has been particularly helpful in exposing the limits of traditional policy recommendations addressing the misuse of common-pool, open-access resources derived straightforwardly from the conventional economic theory of “externalities.” In this regard, the IAD framework has structured the investigation of a wide diversity of settings and helped scientists interested in collective-action problems to better understand why some social arrangements are sustainable over time and others collapse (e.g., Ostrom 1990, 2003, 2005;

Schlager 1990; McKean 1992; Tang 1994; Ostrom et al. 1994; Isaac et al 1994; Hackett, et al 1994; Wade 1996; Baland & Platteau 1996; Agrawal 2001; among many others).

Among the lessons scholars have learned from the vast empirical research of the last several decades is the recognition that problems of overharvesting and misuse of ecological systems are rarely due to a single cause (Ostrom, 2007). Specifically, in relation to how different incentives confronting individuals affect their decisions in collective-action situations, both field and laboratory research have shown that participants' behaviors are affected not only by the structural characteristics of the outcomes (high or low, certain or probable, et cetera) and of the group of people involved (large or small group, with or without communication, with or without a leader, et cetera), but also by the specific content or context of the interaction situation (investment decisions, social events, environmental issues, et cetera) (Kollock, 1998; Komorita & Parks, 1995; Kopelman et al., 2002; Lepyard, 1995; Van Lange, Liebrand, Messick, & Wilke, 1992; Ostrom, Gardner & Walker, 1994; Poppe, 2005).

Yet, a core challenge confronting commons researches is how to explain norm-governed action without falling into the theoretical trap that all that is needed is to assume that individuals learn and use norms (Ostrom, 2005), but also without abandoning the action frame of reference and supplying purely functionalist explanations of these norms (cf. Heath, 2008). In addition, there is also the challenge of how to explain the variety of behavioral responses to *similar* incentive structures and contexts of action without falling into the theoretical trap that all that is needed is to assume that individuals just differ culturally or idiosyncratically from each other.

In this paper, I suggest how the constructivist developmental framework in psychology can help researchers studying social dilemmas to face these challenges. A key step in this direction is to bring Koestler's (1973) concepts of *holon* and *holarchy*<sup>1</sup>—which have lately become central to the IAD framework—to make a sense of what otherwise seems to be an

---

<sup>1</sup> Arthur Koestler (1973) coined the term *holon* to refer to that which, being a *whole* in one context, is simultaneously a *part* in another. *Holarchy* is a self-organizing, open-ended, nested hierarchy which has holons as its components. The sequence from elementary particles to atoms to molecules to cells to organs to organisms is often cited as illustration of holons' capacity of self-organization in a growing holarchy.

incommensurable diversity of personality modalities, attitudes and perspectives, normative and value orientations ensuing from contingent features of individual biographies, social narratives and cultural traditions.

Supported by substantial empirical research, the constructivist developmental framework assembles the wide diversity of human interiority into a coherent holarchy of perspective structures, along which the continual restructuring of the sociocognitive inventory is described as a process of *dialectical sublation* (Habermas, 1990) of perspectives wherein higher-order structures gradually replace the lower ones while preserving them in a reorganized form. To the extent that successive perspectives are in fact assessable *discrete whole-part totalities* (i.e. holons) that have a bearing on discernible behavioral patterns, as the developmental framework holds, it should be possible to examine the interplay among institutional incentives and the subjective makeup of users and managers of focal SESs through experimental research informed by psychosocial assessment.<sup>2</sup>

The policy implication of this suggestion is that, according to the findings of developmental psychology, and in order to produce the expected results, the institutional incentives should be tuned to the characteristics of each perspective structure, viewed as discernible psychosocial centralization *stages*, due to unique motivational needs, means, attitudes, and intentionality that differ within each stage. In addition, the research findings within the developmental framework suggest that enhanced solutions to commons co-governance dilemmas should pass through the proposition of institutions designed to facilitate a swifter flow of humans on the holarchy of interior growth. Support for this evolutionary perspective of cultural intersubjectivity comes from an increasing consensus of scholars familiar with the findings in this field that human interior development *tends* toward heightened behavioral freedom; declining egocentrism; growing inclusiveness and care for others; greater mutual understanding through communication; expanded cooperation; and strengthened

---

<sup>2</sup> The extension of the concept of holon to address norm-governed action brings a number of subtle traps, particularly relating to the possible confusion between *dominator hierarchies* and *growing holarchies*. Considering the present focus, it is worth mentioning that the prior distinction between, first, the interior (subjective) and the exterior (objective) facets of holons, and, then, between their individual (singular) and collective (plural) presentations seems a key step in avoiding the mentioned confusion (Wilber, 1995; 2000).

moral capacity to honor normative commitment, without the need of external enforcement or punishment.

However, while the conceptual insertion of the constructivist developmental point of view into the IAD framework's scheme fits smoothly, as we shall see, its full integration as genuine basis for explaining rational action in morally relevant conflicts of action requires a demonstration that descriptions of different forms of moral reasoning, forms of reciprocity, modes of dialogue, interaction types, and social perspectives do in fact satisfy the conditions of a stage model conceived in terms of a logic of development. But a demonstration of this type, which we find in Habermas's (1987; 1990) developmental account of our capacity to coordinate interaction through *communicative action*, also requires facing empiricist objections to cognitivist approaches in moral theory, and the adoption of "criteria of coherence to govern a *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).

I thus dedicate this paper to introducing these epistemological issues implicated in this proposal for integrating of the constructivist developmental point of view into the IAD framework. Following a brief reference to the most common objections to the developmental approach, I specify the epistemological matter and introduce only the most general terms of Habermas's *nonfoundationalist* epistemological underpinning as the basis (1) for moving beyond strategic action, without abandoning the action-theoretic frame of reference, and (2) for grounding the structural relationships Habermas postulates among *moral stages*, *social perspectives*, and *stages of interaction*. Next, I illustrate these structural connections accompanied by a short clarification of how Habermas counts on them for grounding moral judgment, action types, and forms of interaction in a logic of development using Selman's *stages of social perspective taking* and Laurence Kohlberg's *stages of moral reasoning* as connecting links and instances of *indirect* empirical validation of his *discourse theory of ethics*. At this juncture, I adjoin the stages in Clare Graves's (1970) emergent-cyclical conception of *adult* personality systems development to the connections Habermas establishes between the

stages in Selman's and Kohlberg's models and the stages that occur in his own reconstruction of the development of communicative action. This step is based on the correlations Graves himself establishes between the stages in Kohlberg's model and the ones described in his own. Keeping in mind all this series of correlations, supported by a whole program of philosophical justification rooted in the paradigmatic change brought about with the so-called *linguistic turn*, I illustrate the integration of the paradigm of psychological structuralism into IAD framework by means of simple schematic representations. Experimental results suggesting the robustness of Graves's constructs for predicting behavior of Brazilian participants in three different collective-action dilemmas under variable institutional conditions can be found in Meyer and Braga (2009).

### **Moral noncognitivism: on the epistemological barrier to the acknowledgement of the developmental point of view**

A number of factors have been preventing the recognition of the implications of many reputed developmental models to approach rational action in social sciences. A basic difficulty ensues from the common belief that psychosocial development refers to childhood and adolescence, that is, to the first twenty years of life. "Traditionally," as Marchand (2005), puts it; "experts in developmental psychology analyzed the growth of the child and of the adolescent, holding that development ends before adult life begins" (cf. e.g. Inhelder and Piaget 1955; Piaget, 1970/1972). Were that the case, this fact would evidently move the developmental framework out of serious consideration for addressing social dilemmas, since the actors involved in most relevant situations are typically adult humans. Currently, however, researches are coming to an increasing understanding that human subjective and intersubjective developments have indeed the potential for evolving all the way through the adult life (Graves, 1971; Riegel 1973; Arlin 1975; Basseches 1980, Kramer 1983; Pascual-Leone 1984; Commons, Richards and Armon 1984; Commons, Sinnott, Richards and Armon 1989; Sinnott 1984, 1989; among others).

A second difficulty is that each of the numerous facets or streams of consciousness comprising the overall *self* appears to have its own internal drives or laws of transformation

toward greater complexity and integration. Hence, when considered in its entirety, the overall self of particular individuals does not show a sequential or stage-like development, but appears instead as a rather fluid and flowing affair, with much overlapping and interweaving, resulting in a meshwork or dynamic spiral of consciousness unfolding (Wilber, 2000b, p. 34). Hence, the simple intuition of what seems to be an almost infinite number of multiple modalities of individual personalities stirs a natural sense of incommensurability supporting ordinary relativistic objections against the stage developmental framework in general. However, modern psychological structuralism takes all that intertwining into account and entails careful methodological design for assessing particular streams of consciousness and specific self-related competencies, which are defined as capacities not only to solve but also to recognize the very existence of particular types of problems (e.g. empirical-analytic, moral-practical or interpersonal relationship). Along these lines, as Wilber (ibid) reports, the bulk of research has continued to find that each self-related developmental line itself tends to unfold in a stage-like, sequential, and nested hierarchical fashion, and that self's *center of gravity*, so to speak, tends to hover around one basic level of consciousness at any time (p. 35). Furthermore, according to him, "One of the striking things about the present state of developmental studies is how similar, in broad outline, most of its models are" (p. 5). In fact, by comparing a sizeable number of developmental models and theories, also Richards and Commons (1990) indicate that "The stage sequence [in all of that theories and modes] can be aligned across a *common* developmental space," and that, "The harmony of alignment shown suggests a possible reconciliation of [these] theories..." (p. 160; see also Commons, 1981).

