The discourse theory of ethics, for which I have proposed a program of philosophical justification,¹ is not a self-contained endeavor. Discourse ethics advances universalistic and thus very strong theses, but the status it claims for those theses is relatively weak. Essentially, the justification involves two steps. First, a principle of universalization (U) is introduced. It serves as a rule of argumentation in practical discourses. Second, this rule is justified in terms of the substance of the pragmatic presuppositions of argumentation as such in connection with an explication of the meaning of normative claims to validity. The universalization principle can be understood on the model of Rawls's “reflective equilibrium” as a reconstruction of the everyday intuitions underlying the impartial judgment of moral conflicts of action. The second step, which is designed to set forth the universal validity of (U), a validity that extends beyond the perspective of a particular culture, is based on a transcendental-pragmatic demonstration of universal and necessary presuppositions of argumentation. We may no longer burden these arguments with the status of an a priori transcendental deduction along the lines of Kant’s critique of reason. They ground only the fact that there is no identifiable alternative to our kind of argumentation. In this respect, discourse ethics, like other reconstructive sciences, relies solely on hypothetical reconstruction for which plausible confirmation must be sought.² Initially, of course, the place to look for such confirmation is at the level on which discourse ethics competes with other moral theories. But a theory of this kind is also open to, indeed dependent upon, indirect validation by other theories that are consonant with it. I view the theory of the development of moral consciousness advanced by Lawrence Kohlberg and his coworkers as an example of the latter kind of validation.³

Kohlberg holds that the development of the capacity for moral judgment from childhood to adolescence and adult life follows an invariant pattern. The normative reference point of the developmental path that Kohlberg empirically analyzes is a principled morality in which we can recognize the main features of discourse ethics. From the standpoint of ethics, the consonance between psychological theory and normative theory consists in this case in the following: Opponents of universalistic ethics generally bring up the fact that different cultures have different conceptions of morality. To oppose relativistic objections of this kind, Kohlberg’s theory of moral development offers the possibility of (a) reducing the empirical diversity of existing moral views to variation in the contents, in contrast to the universal forms, of moral judgment and (b) explaining the remaining structural differences between moralities as differences in the stage of development of the capacity for moral judgment.

Yet the consonance in the results of these two theories would seem to be deprived of its significance by the internal links between them. Kohlberg’s theory of moral development makes use of the insights of philosophical ethics in the description of the cognitive structures underlying principle moral judgment. In making a normative theory like that of Rawls an integral part of an empirical theory, the psychologist at the same time subjects it to indirect testing. Empirical corroboration of the assumptions derived from developmental psychology extends, then, to all components of the theory from which the confirmed hypotheses were derived. This means we must give precedence to the moral theory that survives such a test better than others. I consider reservations about the circular character of this verification process to be unfounded.

¹ I would like to thank Max Miller and Gertrud Nunner-Winkler for their critical comments on a draft of this essay.
The empirical corroboration of an empirical theory that presupposes as valid the fundamental assumptions of a normative theory cannot, I admit, pass for an independent corroboration of the normative theory. But independence postulates have been shown to be too strong in several respects. For example, the data used to test an empirical theory cannot be described independently of the language of the same theory. Similarly, two competing empirical theories cannot be evaluated independently of the paradigms furnishing their basic concepts. On the meta- or intertheoretical level, the only governing principle is that of coherence. We want to find out what elements fit together, which is a bit like doing a jigsaw puzzle. The reconstructive sciences designed to grasp universal competences break through the hermeneutic circle in which the Geisteswissenschaften, as well as the interpretive social sciences, are trapped. But the hermeneutic circle closes on the meta-theoretical level even for a genetic structuralism, that, like the theories of moral development derived from Piaget, attempts to deal with problems posed in an ambitious universalist form. In such a case, looking for independent proof is a waste of time. It is only a question of seeing whether the descriptions produced with the aid of several theoretical spotlights can be integrated into a relatively reliable map.

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, philosophy and science must change their self-perceptions. This division of labor is no more compatible with the claim to exclusivity previously raised by the program of unified science on behalf of the standard form of the nomological empirical sciences than it is with transcendental philosophy and its foundationalist aim of ultimate justification. Once transcendental arguments have been disengaged from the language game of the philosophy of reflection and reformulated along Strawsonian lines, recourse to the synthetic functions of self-consciousness loses its plausibility. The objective of transcendental deduction loses its meaning, and the hierarchical relation claimed to exist between the a priori knowledge of foundations and the a posteriori knowl-

edge of phenomena loses its basis. A reflective grasp of what Kant had captured with the image of the subject's constitutive achievements, or as we would say today, the reconstruction of the general and necessary presuppositions under which subjects capable of speech and action reach understanding about something in the world—such striving for knowledge on the part of the philosopher is no less fallible than anything else that has ever been exposed to the grueling and cleansing process of scientific discussion and has stood up, at least for the time being.

A nonfoundationalist self-understanding of this kind does more, however, than simply relieve philosophy of tasks that have overburdened it. It not only takes something away from philosophy; it also provides it with the opportunity for a certain naiveté and a new self-confidence in its cooperative relationship with the reconstructive sciences. A relationship of mutual dependence becomes established. Thus, to return to the matter at hand, not only does moral philosophy depend on indirect confirmation from a developmental psychology of moral consciousness; the latter in turn is built on philosophical assumptions. I will investigate this interdependence by using Kohlberg as an example.

I The Fundamental Philosophical Assumptions of Kohlberg's Theory

Coming from the tradition of American pragmatism, Lawrence Kohlberg has a clear conception of the philosophical bases of his theory. His philosophical views on the "nature of moral judgment" were originally inspired by G. H. Mead. But since the publication of A Theory of Justice (1971), he has used John Rawls's ethics, which takes its bearings from Kant and modern natural-law theory, to sharpen them: "These analyses point to the features of a 'moral point of view,' suggesting truly moral reasoning involves features such as impartiality, universalizability, reversibility and prescriptivity." The premises Kohlberg borrows from philosophy can be grouped under three headings: cognitivism, universalism, and formalism.
Moral Consciousness and Communicative Action

In section 1 below I will elaborate on why I think discourse ethics is better suited than any other ethical theory to explain the moral point of view in terms of premises of cognitivism, universalism, and formalism. In section 2 I want to show that the concept of constructive learning used by Piaget and Kohlberg is necessary to discourse ethics as well. Discourse ethics is thus well suited to the description of cognitive structures that emerge from learning processes. Section 3 shows how discourse ethics can complement Kohlberg's theory by virtue of its connection with a theory of communicative action. In the sections that follow I will use this internal relation to establish a plausible basis for a vertical reconstruction of the developmental stages of moral judgment.

All cognitivist moral theories in the tradition of Kant take into account the three aspects Kohlberg focuses on to explain the idea of the moral. The advantage of the position Apel and I defend is that the basic cognitivist, universalist, and formalist assumptions can be derived from the moral principle grounded in discourse ethics. I have previously proposed the following formulation of this principle:

(U) For a norm to be valid, the consequences and side effects that its general observance can be expected to have for the satisfaction of the particular interests of each person affected must be such that all affected can accept them freely.

Cognitivism Since the universalization principle is a rule of argumentation enabling us to reach consensus on generalizable content, the justification of (U) demonstrates at the same time that moral-practical issues can be decided on the basis of reasons. Moral judgments have cognitive content. They represent more than expressions of the contingent emotions, preferences, and decisions of a speaker or actor. Discourse ethics refutes ethical skepticism by explaining how moral judgments can be justified. Any developmental theory of the capacity for moral judgment must presuppose this possibility of distinguishing between right and wrong moral judgments.

Universalism It follows directly from (U) that anyone who takes part in argumentation of any sort is in principle able to reach the same judgments on the acceptability of norms of action. By justifying (U), discourse ethics rejects the basic assumptions of ethical relativism, which holds that the validity of moral judgments is measured solely by the standards of rationality or value proper to a specific culture or form of life. A theory of moral development that attempts to outline a general path of development would be doomed to failure from the start if moral judgments could not claim universal validity.