Yet, when it comes to the subject of morally relevant conflicts of action, as in social dilemmas, the acknowledgment of the developmental framework is hindered most of all by the common idea that "value judgments" or "moral questions" are rationally undecidable. As Heath (2001) indicates, a critical consequence of this view, which is often unstated, is that "most social theorists simply assume that any agent who acts on the basis of a moral principle, or social norm, is not rationally justified in doing so" (p. 2). According to him, "This is what underlies the widespread tendency among social theorists 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” (ibid). Heath further points to how the presumption of non-rationality makes it tempting to abandon the action frame of reference and supply purely functionalist explanations for the coherence of norm systems and the adaptability of norm-governed action. This trend is noticeable in the current blast of interest in the sociobiological evolutionary framework for explaining human sociability and adherence to norms—visible also in connection with the IAD framework’s approach to normative agreement vis-à-vis the co-governance of social-ecological systems.

As suggested, the mentioned hindrance is epistemological in nature. In Heath’s words, “The traditional reason for thinking that normative commitments are irrational, or unjustifiable, depends upon a rather specific conception of rationality and justifiability known as *foundationalism*” (ibid, p. 2). As a theory of justification, foundationalism is an attempt to answer the so-called *regress argument*—a fundamental problem in epistemology which suggests that any attempt to justify a given statement inferentially gives rise to an infinite regress of new arguments that can be introduced in support of that statement, but which will contain premises that themselves stand in need of justification. As Heath (ibid, p. 197) explains, the only way to break out of this regress is to use the conclusion as premise (i.e. reason in circle), or simply break off the chain of reasons (i.e. make an undefended assumption).

Foundationalism represents an instance of the later strategy as it holds that there is a class of “basic” beliefs (also called *foundational* beliefs) which are intrinsically (i.e. noninferentially) justified by virtue of their *empirical* content. Hence, foundational beliefs are said to be self-justifying or self-evident to the extent that they are not justified by beliefs or constructs other than sensorial perception. Clearly, since value judgments and moral arguments cannot be grounded in any direct experience of the physical world, the foundationalist epistemology implies that moral judgments are essentially noncognitive, as ordinarily understood. This is why, as one of the first to bring a nonfoundationalist conception of rationality to the task of understanding the logic of social action, Jürgen Habermas referred

to this view as *moral noncognitivism*. Moral noncognitivism in turn severs the internal connection between norms and justifying grounds, i.e., it wipes out what constitutes, in Habermas's (1993) terminology, the *rational foundation of normative validity*. In this way, foundationalism ultimately denies that it is possible to justify normative claims, and constitutes the epistemological theory behind different forms of *moral relativism* and *skepticism*.

This epistemological underpinning has obvious implications in determining the way we are supposed to explain rational choice in morally relevant conflicts of action, and in particular the role of communication in producing normative commitment in those situations. Basically, moral argumentation is assumed to be incapable of producing any greater level of agreement than that which agents already bring with them. As a result, linguistic communication is subtly reduced to a background for strategic interactions, providing agents with the information—such as common knowledge of preferences, action alternatives and action-outcome linkages, other participants' reputation as cooperators, expectations of sanctions, and so on—against which the agents determine maximizing strategies.

While the instrumental conception of rationality does not itself presuppose or depend upon any sort of moral noncognitivism (Heath, 2001), the foundationalist standpoint is probably what explains, for instance, Hardin's (1968) emphatic repudiation of moral argumentation for addressing commons dilemmas.

## **Beyond foundationalism: toward stages of moral judgment**

Heath (2001) indicates that a common response to the relativist point of view on morals has been to accept the formal component of the foundationalist analysis, seeking only to deny the narrow empiricist list of belief-types that are claimed to be capable of "objective validation." He mentions, for instance, that naturalist, realist, and intuitionist theories of morality "all attempt to show that moral judgments can be 'grounded' in some class of noninferentially justified beliefs that will be uniform across individuals and cultures" (p 198). However, he also points out that all of these theories suffer from well-known difficulties, so that the relativist position seems quite strong in this context. Conversely, rooted in the new

paradigm of epistemology brought about with the linguistic turn, Habermas's strategy in responding to the relativist stance on morality denies the force of the regress argument entirely, and is governed by a nonfoundationalist defense of the cognitivist conception of moral judgment.

Following Heath's outline, Habermas's discourse-theoretic view has two basic components: first, Habermas claims that noncognitivist's concerns about the truth-aptness of moral judgments has significance only if one assumes that truth represents some kind of correspondence relationship between sentences and states of affairs in the world. But if one denies that this sort of "objectivity" plays any role in vindicating the truth-claims associated with beliefs, then our ability to justify beliefs has nothing to do with their reference to the physical world. Similarly, when the relativist questions the ultimate justifiability of moral judgments, the argument is persuasive only if it presupposes a *monological* conception of rational justification, that is, when justification is tacitly treated as a process involving only agent's cognitive states and the objects of representation—and that has the effect of reducing all public practices of justification to either secondary or derivative. But if one assumes, as does Habermas, that justification is always *dialogical*, that is, a process involving an attempt to justify a claim to some other person, so that justification *to others* is taken as the primary phenomenon, then there is no longer any a priori reason to think that moral questions are any less decidable than empirical or scientific ones.<sup>3</sup>

In summary, Habermas suggests that one can defeat moral noncognitivism by rejecting the traditional project of analytic epistemology, including both the received (correspondence) theory of truth and the received (foundationalist) view of justification. Heath also suggests that one reason that some theorists have been inclined to take this more radical

---

<sup>3</sup> Heath (2001) notes, nonetheless, that Habermas retains a residual adherence to the idea that "truth" has some connection with description, so that he introduces the idea of "rightness" to capture the imperative dimension of morality. To clarify, Habermas's distinction between *truth*, *rightness*, and *truthfulness* is connected with his theory of modernization and the *differentiation of value sphere*, defined as a reference system of "three worlds" (objective, social, and subjective) that communicative actors make the basis for of their effort to reach mutual understanding. Heath suggests that Habermas's "multidimensional" theory is unnecessary, seeing that a "deflationary" (i.e. dialogical) view of truth already eliminates the suggestion that truth involves some kind of correspondence relation between mind and world. As he puts it, adopting a deflationary theory "allows the theorist to say that moral statements as straightforwardly true, without suggesting that there are moral facts [or possible worlds] out there to which they correspond."

step is that “foundationalism does not offer a very persuasive justification for any kind of belief, including empirical ones” (2001, p. 198). For this reason, as he indicates, the foundationalist epistemology has become increasingly discredited in philosophical circles, largely as a consequence of the linguistic-turn and its correspondence of language with behavior. But 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.

Now, I bring this epistemological puzzle to the fore because a cognitivist perspective on value judgments seems vital for the recognition of the implications that stage developmental psychology may have in addressing collective-action dilemmas. As Habermas (1990) indicates, “Any developmental theory of the capacity for moral judgment must presuppose [the] possibility of distinguishing between right and wrong moral judgments” (p. 120). The cognitivist view admits that moral motivation has its sources in the *affective* psychological development of individuals and the formation of personal identity, which are contingent on socialization into forms of communal life that foster and reinforce sensitivity and openness to the claims of others. However, basic *universalist* and *formalist* assumptions sets Habermas’s discourse theory of ethics in opposition to the fundamental assumptions of both *moral relativism* and *material ethics* (i.e., ethic theories oriented to issues of happiness, and which tend to ontologically favor some particular ethical life or other). Such basic assumptions seem similarly critical for the recognition of the significance of the developmental framework in our present context. Again, in Habermas’s words, “A theory of moral development that attempts to outline a general path of development would be doomed to failure from the start if moral judgment could not claim universal validity” (ibid, p. 121). In this vein, the universalist-formalist framework of discourse ethics works like a rule that, first, “eliminates as nongeneralizable content all those concrete value orientations with which particular biographies or forms of life are permeated,” and then “establishes a *procedure* based on presuppositions and designed to guarantee the impartiality of the process of judging” (ibid, p. 122).

Hence, as the experiential basis of obligation or sense of duty, the moral feelings represent, in Habermas's view, more than expressions of contingent emotions, preferences, and decisions of individual actors, and more than the contingent features of their traditions or culture. As he explains, while "mere" conventions bind, so to speak, in a groundless fashion by custom alone, duties, by contrast, derive their binding force from the validity of norms of interaction that claim to rest on good reasons. This is why, according to Habermas (1994, p. 41), we feel obligated only by norms of which we believe that, if called upon to do so, we could explain why they both deserve and admit of recognition on the part of their addressees (and of those affected). In other words, according to Habermas's discourse ethics, "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." Sanctions, he says, however much they are internalized, "are not constitutive of normative validity; they are *symptoms* of and already felt, and thus antecedent, violation of a normatively regulated context of life" (ibid, p. 42, emphasis in the original). Similarly, "It is not because 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).