Formalism (U) works like a rule that eliminates as nongenerizable content all those concrete value orientations with which particular biographies or forms of life are permeated. Of the evaluative issues of the good life it thus retains only issues of justice, which are normative in the strict sense. They alone can be settled by rational argument. With its justification of (U), discourse ethics sets itself in opposition to the fundamental assumptions of material ethics. The latter is oriented to issues of happiness and tends to ontologically favor some particular type of ethical life or other. By defining the sphere of the normative validity of action norms, discourse ethics sets the domain of moral validity off from the domain of cultural value contents.

Yet this does not exhaust the contents of discourse ethics. While its universalization principle furnishes a rule of argumentation, the principle of discourse ethics (D) expresses the fundamental idea of moral theory that Kohlberg borrowed from G. H. Mead's communication theory as the notion of "ideal role taking." This principle postulates the following:

(D) Every valid norm would meet with the approval of all concerned if they could take part in a practical discourse.
Discourse ethics does not set up substantive orientations. Instead, it establishes a procedure based on presuppositions and designed to guarantee the impartiality of the process of judging. Practical discourse is a procedure for testing the validity of hypothetical norms, not for producing justified norms. It is this proceduralism that sets discourse ethics apart from other cognitivist, universalist, and formalist ethical theories, and thus from Rawls's theory of justice as well. (D) makes us aware that (U) merely expresses the normative content of a procedure of discursive will formation and must thus be strictly distinguished from the substantive content of argumentation. Any content, no matter how fundamental the action norms in question may be, must be made subject to real discourse (or advocatory discourses undertaken in their place). The principle of discourse ethics prohibits singling out with philosophical authority any specific normative contents (as, for example, certain principles of distributive justice) as the definitive content of moral theory. Once a normative theory like Rawls's theory of justice strays into substantive issues, it becomes just one contribution to practical discourse among many, even though it may be an especially competent one. It no longer helps to ground the moral point of view that characterizes practical discourses as such.

The fundamental assumptions of cognitivism, universalism, and formalism discussed above are already contained within a procedural definition of the moral. Such a definition also allows for a sufficiently sharp demarcation within moral judgment between cognitive structures and contents. The discursive procedure, in fact, reflects the very operations Kohlberg postulates for moral judgments at the postconventional level: complete reversibility of the perspectives from which participants produce their arguments; universality, understood as the inclusion of all concerned; and the reciprocity of equal recognition of the claims of each participant by all others.

Discourse ethics singles out (U) and (D) as characteristics of moral judgment that can serve as normative points of reference in describing the development of the capacity for moral judgment. Kohlberg distinguishes six stages of moral judgment. They can be regarded as gradual approximations in the dimensions of reversibility, universality, and reciprocity to structures of impartial or just judgments about morally relevant conflicts of action. Below I quote Kohlberg's summary of his moral stages.

**Level A, preconventional level**

*Stage 1*, the stage of punishment and obedience

**Content:** Right is literal obedience to rules and authority, avoiding punishment, and not doing physical harm.

1. What is right is to avoid breaking rules, to obey for obedience' sake, and to avoid doing physical damage to people and property.
2. The reasons for doing right are avoidance of punishment and the superior power of authorities.

*Stage 2*, the stage of individual instrumental purpose and exchange

1. What is right is following rules when it is to someone's immediate interest. Right is acting to meet one's own interests and needs and letting others do the same. Right is also what is fair; that is, what is an equal exchange, a deal, an agreement.
2. The reason for doing right is to serve one's own needs or interests in a world where one must recognize that other people have their interests, too.

**Level B, conventional level**

*Stage 3*, the stage of mutual interpersonal expectations, relationships, and conformity

**Content:** The right is playing a good (nice) role, being concerned about the other people and their feelings, keeping loyalty and trust with partners, and being motivated to follow rules and expectations.