Heath (2001) notices that at the heart of Habermas's approach is an essentially contractualist intuition, namely, that only rationally compelling forms of obligation are those that have been freely incurred by agents. However, the overtly voluntaristic aspects of contractualism is avoided through the claim that agents enter into rationally motivated agreement when attempting to adopt a rule to govern their interactions whenever they participate in *linguistically organized social practices*. Resting on the epistemological paradigm brought about with the linguistic turn, Habermas incorporates *speech act theory* into his model of social action to demonstrate that the mere use of language, as the background practice that secures the intelligibility of speech acts, imposes pragmatic constraints or procedural commitments that we recognize, post hoc, as morally significant. As Heath puts it, "Habermas can therefore be said to 'ground' morality in argumentation" (ibid, p. 281), which is characterized not as a set of logico-semantic relations between linguistic expressions, but as

“a social practice structured by procedural commitments to open participation, free expression, immunization against force, and so on” (p. 282).

Clearly, Habermas recognizes that, when stated baldly, this view might lead to the absurd inference that all social practices are grounded in rational consensus (Heath, 2001), thus conflicting with both “the repressive character evidenced in the fact that norms, demanding obedience, take effect in the form of social control” (cf. Habermas, 1987, p. 39), and “the fact that argumentation plays a much less important public role in some societies than it does in others” (Heath, 2001, p. 303). Habermas relies on transcendental arguments to demonstrate that the morally relevant presuppositions set in his discourse theory of ethics are indispensable to any argumentation, i.e., that agents cannot simply choose to revise particular rules governing practical discourse, and that different groups cannot have entirely different conceptions of what counts as rational argumentation (Heath, 2001).<sup>4</sup> Also, it is precisely in Habermas’s explanation for these empirical differences -- in the extent to which different societies or groups depend upon explicitly discursive procedures to secure social integration -- that his theory turns out to be fundamentally *stage developmental*. Yet, before entering Habermas’s route for grounding moral stages, social perspectives, and types of interaction and reciprocity in a logic of development, I must first provide at least a brief outline of the typological character of his account of rational action.

## **Beyond instrumental rationality and strategic action: toward mutual understanding and moral consciousness**

As Heath (2001, p. 13) observes, “The first thing to notice about Habermas’s theory of communicative action is that it is a *typological* theory” (emphasis in the original). This remark is particularly important considering the present purpose of bringing the developmental point of view on morality together with the IAD framework. Hence, in setting forth his theory of communicative action, Habermas does not reject the instrumental conception of rationality and

---

<sup>4</sup> While a detained discussion of Habermas’s transcendental argumentation falls beyond the scope of this article, it is worth mentioning that limitative results of game theory show that strategic action is unable to generate a stable social order, and so “agents will sooner or later have to fall back upon general resources of communication in order to achieve a fully institutionalized interaction pattern” (Heath, 2001, p. 281).

replace it with an alternative, “communicative” conception. As Heath (ibid) explains, Habermas takes as his point of departure that agents have available to them a set of different, often incommensurable standards of choice. Communicative action turns out to be action governed by a particular standard, namely, that of reaching *understanding*, while instrumental action is action governed by a different standard: that of reaching *success* in respect of the intended consequences or outcomes of a chosen action.

According to Habermas’s typology, *instrumental action* and *speech acts* form two “elementary forms of action.” From this, the introduction of a second agent generates *social action*, understood as a complex phenomenon constructed out of the interaction of the two elementary forms of action. According to this view, rational agents engaged in social action are always in a position where they face a problem of *interdependent expectations*, which can be resolved by drawing upon the resources of either instrumental action or speech. When the actors are *primarily* interested in the consequences or outcomes of their actions, social action takes the form of *strategic action*, in the standard game-theoretic sense. But when speech is used to resolve the coordination problem derived from the regress of anticipations it generates the form of action that Habermas characterizes as *communicative action* (cf. Habermas, 1990, p. 133).

This basic scheme is indicated by the straight lines in the Figure 1. The upward oblique line indicates that communicative action is *not* the same as speech. Just like strategic action, it also presupposes the basic teleological structure of action inasmuch as the actors are assumed to continue interestedly in carrying out their plans and bring about certain states of affairs in the world. In Habermas’s (ibid) words, the two social action types differ in that “for the *model of strategic action*, a structural description of the action directly oriented toward success is sufficient, whereas the *model of action oriented toward reaching understanding* must specify the propositions of an agreement, to be reached communicatively, that allows alter to link his action to ego’s” (where alter is a second person and ego is the first person) (p. 134). In other words, when engaged in communicative action, actors are assumed to be “prepared to harmonize their plans of action through internal means, committing themselves to pursuing

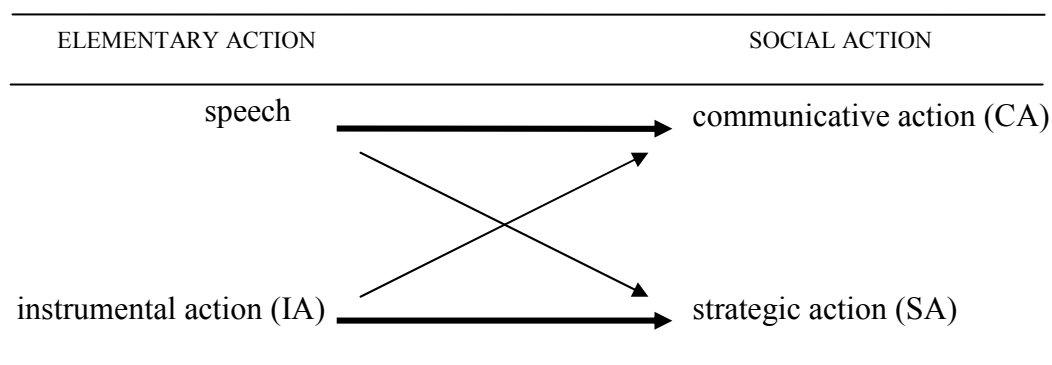
their goals only on the conditions of an agreement—one that already exists or one to be negotiated—about definitions of the situation and prospective outcomes” (ibid).

As indicated in the previous section, the discourse theory of ethics maintains that the mere use of language as an explicit coordination mechanism (rather than consistently aligning beliefs through strategic reasoning) imposes morally relevant constraints on types of goals that agents can pursue, and the means that they can employ (Heath, 2001). According to Habermas’s view, these constraints are rooted in the development of language as the aspect of *public accountability* that renders speech intelligible in the first place. Hence, they are *not* merely intralinguistic, but are rather suggested to constitute the very foundation of social order, and to incorporate certain commitments that impose direct constraints on an agent’s future conduct (Habermas, 1996, p. 8).

On the other hand, the descending oblique line in Figure 1 indicates that strategic action is *not* a simple generalization of instrumental action. Like communicative action, it relies upon linguistic recourses as well. However, in this case, language provides only the background for information exchange, so that the intralinguistic objectives of speech acts are subordinated to each agent’s individual projects and maximizing strategies. Hence, in the model of strategic action, communication “does not exercise any constraint on the range of action alternative available to agents” (Heath, 2001, p. 24).

Clearly, this is why, insofar as we stick with the canonical assumption of self-interest of standard game theory, Nash’s (1950, 1951) exclusion of communication for distinguishing

Figure 1: Elementary action types combine to produce social action types.



Source: Adapted from Heath (2001, p. 25)

between cooperative and noncooperative games turns out to be really a superfluous condition, as Harsanyi and Selten (1988, p. 3) have suggested. But even if we acknowledge that individuals differ with regard to the extent they take the interest of others into account in the decisions they make, the intrinsic valuation they may place on particular types of actions or reaching particular types of outcomes (Ostrom, 2005), then commitment to an instrumental conception of rationality still obscures, and mystifies, in fact, the understanding of the role of communication in fostering normative commitment. The recognition of this limitation is precisely the motivation under Habermas's (1984/1987) central claim that instrumental models do not provide a sufficient basis for a *general* theory of rational action.

In this regard, it is worth mentioning that Heath (2001) provides a straightforward game-theoretic demonstration in support of Habermas's rationale for introducing communicative action as a second type of rational action. Heath's limitative results suggest, in his words, that "As far as the instrumental model of rationality is concerned, the fact that agents are able to communicate successfully is completely mysterious" (p. 81). According to him, "This means that any attempts to expand the notion of rational action to account for communication starts out with a certain *prima facie* plausibility" (*ibid*).

Now, considering (1) that the sort of case that Habermas has in mind is precisely of agents discussing the validity of alternative rules to govern their interactions in morally relevant conflicts of action, and (2) that his typological theory gets rid of nothing from the standard instrumental approach in order to add communicative action (Figure 1), there seems to be no cogent explanation for social scientists interested in social dilemmas not paying adequate attention to the implications of the developmental perspective in this field, other than a structural or institutional bias against the cognitivist view on morals.

## **On grounding stages of moral reasoning and norm-governed action in a logic of development**

In what follows, I roughly outline Habermas's developmental account of our capacity to coordinate interaction through communicative action. The focus turns on pointing to the postulated structural relationships that impart plausibility to the parallels drawn in Table 1.

In broad lines, Habermas's argument is aimed at showing that the stages that occur in his historical reconstruction of the development of communicative action, which takes the form of an interpretation of work by Emile Durkheim and George Herbert Mead,<sup>5</sup> are recapitulated in the ontogenesis of our capacity to speech and action, and isomorphic to the stages described in Laurence Kohlberg's model of the development of sociocognitive and moral reasoning.<sup>6</sup> The connecting links are provided by Selman's account of sociocognitive development in relation to stages of social perspective taking, which Habermas reformulates in terms of structures of social interaction. "The point of this chain of argument is to connect structures of moral judgment to structures of social interaction in such a way that their developmental-logical features stand out more clearly" (McCarthy, 1990, p. ix).