1. What is right is living up to what is expected by people close to one or what people generally expect of people in one's role as son, sister, friends, and so on. "Being good" is important and means having good motives, showing concern about others. It also means keeping mutual relationships, maintaining trust, loyalty, respect, and gratitude.
2. Reasons for doing right are being to serve one's own needs and those of others, caring for others, and because if one puts oneself in the other person's place one would want good behavior from the self (Golden Rule).
Stage 4, the stage of social system and conscience maintenance
Content: The right is doing one's duty in society, upholding the social order, and maintaining the welfare of society or the group.

1. What is right is fulfilling the actual duties to which one has agreed.
Laws are to be upheld except in extreme cases where they conflict with other fixed social duties and rights. Right is also contributing to society, the group, or institution.

2. The reasons for doing right are to keep the institution going as a whole, self-respect or conscience as meeting one's defined obligations, or the consequences: "What if everyone did it?"

Level C, postconventional and principled level
Moral decisions are generated from rights, values or principles that are (or could be) agreeable to all individuals composing or creating a society designed to have fair and beneficial practices.

Stage 5, the stage of prior rights and social contract or utility
Content: The right is upholding the basic rights, values, and legal contracts of a society, even when they conflict with the concrete rules and laws of the group.

1. What is right is being aware of the fact that people hold a variety of values and opinions, that most values and rules are relative to one's group. These "relative" rules should usually be upheld, however, in the interest of the impartiality and because they are the social contract. Some nonrelative values and rights such as life, and liberty, however, must be upheld in any society and regardless of majority opinion.

2. Reasons for doing right are, in general, feeling obligated to obey the law because one has made a social contract to make and abide by laws, for the good of all and to protect their own rights and the rights of others. Family, friendship, trust, and work obligations are also commitments or contracts freely entered into and entail respect for the rights of others. One is concerned that laws and duties be based on rational calculation of overall utility: "the greatest good for the greatest number."

Stage 6, the stage of universal ethical principles
Content: This stage assumes guidance by universal ethical principles that all humanity should follow.

1. Regarding what is right, Stage 6 is guided by universal ethical principles. Particular laws or social agreements are usually valid because they rest on such principles. When laws violate these principles, one acts in accordance with the principle. Principles are universal principles of justice: the equality of human rights and respect for the dignity of human beings as individuals. These are not merely values that are recognized, but are also principles used to generate particular decisions.

Kohlberg conceives the transition from one stage to the next as learning. Moral development means that a child or adolescent rebuilds and differentiates the cognitive structures he already has so as to be better able to solve the same sort of problems he faced before, namely, how to solve relevant moral dilemmas in a consensual manner. The young person himself sees this moral development as a learning process in that at the higher stage he must be able to explain whether and in what way the moral judgments he had considered right at the previous stage were wrong. Kohlberg interprets this learning process as a constructive achievement on the part of the learner, as would Piaget. The cognitive structures underlying the capacity of moral judgment are to be explained neither primarily in terms of environmental influences nor in terms of inborn programs and maturational processes. They are viewed instead as outcomes of a creative reorganization of an existing cognitive inventory that is inadequate to the task of handling certain persistent problems.

Discourse ethics is compatible with this constructivist notion of learning in that it conceives discursive will formation (and argumentation in general) as a reflective form of communicative action and also in that it postulates a change of attitude for the transition from action to discourse. A child growing up, and caught up, in the communicative practice of everyday life is not able at the start to effect this attitude change.

In argumentation, claims to validity that heretofore served actors as unquestioned points of orientation in their everyday communication are thematized and made problematic. When this happens, the participants in argumentation adopt a hypothetical attitude to controversial validity claims. The validity of a contested norm is put in abeyance when practical discourse begins. The issue is then whether or not the norm deserves to be recognized, and that issue will be decided by a contest between proponents and opponents of the norm. The attitude
change accompanying the passage from communicative action to discourse is no different for issues of justice than for issues of truth. In the latter case, what had previously been considered facts in naive dealings with things and events must now be regarded as something that may or may not be the case. Just as "facts" are thus transformed into "states of affairs," so social norms to which one is accustomed are transformed into possibilities for regulation that can be accepted as valid or rejected as invalid.

If by way of a thought experiment we compress the adolescent phase of growth into a single critical instant in which the individual for the first time—yet perversely and intransigently—assumes a hypothetical attitude toward the normative context of his lifeworld, we can see the nature of the problem that every person must deal with in passing from the conventional to the postconventional level of moral judgment. The social world of legitimately regulated interpersonal relations, a world to which one was naively habituated and which was unproblematically accepted, is abruptly deprived of its quasi-natural validity.

If the adolescent cannot and does not want to go back to the traditionalism and unquestioned identity of his past world, he must, on penalty of utter disorientation, reconstruct, at the level of basic concepts, the normative orders that his hypothetical gaze has destroyed by removing the veil of illusions from them. Using the rubble of devalued traditions, traditions that have been recognized to be merely conventional and in need of justification, he erects a new normative structure that must be solid enough to withstand critical inspection by someone who will henceforth distinguish soberly between socially accepted norms and valid norms, between de facto recognition of norms and norms that are worthy of recognition. At first principles inform his plan for reconstruction; these principles govern the generation of valid norms. Ultimately all that remains is a procedure for a rationally motivated choice among principles that have been recognized in turn as in need of justification. In contrast to moral action in everyday life, the shift in attitude that discourse ethics requires for the procedure it singles out as crucial, the transition to argumentation, has something unnatural about it: it marks a break with the ingenious straightforwardness with which people have raised the claims to validity on whose intersubjective recognition the communicative practice of everyday life depends. This unnaturalness is like an echo of the developmental catastrophe that historically once devalued the world of traditions and thereby provoked efforts to rebuild it at a higher level. In this sense, what Kohlberg conceives as a constructive learning process operating at all levels is built into the transition (which has become routine for the adult) from norm-guided action to norm-testing discourse.

Having dealt with the normative reference point of moral development in Kohlberg’s theory in section 1 and with his concept of learning in section 2, I will now turn to an analysis of his stage model. Here too Kohlberg follows Piaget by setting up a model of the developmental stages of a specific competence, in this case the capacity for moral judgment. Kohlberg describes this model in terms of three strong hypotheses:

- The stages of moral judgment form an invariant, irreversible, and consecutive sequence of discrete structures. This assumption precludes the possibility that different experimental subjects will reach the same goal by different developmental paths, that the same subject will regress from a higher to a lower stage, and that stages will be skipped in the course of a subject’s development.

- The stages of moral judgment form a hierarchy in that the cognitive structures of a higher stage dialectically sublate those of the lower one, that is, the lower stage is replaced and at the same time preserved in a reorganized, more differentiated form.

- Every stage of moral judgment can be characterized as a structured whole. This assumption precludes the possibility that at a given point in time an experimental subject will have to judge different moral content at different levels. Not pre-
concluded are so-called décalage (realignment) phenomena, which indicate a step by step anchoring of newly acquired structures.

The key component of this model is obviously the second hypothesis. The other two hypotheses can be toned down or modified, but the notion of a path of development which can be described in terms of a hierarchically ordered sequence of structures is absolutely crucial to Kohlberg’s model of developmental stages. For Kohlberg as for Piaget, synonymous concepts. Kohlberg justifies the developmental logic of his six stages of moral judgment by correlating them with corresponding sociomoral perspectives. I quote his summary of these perspectives below:

1. This stage takes an egocentric point of view. A person at this stage doesn’t consider the interests of others or recognize they differ from actor’s, and doesn’t relate two points of view. Actions are judged in terms of physical consequences rather than in terms of psychological interests of others. Authority’s perspective is confused with one’s own.

2. This stage takes a concrete individualistic perspective. A person at this stage separates own interests and points of view from those of authorities and others. He or she is aware everybody has individual interests to pursue and these conflict, so that right is relative (in the concrete individualistic sense). The person integrates or relate conflicting individual interests to one another through instrumental exchange of services, through instrumental need for the other and the other’s goodwill, or through fairness giving each person the same amount.