As indicated previously, moral cognitivism and the dialogical model of justification play a central role in this reconstruction. Thus, by defining *discourse* as reflective form of communication action, Habermas situates the morally relevant presuppositions of practical argumentation as the tail end and point of reference in a constructivist *learning* process, in which complex forms of social action have given rise to competences resting on repeatedly reorganized sociocognitive inventories and perspective structures that have, in turn, permitted the emergence of more sophisticated forms of action. Viewed within the development of a complex structure of perspectives that culminates in a *decentered understanding of the world* (more on this below) displayed by subjects who act with an orientation toward reaching understanding, Habermas distinguishes *stages of interaction* in terms of different *achievements of coordination*, expressing a development that is *directed* and *cumulative*. Habermas is primarily concerned in pointing to the existence of certain *breaks* along the ontogenesis of our capacity to coordinate interaction which support the view that successive perspective structures are in fact *discrete totalities* (1990, p. 168). Along these lines, the developmental account of discourse ethics is made compatible with the constructivist notion of

---

<sup>5</sup> This phylogenetic account is in the fifth chapter of *The Theory of Communicative Action* (1984/1987, v. 2).

<sup>6</sup> The ontogenetic ground of Habermas's onto-phylogenetic parallel is developed in the fourth chapter of *Moral Consciousness and Communicative Action* (1990). The interested reader may want to consult in addition the third chapter of *Justification and Application* (Habermas, 1993).

learning that likewise informs Kohlberg's conception of the transition from one stage of moral reasoning to the next. "The use of these [action-theoretic] analytic tools for the reconstruction of the Kohlberg framework is thus intended to demonstrate the explanatory fruitfulness of the theory of communicative action, and to lend support to the broader developmental account offered in the analysis of Mead and Durkheim" (Heath, 2001, p. 45).

Put briefly, Kohlberg's framework postulates three major levels of ability, namely: *preconventional*, *conventional*, and *postconventional* reasoning. Each level includes two stages, thus comprising *six stages of moral judgment* (see Table 1). At the preconventional level, while the ability to distinguish between natural consequences of an action and those that follow from the disappointment of expectations is present, the child adopts a uniformly instrumental orientation toward both, and rules are perceived as strictly external constraints. In the first stage of the preconventional level, the reasons for doing right are avoidance of punishment and the superior power of authorities. At the second stage, the reason for doing right is to serve one's own needs or interests in a world where one must recognize that other people also have their own interests. At the conventional level, social rules have been "internalized" either in the form of direct commands or role expectations. On the whole, the conventional attitude toward these rules is one of pure conformity. At the first stage of the conventional level, the reasons for doing right are needing to be good in one's own eyes and those of others, caring for others, and the basic Golden Rule. At the second stage, the reasons for doing right are to keep the accepted institutions going as a whole, and self-respect linked to the conscience of meeting one's defined obligation, or that of the consequences of coordination failures (often expressed in some reasoning like: "What if everyone did it"?). At the postconventional level, norm-conformity is maintained on the grounds of abstract principles or values. At the first stage, the reasons for doing right are, in general, feeling obligated to obey the law from a perspective of sense of duty, that is, in the condition of a participant in a social contract, which is understood as being rationally motivated for the good of all. Finally, at the second stages of the postconventional level, the reason for doing right is

just that, as a rational person, one has seen (i.e., realized) the validity of principles and has become committed to them (cf. Kohlberg, 1981, pp. 409ff; see also Kohlberg, 1969 and 1976).

Overall, Habermas (1990, p. 123) suggests that Kohlberg's six stages of moral judgments "can be regarded as *gradual approximations* in the dimension of reversibility, universality, and reciprocity to structures of impartial or just judgments about morally relevant conflicts of action," which are implicated in the procedural presuppositions of *discourse*, as a (third) stage of interaction (Table 1). However, Habermas notices that Kohlberg saddles his representation of sociomoral perspectives with the job of grounding moral stages in a logic of development.<sup>7</sup> In order to keep such grounding strictly within the action-theoretic frame of reference, Habermas first singles the *justice conceptions* derived from the sociocognitive inventory at any particular point out from Kohlberg's sociomoral perspectives, and then replaces them with the stages of perspective taking developed by Robert Selman (1980) (see Table 1). Habermas next reformulates Selman's stages of social perspective in terms of structures of social interaction, which set the parameters for the constructive learning of basic sociocognitive concepts in children and adolescents. Viewed in terms of a *progressively decentered understanding of the world*, as indicated previously, the *stages of interaction* are described in terms of *perspective structures* (Table 1) implemented through different *types of action* (Figure 1) and *forms of reciprocity* involving corresponding *structures of behavioral expectations* (omitted in Table 1).

"To the extent to which these perspectives, embodied and integrated in interactions, fit readily into the scheme of a logic of development," Habermas says, "it will be possible to ground stages of moral judgments by tracing Kohlberg's moral stages first to social perspectives and ultimately to stages of interaction" (ibid, p. 132). As indicated, Habermas sets out a case that the proposed *hierarchy of action types* (Table 1) does indeed reflect a logic of development by pointing to the existence of certain breaks or leaps in the ontogenesis

---

<sup>7</sup> As Habermas explains, while the correlations Kohlberg establishes between these sociomoral perspectives and the stages of moral reasoning seem intuitively correct, this plausibility is achieved "at cost of a description in which the sociocognitive conditions of moral judgment have already been blended with the structures of those judgments."

of the complex structure of perspectives that support our capacities to speak and act, and that culminates in the decentered understanding of the world and the procedural presuppositions of discourse ethics.

In order to keep the narrative focused on our primary motivation for bringing the development point of view into contact with the IAD framework—namely, that of examining experimentally whether normative commitments achieved on the grounds of communication alone for the co-governance of common-pool resources and provision of public goods are or not facilitated by higher levels of psychosocial development—I will skip the details of Habermas's reconstruction. I do this to single out the reasons that, according to him, the ability to act from the perspective of a strict concept of morality (as an autonomous and rational sense of duty) can evolve only at the *postconventional* level, while the ability for acting strategically requires only an actualization of the structure of perspective applying to the *preconventional* level, without requiring any further reorganization of the sociocognitive inventory.

To show that this occurs, Habermas first redefines the preconventional types of action in terms of *forms of reciprocity* linked to different *structures of behavioral expectations* (not shown in Table 1). In this fashion, interaction controlled by authority is redefined in terms of an *asymmetrical form of reciprocity* which tends to obtain whenever the authority to control others' contributions to the interaction is unequal, as in the family. Conversely, the *symmetrical form of reciprocity* obtains when the participants exercise mutual control over their contributions to the interaction, as in egalitarian friendship, for example (p. 147). These differentiations are correspondingly reflected in two different forms of action coordination: *authority-governed complementarity*, and *interest-governed reciprocity*, to which actors can resort in the face of both cooperative and competitive relationships.

Habermas suggests that authority-governed complementary and interest-governed symmetrical social relations define two different types of interaction that can embody the *same* perspective structure, namely: the reciprocity of action perspectives typical of Selman's level 2 of perspective taking (Table 1). According to Selman, the sociocognitive inventory of children

at this level—i.e. analogously structured concepts of behavioral expectations, authority, motives for action, and the ability to act—enables them to control interactions by deception if necessary. An asymmetry between the developmental requisites for strategic action and action oriented toward reaching understanding starts to emerge as we recognize that in cooperative relationships, the participants renounce the use of deception, whereas in authority-governed relationships, the dependent partner cannot resort to deception, even in cases of conflict. “Hence, the option of influencing alter’s behavior by means of deception exists only when ego construes the social relationship as symmetrical and interprets the action situation in terms of conflicting needs” (ibid, p. 148).

As shown in Table 1, Habermas correlates the justice concept based on the complementarity of order and obedience, which is built into Kohlberg’s first stage of moral reasoning, with the considerations that will guide action when one sees oneself as dependent, and tries to resolve the conflict between ego’s needs and alter’s demands by avoiding threatened sanctions. On the other hand, the concept of justice based on symmetry of compensation, set in Kohlberg’s second moral stage, emerges only when one starts to see power as distributed equally, and may try to avail oneself of the possibilities of deception that exist in symmetrical relations. Habermas then brings up results from Flavell’s (1968) experiment in order to trace the reorganization of the preconventional stage of interaction and show how strategic action comes to be differentiated from competitive behavior.

In Flavell’s experiment, two cups concealing different amounts of money are put side down on a table. Each cup bears a label in plain view indicating the payoff value supposedly hidden under the cup. The participants are shown that the relationship between the inscription and the actual amount hidden can be varied at will. Ego’s task is to secretly distribute the payoffs in such a way that alter’s will fail to guess where the greater amount is hidden. The point of the game is clear: alter will try to win as much as she can, and ego will try to prevent this by means of deception

Habermas points out that if the participants in the experiment have the perspective structure of Selman’s level 2 (see Table 1) they will choose what Flavell called strategy B.