3. This stage takes the perspective of the individual in relationship to other individuals. A person at this stage is aware of shared feelings, agreements, and expectations, which take primacy over individual interests. The person relates points of view through the “concrete Golden Rule,” putting oneself in the other person’s shoes. He or she does not consider generalized “system” perspective.

4. This stage differentiates societal point of view from interpersonal agreement or motives. A person at this stage takes the viewpoint of the system, which defines roles and rules. He or she considers individual relations in terms of place in the system.

5. This stage takes a prior-to-society perspective—that of a rational individual aware of values and rights prior to social attachments and contracts. The person integrates perspectives by formal mechanisms of agreement, contract, objective impartiality, and due process. He or she considers the moral point of view and the legal point of view, recognizes they conflict, and finds it difficult to integrate them.

6. This stage takes the perspective of a moral point of view from which social arrangements derive or on which they are grounded. The perspective is that of any rational individual recognizing the nature of morality or the basic moral premise of respect for other persons as ends, not means.

Kohlberg describes these sociomoral perspectives in such a way that their correlations with stages of moral judgment seem intuitively correct. This plausibility is achieved, however, at the cost of a description in which the sociocognitive conditions of moral judgment have already been blended with the structures of those judgments. Moreover, the sociocognitive conditions lack sufficient analytical rigor to make it immediately evident why this sequence represents a hierarchy in the sense of a logic of development. Perhaps these reservations can be cleared up if we replace Kohlberg’s sociomoral perspectives with the stages of perspective taking developed by Robert Selman. This is indeed a step in the right direction, but it does not suffice for a justification of moral stages, as I will presently show.

It remains to be demonstrated that Kohlberg’s descriptions of moral stages do in fact satisfy the conditions of a stage model conceived in terms of a logic of development. This is a problem of conceptual analysis. It is my impression that empirical research will not advance our understanding here until we have an interesting and sufficiently precise proposal for a solution of the problem in the form of a hypothetical reconstruction. What follows is an attempt to establish whether discourse ethics can contribute to the solution of this problem.

Discourse ethics uses transcendental arguments to demonstrate that certain conditions are unavoidable. Such arguments are geared to convincing an opponent that he makes performative use of something he expressly denies and thus gets caught up in a performative contradiction. In grounding (U) I am specifically concerned to identify the pragmatic presup-
positions indispensable to any argumentation. Anyone who participates in argumentation has already accepted these substantive normative conditions—there is no alternative to them. Simply by choosing to engage in argumentation, participants are forced to acknowledge this fact. This transcendental-pragmatic demonstration serves to make us aware of the extent of the conditions under which we always already operate when we argue; no one has the option of escaping to alternatives. The absence of alternatives means that those conditions are, in fact, inescapable for us.

This “fact of reason” cannot be deductively grounded, but it can be clarified if we take the further step of conceiving argumentative speech as a special case—in fact, a privileged derivative—of action oriented toward reaching understanding. Only when we return to the level of action theory and conceive discourse as a continuation of communicative action by other means can we understand the true thrust of discourse ethics. The reason we can locate the content of (U) in the communicative presuppositions of argumentation is that argumentation is a reflective form of communicative action and the structures of action oriented toward reaching understanding always already presuppose those very relationships of reciprocity and mutual recognition around which all moral ideas revolve in everyday life no less than in philosophical ethics. Like Kant’s appeal to the “fact of reason,” this thrust of discourse ethics has a naturalistic ring to it, but it is by no means a naturalistic fallacy. Both Kant and the proponents of discourse ethics rely on a type of argument that draws attention to the inescapability of the general presuppositions that always already underlie the communicative practice of everyday life and that cannot be picked or chosen like makes of cars or value postulates. This type of argument is made from the reflective point of view, not from the empiricist attitude of an objectivating observer.