Following the strategy B, alter chooses the cup labeled “lower payoff,” as she reasons that ego wants to fool her by *not* concealing the higher payoff under the cup labeled “higher payoff.” On the other hand, participants who are able to engage in Selman’s level 3 of perspective taking will choose Flavell’s strategy C, which is a mixing strategy emerging from the recognition that alter sees through ego’s strategy B. As this mutual (symmetrical) recognition establishes an infinite regress of anticipations, strategy C comes out from alter’s realization that the chances of losing is as great as the chance of winning, no matter what she decides to do.

Habermas suggests that strategy C is characteristic of a type of action that is possible only at the *conventional* stage of interaction (see Table 1), because it requires a coordination of observer and participant perspectives that is lacking at Selman’s level 2, but necessary for the restructuring of pre-conventional competitive behavior into strategic action. It is this shift that, according to Habermas, allows ego to attribute stability over time to alter’s pattern of attributes and preferences, so that alter stops being perceived as someone whose actions are determined by shifting needs and interests and begins being viewed as a subject who intuitively follows rules of rational action. “Beyond this, however, *no* structural change in the sociocognitive inventory is required. In all other respects the pre-conventional inventory is adequate for the strategic actor” (1990, p. 150).

On the other hand, as Habermas’s puts it, the passage to normatively regulated action cannot be adapted so economically to the conventional stage of interaction. According to him, pre-conventional modes of coordinating action come under pressure in areas of behavior *not* dominated by competition, wherein deception is precluded. In these situations the sociocognitive inventory does require a global reconstruction to make room for a mechanism of nonstrategic coordination of action. As Habermas explains, this mechanism must be independent of both authority relations to an actual reference person and of direct links to self-interests, so that “this stage of conventional but nonstrategic action requires basic sociocognitive concepts revolving around the notion of a suprapersonal will” (*ibid*, p. 152). Habermas then goes on to discuss the structural breaks underlying his justification of the

developmental sequence associated with the emergence of different concepts and institutions embodying the idea of a suprapersonal authority, such as *loyalty* to social roles and *legitimacy* of rules (see Table 1). Concepts and intuitions of this kind provide the elements for constituting a social world of legitimately ordered interpersonal relations and for judging actions according to whether or not they conform to or violate socially recognized norms. At the conventional level, these judgments connect in turn with the justice concepts of *conformity to roles* and *conformity to systems of norms*, as shown in Table 1.

At this point, Habermas indicates that the complex structure of perspectives underpinning normatively regulated interactions—which includes the differentiation of a formal *three-world reference system* (objective, social, and subjective) to which correspond three different *attitudes toward the world* (objectifying, norm-conformative, and expressive), and the three basic modes of language use (first, second, and third-person *communicative roles* of speakers and hearers)—satisfy the structural preconditions of a communicative action in which individual plans of action are coordinated by means a mechanism for reaching understanding through communication (ibid, p. 158).

As indicated previously, the *third stage of interaction*, i.e., discourse (Table 1), takes form only when communicative action becomes fully reflexive. At this stage, the complexity of the perspective structure undertakes a further growth in order to make room for the *hypothetical attitude* that characterizes the decentered understanding of the world and allows participants in argumentation to leave behind the horizon of unquestioned, intersubjectively shared, nonthematized certitudes of a quasi-natural social world in order to focus on and test validity claims that are initially raised implicitly in communicative action and are naively carried out along with it. According to Habermas, the structural leap is marked by the synthesis of the two systems of world perspectives and speaker perspectives. “On the one hand, the system of world perspectives, which has been refracted, as it were, by the hypothetical attitude, is [now] constitutive of claims of validity that are thematized in argumentations. On the other hand, the system of fully reversible speaker perspectives is constitutive of the framework within which participants in argumentation can reach rationally motivated agreement” (p. 159).

In discourse, then, the two systems that had been fully developed at the second stage of the conventional level are put in relationship to one another. This enables participants in communicative action that want to reach a *shared* understanding *about something* in the objective, social, or subjective worlds to adopt, when necessary, an *objectifying* attitude to a given state of affair, a *norm-conformative* attitude to legitimately ordered interpersonal relations, and a *expressive* attitude to their own lived experiences, but also vary these attitudes in relation to each of the three words. This flexibility makes it possible to confront external nature not only in a objectifying attitude but also in a norm-conformative or an expressive one, to confront society not only in a norm-conformative attitude but also in a objectifying or an expressive one, and to confront inner nature not only in an expressive attitude but also in a objectifying or a norm-conformative one (ibid, p. 138-9). At the same time, the prior polarity involving communicative action and strategic action is overcome in discourse, as the success-orientation of competitors is assimilated into argumentation. As Habermas explains, what happens in argumentation is that “proponents and opponents engage in a *competition with arguments* in order to convince one another, that is, in order to reach a consensus” (ibid, p. 160). Actually, the condition that arguments are not regressively reduced to mere means of influencing each other, as often presumed along with an exclusionary instrumental conception of rationality, is what distinguishes the communicative from the strategic use of communication. “In discourse,” Habermas says, “what is called the *force* of the better argument is wholly unforced” (ibid).

Thus, “In the light of hypothetical claims to validity the world of existing states of affairs is theorized, that is, becomes matter of theory, and the world of legitimately ordered relations is moralized, that is, becomes a matter of morality” (ibid, p. 161). This “moralization of society” undermines the normative power of the factual, so that institutions that have lost their quasi-natural character can be turned into “so many instances of problematic justice” (ibid). A new reorganization of the fundamental sociocognitive concepts available at the stage of role behavior and normatively governed interaction becomes necessary in order to rationally justify the “uprooted and now free-flowing systems of norms” (ibid). At the postconventional level,

norms of action are subordinated to *principles*, or higher-order norms. “The notion of the legitimacy of norms of action is now divided into the components of mere de facto recognitions and worthiness to be recognized” (ibid). Correspondingly, a parallel differentiation occurs in the concept of duty, where “the respect for the law is no longer considered an ethical motive per se” (ibid). To dependence on existing norms, is opposed “the demand that the agent make the validity rather than the social currency of a norm the determining ground of his action” (ibid). That is, to heteronomy it is opposed autonomy (Table 1).

In short, Habermas claims that a strict (cognitivist, universalist, formalist) concept of morality can evolve only at postconventional stage, for “only at the postconventional stage is the social world uncoupled from the stream of cultural givens” (ibid, p. 162). To be sure, it is precisely the sight of plural relativism, which comes into view at the postconventional stage, which makes the autonomous justification of morality an unavoidable problem (ibid).

Now, if Habermas’s action-theoretic account of development of the sought-after moral point of view admittedly requires distinctions not easy to operationalize, the difficulty to understand how the conceptions of justice emerge from the sociocognitive inventory of the corresponding stages of interaction can be facilitated by a key insight. This insight, which Habermas properly attributes to Durkheim, is that there is no *specific* socialization process through which agents acquire moral dispositions. As Habermas puts it, “In trying to understand this process, one has to take into account that the normatively regulated fabric of social relations is moral *in and of itself*, as Durkheim has shown” (ibid, p. 164, emphasis in the original). In Heath’s (2001) words, “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” ( p. 8). At any rate, as clearly Habermas recognizes, a hypothetical reconstruction of this type can serve at best as a guide for further research. It is in with this intent that I bring Clare Graves’s emergent-stage conception of adult personality systems development into contact with the IAD framework.

Table 1 – Stages of interaction, social perspectives, moral stages, and value systems

Types of action (Habermas)	COGNITIVE STRUCTURES					Value systems (Graves)
	Perspective Structure	Concept of authority	Concept of motivation	Social perspective/ concept of justice	Stage of moral reasoning (Kohlberg)	
<p><b><u>Preconventional</u></b> Interaction controlled by authority</p> <hr/> <p>Cooperation based on self-interest</p>	<p>Reciprocal interlocking of action perspectives (Selman's Level 2)</p>	<p>Authority of reference persons: externally sanctioning will</p>	<p>Loyalty to reference persons: orientation toward reward &amp; punishment</p>	<p>Egocentric / Complementarity of order and obedience</p> <hr/> <p>Egocentric / Symmetry of compensation</p>	<p>1. Punishment and obedience</p> <hr/> <p>2. Naive instrumental hedonism</p>	<p><b>3<sup>rd</sup> → 4<sup>th</sup></b></p> <hr/> <p><b>3<sup>rd</sup> → 4<sup>th</sup></b></p>
<p><b><u>Conventional</u></b> Role behavior</p> <hr/> <p>Normatively governed interaction</p>	<p>Coordination of participant and observer perspectives (Selman's Level 3)</p>	<p>Internalized authority of supra-individual will: loyalty</p> <hr/> <p>Internalized authority of and impersonal collective will: legitimacy</p>	<p>Duty versus inclination</p>	<p>Primary group perspective / Conformity to roles</p> <hr/> <p>Collectivity's perspective / Conformity to existing systems of norms</p>	<p>3. Good boy good girl morality</p> <hr/> <p>4. Law and order morality</p>	<p><b>4<sup>th</sup></b></p> <hr/> <p><b>4<sup>th</sup> → 5<sup>th</sup></b></p>
<p><b><u>Postconventional</u></b> Discourse</p>	<p>Integration of speaker and world perspectives (Habermas's <i>decentered understanding of the world</i> orientation)</p>	<p>Ideals versus social validity</p>	<p>Autonomy versus heteronomy</p>	<p>Principled perspective / Orientation toward principles of justice</p> <hr/> <p>Procedural perspective / Orientation toward procedures for justifying norms</p>	<p>5. Morality of democratic contract</p> <hr/> <p>6. Morality of individual principles</p>	<p><b>5<sup>th</sup></b></p> <hr/> <p><b>6<sup>th</sup> → 7<sup>th</sup></b></p>

Source: Author's configuration, adapted from Habermas (1990, p. 166) and Graves (2005, p. 443).