The transcendental mode of justification reflects the fact that practical discourse is embedded in contexts of communicative action. To that extent discourse ethics points to, and itself depends upon, a theory of communicative action. We can expect a contribution to the vertical reconstruction of stages of moral consciousness from the theory of communicative action, for the latter focuses on structures of linguistically mediated, norm-governed interaction, structures that integrate what psychology analytically separates, to wit, perspective taking, moral judgment, and action.

Kohlberg saddles sociomoral perspectives with the job of grounding a logic of development. His social perspectives are supposed to express capacities for social cognition. The stages of social perspective do not, however, match Selman’s stages of perspective taking. It may be wise to separate two dimensions that are associated in Kohlberg’s description: the perspective structures themselves and the justice conceptions derived from the sociocognitive inventory at any particular point. There is no need to bring in these normative aspects on the sly, because a moral dimension is intrinsic to the basic concepts of the “social world” and “norm-governed interaction.”

Kohlberg’s construction of social perspectives also seems to be based on the concepts of a conventional role structure. At stage 3 the child learns these role structures in their particularity and generalizes them at stage 4. The axis around which the social perspectives turn, as it were, is the social world as the sum total of the interactions a social group considers legitimate because they are institutionally ordered. At the first two stages the child does not yet have these concepts, and at the last two stages he has reached a standpoint at which he leaves concrete society behind and from which he can test the validity of existing norms. With this transition the fundamental concepts in which the social world was constituted for the young person are transformed directly into basic moral concepts. I would like to trace these links between social cognition and morality, using the theory of communicative action. The attempt to clarify Kohlberg’s social perspectives in this framework should have a number of advantages.

The idea of action oriented toward reaching understanding implies two notions that require clarification: that of the social world and that of norm-guided interaction. The sociomoral perspective developed at stages 3 and 4 and used reflectively at stages 5 and 6 can be set within a system of world perspectives that, in conjunction with a system of speaker perspectives, underlies communicative action. Moreover, the nexus between con-
ceptions of the world and claims to validity opens the possibility of linking the reflective attitude toward the social world (Kohlberg's "prior-to-society perspective") with the hypothetical attitude of a participant in argumentation who thematizes corresponding normative validity claims. This enables us to explain why the moral point of view, conceived in terms of discourse ethics, emerges when the conventional role structure is made reflexive.

This action-theoretic approach suggests that we should understand the development of sociomoral perspectives in the context of the decentering of the young person's understanding of the world. It also draws attention to the structures of interaction themselves, which set the parameters for the constructive learning of basic sociocognitive concepts in children and adolescents. The concept of communicative action is well suited to serve as a point of reference for the reconstruction of stages of interaction. These stages of interaction can be described in terms of the perspective structures implemented in different types of action. To the extent to which these perspectives, embodied and integrated in interactions, fit readily into the scheme of a logic of development, it will be possible to ground stages of moral judgment by tracing Kohlberg's moral stages first to social perspectives and ultimately to stages of interaction. This is my objective in what follows.

I propose to begin by reviewing in part II some of the tenets of the theory of communicative action in order to show that the concept of the social world forms an integral part of a decentered understanding of the world, which in turn forms the basis of action oriented toward reaching understanding. In part III, I will focus on two specific stages of interaction. Studies of perspective taking by Flavell and Selman will serve as my point of departure. The bulk of part III will trace the restructuring of preconventional types of action in two directions: strategic action and normatively regulated action. Part IV deals with a conceptual analysis of how the introduction of a hypothetical attitude into communicative action makes the exacting form of communication called discourse possible, how the moral point of view arises when a reflective stance is taken vis-à-vis the social world, and finally, how stages of moral judgment can be traced back to stages of interaction via social perspectives. This justification of moral stages in terms of a logic of development will have to prove its mettle in subsequent empirical research. For now I will use my reflections only to illuminate some of the anomalies and unresolved problems with which Kohlberg's theory is currently faced (part V).

II The Perspective Structure of Action Oriented toward Reaching Understanding

In section I I review some conceptual aspects of action oriented toward reaching understanding. Section 2 provides an outline of how the related concepts of the social world and normatively regulated action emerge from a decentered understanding of the world.