## **Opening the IAD framework to interior holarchies and hermeneutics: toward further experimental research**

Although Habermas's nonfoundationalist epistemology and action-theoretic account of the ontogenesis of communicative action are both of great relevance for the matter at hand, the assessment of Kohlberg's moral stages does not seem to offer an adequate basis for experimental research aimed at analyzing the interplay among institutional incentives, communication opportunities, and groups' diverse capacities to coordinate interactions and achieve cooperative outcomes in collective-action dilemmas. This allegation has to do with a number of "anomalies" and problems around which the debates surrounding Kohlberg's approach frequently revolve, plus the fact the Kohlberg's stages refer to the development of the growing child to adolescence and preclude any stages of adult learning.

Habermas himself returns to four primary problems, counting (1) the lack of experimental evidence for the existence of Kohlberg's hypothetical stage 6 of moral judgment, (2) the cases of regression that occur in the postadolescent period, (3) the question of accommodating relativists or value skeptics as a group in Kohlberg's stage model, and (4) the question of whether structuralist theory can be combined with the findings of ego psychology in a way that would do justice to the psychodynamic aspect of the formation of judgments (Habermas, 1990, p. 171).

He suggests that the nature of these problems may possibly become clearer if we take into account the following important issues: "first, the [increased] degrees of [behavioral] freedom the adolescent attains when he makes the transition from normatively regulated action to discourse and achieves detachment from the social world of quasi-natural embeddedness, second, the problems of mediating between morality and ethical life that arise when the social world is moralized and cut off from the certainties that the lifeworld provides, third, the escape route that the adolescent takes when he distances himself from the devaluated traditional world of norms but stops there without taking the further step of reorganizing the sociocognitive inventory of the conventional stage as a whole, and fourth, the discrepancies between moral judgment and moral action that result from a failure to

separate the attitude oriented toward success from the attitude oriented toward reaching understanding" (ibid).

Habermas then addresses each of these issues systematically, as a contribution to theory in construction. Yet, at this point, I will leapfrog the theoretical discussion in order to merely indicate the practical significance of the stage developmental point of view for advancing our understanding of rational behavior in social dilemmas, coupled with the variable extent of normative commitment that can be evaluated by means of assessing Clare Graves's constructs (Table 2) in experimental research oriented with the IAD framework.

I will *not* justify the alternative to Graves's model by means a defense of the accurateness of the parallels he establishes between the stages in Kohlberg's model and those described in his own (Table 1), or take on justifying the plausibility of the extended correlations with Habermas's account, implied in Table 1. Rather, I will simply point toward a number of qualitative proprieties in Graves's scheme which renders it, in my view, particularly suitable for the experimental examination of groups' capacity to adopt rules for coordinating their interrelation in morally relevant social action conflicts.

First, differently from Selman and Kohlberg, Graves's constructs describe a sequence of emergent worldviews and value-based behavior systems resulting from research with adult subjects, ranging from 18 to 61 years of age, both male and female (Graves, 1971, p. 8). Graves's focus on adult thinking and argumentation revolving on conceptions of health, and mature personality led him to pay considerable attention to both psychodynamic aspects and environmental factors that may arrest development, as suggested by Habermas. As part of the conceptualization of the developmental order put forth in his model, Graves examined how change in publicly stated values and conceptions of mature personality occurs by observing whether his experimental subjects stick with a defense of their original standpoints or revise them in face of both authority and peer argumentation on behalf of different values and conceptions.

Connecting thirty eight measured dimensions, the postulated developmental order in Graves's model accommodates multiple natures of variation: quantitative, trend, cyclical,

cyclical with trend, and system specific. Among them, *increasing behavioral freedom* and *decreasing egocentrism* are essential features in his model, which are explained as resulting from the same dialectical sublation of perspectives to which Habermas refers. Further, as a pioneer with regard to the notion that “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” (1981), Graves’s conception is coherent with Koestler’s concepts of holon and holarchy, recently adopted by the IAD framework, as holarchies are necessarily ever-emergent, i.e., open-ended structures.

Finally, the substance of Graves’s constructs resides precisely on revealing the different set of values individuals may place on actions and outcomes affecting the well-being of others. In this regard, Graves’s model puts forward that people tend to oscillate back and forth between two fundamental stances, much like the relative position of a pendulum in its arc between “me” (agency) and “we” (communion) orientations (Cowan and Todorovic, 2005). According to Graves, this cyclical turn involving the agency and the communion capacities of the self produces two basic families of behavioral systems: namely, *express-self systems* and *sacrifice-self systems*, which have manifest implication for the analysis of situations where the individual and collective interests collide.

Table 2 presents the basic themes, style of thinking, and cyclical aspect concerning the eight stages constituting Graves’s model. That egocentrism decreases is noticeable by observing the change in the basic themes of the express-self systems, and in particular the dramatic shift of attitude toward others exhibited in the assertive expression of the self’s interests at the seventh stage. As suggested in Table 1, this attitude change in the transition from the 6<sup>th</sup> to 7<sup>th</sup> stage seems basically coherent with the morally relevant presuppositions of discourse, as a third stage of interaction.

Although Graves’s characterization of the eighth stage is admittedly uncertain, it is nonetheless noteworthy that it does accommodate relativism, yet does not value skepticism. Actually, value skepticism seems to be an escape route in face of a failure to supersede the kaleidoscopic display brought about with modern pluralism and relativism. In any event,

Graves's constructs are themselves a source of distinct behavioral expectations resting on the different intrinsic values individuals place on actions and outcomes, particularly outcomes reached by others. As such, they can provide a basis for advancing testable hypotheses about participants' behavioral responses in face of intersubjective, morally relevant conflicts of action under variable institutional conditions. Since the on-line assessment tool of Graves's constructs (Hurlbut, 1979) allows fairly quick surveys and supplies quantitative data, the hypothesis testing can utilize standard or sophisticated statistical procedures for checking the strength of the relationships between the stages and the behavioral responses of the participants.

Table 2. Cyclical aspect, way of thinking and themes of the selected Gravesian stages or waves of interior development

Stage or wave	Cyclical aspect	Thinking	Basic theme
8 <sup>th</sup>	Sacrifice-self (communion)	Holistic	<i>Adjust to the realities of one's existence and accept the existential dichotomies as they are and go on living.</i>
7 <sup>th</sup>	Express-self (agency)	Ecological	<i>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.</i>
6 <sup>th</sup>	Sacrifice-self (communion)	Consensus	<i>Sacrifice now in order for all to get now.</i>
5 <sup>th</sup>	Express-self (agency)	Strategic	<i>Express self for what self desires, but in a fashion calculated not to bring down the wrath of others.</i>
4 <sup>th</sup>	Sacrifice-self (communion)	Authority	<i>Sacrifice self now to receive reward later.</i>
3 <sup>rd</sup>	Express-self (agency)	Egocentric	<i>Express self, to hell with others and the consequences, lest one suffer the torment of unbearable shame.</i>
2 <sup>nd</sup>	Sacrifice-self (communion)	Animistic	<i>Sacrifice self to the way of your elders.</i>
1 <sup>st</sup>	Express-self (agency)	Instinctive	<i>Express self as just another animal according to the dictates of one's psychological needs and the environmental possibilities.</i>

Source: Author's configuration based on Graves (2005) and Beck and Cowan (1996)

In what follows, I illustrate the conceptual scheme of the proposed integration of the constructivist developmental point of view into the IAD framework. In the figures ahead, I take on a pictographic representation of Graves's model, based on Beck and Cowan (1996),

to stand for the whole set of connection built in Table 1, including the far-reaching epistemological discussion from which its plausibility is derived.

In Figure 1, the suggestion is that the developmental perspective can be used to access the subjective and intersubjective aspects of both the individual actors and their communities, and that these interior structures of perspectives and sociocognitive inventories affect the patterns of interactions, as Habermas's thoroughly action-theoretic account manifestly indicates.

In Figure 2, the relevant developmental models join the IAD framework to examine the influence of variables affecting institutional choice. The suggestion here is that the different but ordered structures of perspective taking, behavioral expectations, concepts of authority, concepts of motivation, social perspectives, concepts of justice, value-based behavioral systems that characterize different stages of interaction (Table 1) inform how the subjective benefit-cost analysis proceeds. Fundamentally, it is the plurality of subjective perspectives (in the center of Figure 2) that tells how the individuals evaluate the appeal of the institutional change. This, in view of the relaxation of the selfishness axiom, already admitted in the IAD framework and contemporary evolutionary psychology, the distinct character of this proposal is only to suggest that the family of intrinsic values and the interior dispositions of the participants in fact present an emergent stage developmental structure.

In Figure 3, I suggest that the internal drives of human development impose restrictions on how the individuals are able to revise their "mental models" of the action situation. At the same time, the epistemological critique leading to the nonfoundationalist, discursive model of justification (Habermas) and deflationary theory of truth (Heath) diminishes the relevance of the correspondence theory of knowledge implicated in the ordinary conception of mental models, thus broadening significantly the meaning of the scheme represented in Figure 3. This broadening has a great bearing on the interpretation of the role of communication, which would allow the IAD framework to move beyond the instrumental conception of rationality and open itself to the hermeneutic dimension of communicative action.

Information about the psychosocial value systems of the individuals thus helps to predict the likelihood of a group of participants reaching mutual understanding in devising sustainable solutions to their collective action problems through communication. Figure 3 also indicates that the vividness and salience of the information (Frohlich and Oppenheimer, 2001) being shared depends on the subjective perception and evaluation of the situation under discussion. Again, it is worth mentioning that the examination of the relationship between the perceptual readiness and the hypothesized levels of human existence was one of the first lines of investigation Graves pursued to check the behavioral characteristics corresponding to his constructs.

In short, these are the key contributions that the emergent stage conception can offer to refine the analysis and development of better institutions and their incentive structures to deal with collective action problems.

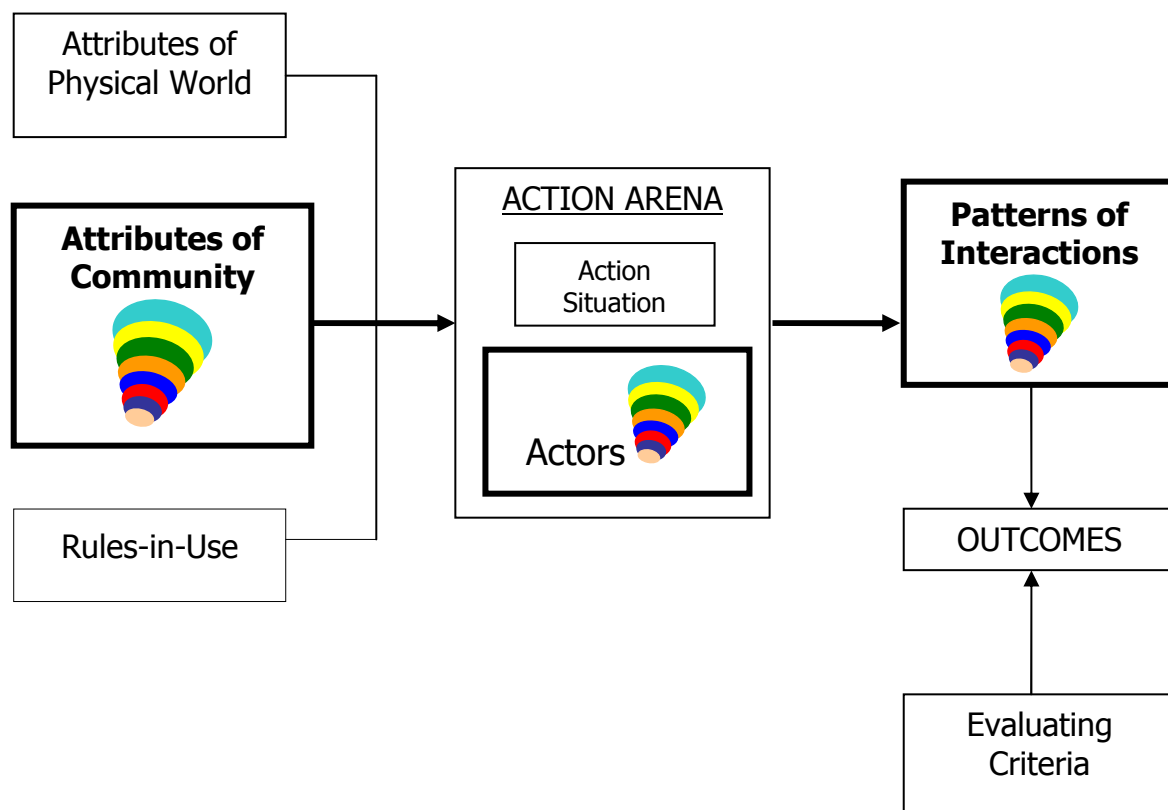


Figure 1 The constructivist stage-developmental perspective meets the IAD framework.  
*Source:* Adapted from Ostrom, Gardner and Walker (2002, p. 37)

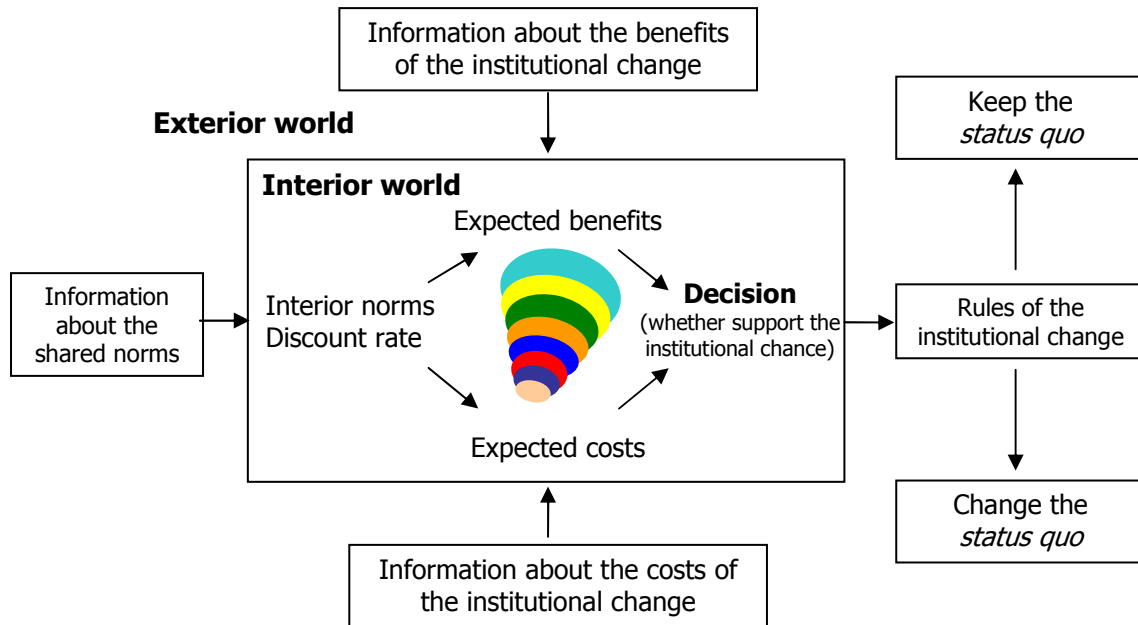


Figure 2. Stages of psychosocial development as weights and restrictions to expectations, norms, time horizon and choice concerning institutional change. Source: Author's configuration, adapted from Ostrom (1997, p. 193)

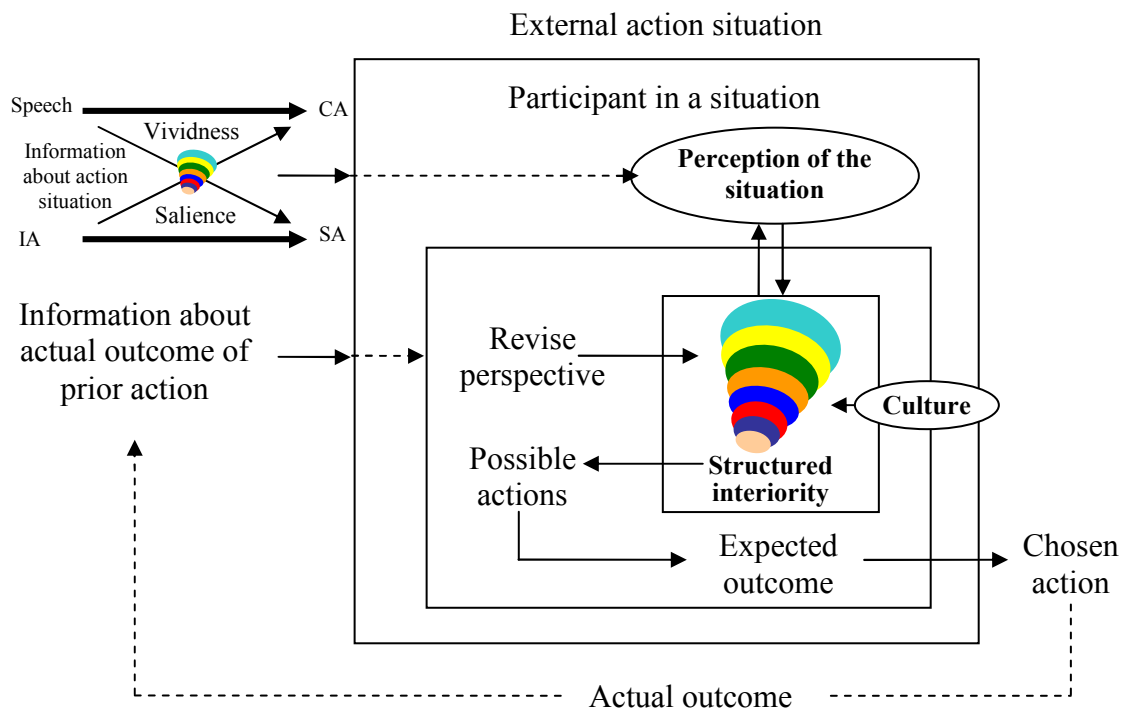


Figure 3. Internal drives and stages affect both the revision of perspectives re the action situation and participants' attitudes toward communication opportunities. Source: Author's configuration, adapted form Ostrom (2005), based on Denzau and North (2000)

## References

- AGRAWAL, A. (2001). Common Property Institutions and Sustainable Governance of Resources. *World Development*, v. 29, p. 1649-1672
- ARLIN, P.K. (1975). Cognitive development in adulthood: a fifth stage? *Developmental Psychology* 11, 602-606
- BALAND, J.M.; PLATTEAU, J.P. 1996. *Halting degradation of natural resources: is there a role for rural communities?* Oxford: Clarendon Press, 1996.
- BASSECHES, M. (1980). Dialectical schemata: a framework for the empirical study of the development of dialectical thinking. *Human Development*, 23, 400-421.
- BECK, E.D.; COWAN, C.C. (1996). *Spiral Dynamics. Mastering Values, Leadership and Change. Exploring the New Science of Memetics*. London : Blackwell Publishers.
- COMMONS, M.L. (1981). 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, pp. 257-291.
- COMMONS, M.L.; RICHARDS, F.A. (1984). 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. v. 1, p. 120-140.
- COMMONS, SINNOTT, RICHARDS and ARMON 1989;
- COWAN, C.; TODOROVIC, N.( 2005). *Spiral Dynamics I: Certification*, NVCC, Santa Barbara.
- 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.
- FLAVELL, J.H., et al (1968). *The Development of Role-Taking and Communication Skills in Children*. New York.
- FROHLICH, N.; OPPENHEIMER, (2001). J.A. Choosing: a cognitive model of economic and political choice. Winnipeg: University of Manitoba, Faculty of Management. (Working Paper).
- GRAVES, C.W. (1970). Levels of Existence: An Open System Theory of Values. *Journal of Humanistic Psychology*, Fall 1970, Vol. 10, No. 2, pp. 131-155.
- GRAVES, C.W. (1971). *A systems conception of personality: remarks by Clare W. Graves on his levels of existence theory*. Santa Barbara, CA: ECLET Publishing (Presented at the Washington School of Psychiatry, Washington, DC).
- GRAVES, C.W. (1981). *Summary statement: the emergent cyclical, double-helix model of the adult human biopsychosocial systems*. Boston, 1981.
- GRAVES, C.W. (2005). "The Never Ending Quest: A Treatise on an Emergent Cyclical Conception of Adult Behavioral Systems and Their Development." Edited and compiled by Christopher C. Cowan and Natasha Todorovic. Santa Barbara, CA: ECLET Publishing.
- HABERMAS, J. (1996). *Between Facts and Norms*. Trans. W. Rehg. Cambridge, MA: MIT Press.
- HABERMAS, J. (1993). *Justification and Application: Remarks on Discourse Ethics*; translated by Ciaran Cronin. MIT Press, Cambridge, Massachusetts, and London, England.
- HABERMAS, J. *Moral consciousness and communicative action*. Cambridge, MA: MIT Press, 1990.
- 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.
- HACKETT, S., SCHLAGER, E.; WALKER, J.M. (1994). The role of communication in resolving commons dilemmas: experimental evidence with heterogeneous appropriators. *Journal of Environmental Economics and Management*, 27(2), 99-126.

- HARDIN, G. (1968). The tragedy of the commons. *Science*. 162:1243–1248.
- HARSANYI, J.C.; SELTEN, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, Mass.: MIT Press.
- HEATH, J. (2001). *Communicative action and rational choice*. Cambridge, MA: MIT Press;
- HEATH, J. (2008). *Following the rules: practical reasoning and deontic constraints*. New York, NY: Oxford University Press.
- HURLBUT, M. A. (1979). *Clare Graves's Levels of Psychological Existence: A Test Design*. North Texas State University, Denton, Texas. PhD dissertation.
- INHELDER, B.; PIAGET, J. (1955). De la logique de l'enfant à la logique de l'adolescent. Paris: Presses Universitaires de France. (English version: The growth of logical thinking from childhood to adolescence. London: Routledge, 1958)
- ISAAC, R.M.; WALKER, J.; WILLIAMS, A.W. (1994). 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.
- KOESTLER, A. (1973). *The ghost in the machine*. New York: Random House.
- KOHLBERG L. (1969). Stage and Sequence: The Cognitive-Developmental Approach to Socialization, in D. A. Goshn, ed., *Handbook of Socialization Theory and Research*, New York: Rand McNally, pp. 347-480.
- KOHLBERG L. (1976). Moral stages and moralization: the cognitive developmental approach. In: LICKONA, T. (Ed.). *Moral development behavior*. New York: Holt, Rinehart and Winston.
- KOHLBERG, L. (1981). *Essays on Moral Development*. San Francisco, v. 1, pp. 426-428.
- KOLLOCK, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 24, 183–214.
- KOMORITA, S. S., & PARKS, C. D. (1995). Interpersonal relations: Mixed motive interaction. *Annual Review of Psychology*, 46, 183–207.
- KOPELMAN, S., WEBER, J. M., & MESSICK, D. M. (2002). Factors influencing cooperation in commons dilemmas: A review of experimental psychological research. In E. Ostrom, T. Dietz, N. Dolsak, P. C.
- KRAMER, D.A. (1983). Postformal operations? A need for further conceptualization. *Human Development*, 26, 91-105.
- LEPYARD, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.
- MANTZAVINOS, C., NORTH, D.C. and SHARIQ, S. (2004). Learning, Institutions, and Economic Performance. *Perspectives on Politics*, vol. 2, nr. 1.
- MARCHAND, H. Some reflections on postFormal thought. *The Genetic Epistemologist*, v. 29, n. 3, 2005.
- McCARTHY, (1990). Introduction. In. Habermas, J. *Moral consciousness and communicative action*. Cambridge, MA: MIT Press, 1990.
- McKEAN, M. A. (1992). Management of Traditional Common Lands (Iriaichi) in Japan. In *Making the Commons Work: Theory, Practice, and Policy*, ed. Daniel W. Bromley, 63-98. San Francisco: Institute of Contemporary Studies.
- MEYER, L.F.F.; BRAGA, M.J. (2009). Fear or greed? Duty or solidarity? Motivations and stages of moral reasoning: experimental evidences from public-goods provision dilemmas. Working paper at the Fourth Workshop on the Workshop in Political Theory and Policy Analysis (WOW4). Indiana University, Digital Library of the Commons.
- NASH, J.F. (1950). Equilibrium Points of n-Person Games. *Proceedings of the National Academy of Science* 36: 48-49.
- NASH, J.F. (1951). Non-Cooperative Games. *Annals of Mathematics* 54: 286-95.
- OSTROM, E. (2003). Toward a behavioral theory linking trust, reciprocity, and reputation. In: OSTROM, E.; WALKER, J. (Eds.). *Trust & reciprocity: interdisciplinary lessons from experimental research*. 19-79.
- OSTROM, E. 1990. *Governing the commons*. Cambridge: Cambridge University Press.

- OSTROM, E. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- OSTROM, E. (2007). A diagnostic approach for going beyond panaceas. *National Academy of Sciences*,.
- OSTROM, E.; GARDNER, R.; WALKER, J.M. 1994. *Rules, games, and common-pool resources*. University of Michigan Press, Ann Arbor, MI.
- PASCUAL-LEONE, J. (1984). Attentional, dialectical, and mental effort: toward an organismic theory of life stages. In M. L. Commons, F. A. Richards & C. Armon (Eds.), *Beyond formal operations* (pp. 182-216). New York: Praeger.
- PIAGET, J. (1970/1972). Intellectual evolution from adolescence to adulthood. *Human Development*, v. 15, p. 1-12.
- POPPE, M. (2005). The specificity of social dilemmas situations. *Journal of Economic Psychology*, v. 26, p. 431-441.
- RICHARDS, F.; COMMONS, M. (1990). 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.
- RIEGEL, K. (1973). Dialectic operations. The final period of cognitive development. *Human Development*, 16, 346-370.
- SCHLAGER, E. (1990). Model specification and policy analysis: the governance of coastal fisheries. [PhD Thesis] Indiana University.
- SINNOTT, J. D. (1984). Postformal reasoning: The relativistic stage. In M. L. Commons, F. A. Richards & C. Armon (Eds.), *Beyond formal operations* (pp. 298-325). New York: Praeger.
- SINNOTT, J. D. (1989) Adult differences in use of postformal operations. In M. L. Commons, J. D. Sinnott, F. A. Richards & C. Armon (Eds.), *Beyond formal operations II* (pp. 239-278). Westport: Praeger.
- TANG, S. Y. (1994). "Institutions and Performance in Irrigation Systems." In *Rules, Games, and Common-Pool Resources*, ed. Elinor Ostrom, Roy Gardner, and James Walker, 225-45. Ann Arbor: University of Michigan Press.
- VAN LANGE, P. A. M., LIEBRAND, W. B. G., MESSICK, D. M., & WILKE, H. A. M. (1992). Introduction and literature review. In W. B. G. Liebrand, D. M. Messick, & H. A. M. Wilke (Eds.), *Social dilemmas* (pp. 3-28). Oxford: Pergamon Press
- WADE, J. (1996). *Changes of Mind*. Suny Series in the Philosophy of Psychology. Albany, State University of New York Press
- WILBER, K. (2000). *Integral Psychology: Consciousness, Spirit, Psychology, Therapy*. Boston, Mass.: Shambhala.
- WILBER, K. (2001). *A Theory of Everything*. Boston, Mass.: Shambhala.