Since I have given a detailed account of the concept of communicative action elsewhere, I will confine myself at this point to reviewing the most important aspects of this formal-pragmatic study.

Orientation toward success versus orientation toward reaching understanding Social interactions vary in terms of how cooperative and stable or, conversely, how conflictual and unstable they are. The question in social theory of what makes social order possible has a counterpart in action theory: How can (at least two) participants in interaction coordinate their plans in such a way that alter is in a position to link his actions to ego's without a conflict arising, or at least without the risk that the interaction will be broken off? If the actors are interested solely in the success, i.e., the consequences or outcomes of their actions, they will try to reach their objectives by influencing their opponent's definition of the situation, and thus his decisions or motives, through external means by using weapons or goods, threats or enticements. Such actors treat each other strategically. In such cases, coordination of the subjects' actions depends on the extent to which their egocentric utility calculations mesh. The
degree of cooperation and the stability is determined by the interest positions of the participants. By contrast, I speak of communicative action when actors are prepared to harmonize their plans of action through internal means, committing themselves to pursuing their goals only on the condition of an agreement—one that already exists or one to be negotiated—about definitions of the situation and prospective outcomes. In both cases the teleological structure of action is presupposed inasmuch as the actors are assumed to have the ability to act purposively and an interest in carrying out their plans. They differ in that for the model of strategic action, a structural description of action directly oriented toward success is sufficient, whereas the model of action oriented toward reaching understanding must specify the preconditions of an agreement, to be reached communicatively, that allows alter to link his actions to ego’s.\textsuperscript{18}

Reaching agreement as a mechanism for coordinating actions The concept of communicative action is set out in such a way that the acts of reaching understanding that coordinate the action plans of several actors, thus forging a complex of interactions out of goal-directed behavior, cannot in turn be reduced to teleological action.\textsuperscript{19} The kind of agreement that is the goal of efforts to reach understanding depends on rationally motivated approval of the substance of an utterance. Agreement cannot be imposed or brought about by manipulating one’s partner in interaction, for something that patently owes its existence to external pressure cannot even be considered an agreement. The generation of convictions can be analyzed in terms of the model of taking a position on the offer contained in a speech act. Ego’s speech act can be successful only if alter accepts the offer contained in it by taking an affirmative position, however implicitly, on a claim to validity that is in principle criticizable.\textsuperscript{20}

Action situation and speech situation If we define action in general as mastering situations, then the concept of communicative action highlights two aspects of this mastering: the teleological one of implementing an action plan and the communicative one of arriving at a shared interpretation of the situation, or more generally, of reaching consensus. A situation denotes a segment of a lifeworld that has been delimited in terms of a specific theme. A theme arises in connection with the interests and objectives of actors. It defines the range of matters that are relevant and can be thematically focused on. Individual action plans help put a theme in relief and determine the current need for consensual understanding that must be met through the activity of interpretation. In these terms the action situation is at the same time a speech situation in which the actors take turns playing the communicative roles of speaker, addressee, and bystander. To these roles correspond first- and second-person participant perspectives as well as the third-person observer perspective from which the 1-thou relation is observed as an intersubjective complex and can thus be objectified. This system of speaker perspectives is intertwined with a system of world perspectives (see below).

The lifeworld as background Communicative action can be understood as a circular process in which the actor is two things in one: an initiator who masters situations through actions for which he is accountable and a product of the traditions surrounding him, of groups whose cohesion is based on solidarity to which he belongs, and of processes of socialization in which he is reared.

The actor stands face to face with that situationally relevant segment of the lifeworld that impinges on him as a problem, a problem he must resolve through his own efforts. But in another sense, the actor is carried or supported from behind, as it were, by a lifeworld that not only forms the context for the process of reaching understanding but also furnishes resources for it. The shared lifeworld offers a storehouse of unquestioned cultural givens from which those participating in communication draw agreed-upon patterns of interpretation for use in their interpretive efforts.

These ingrained cultural background assumptions are only one component of the lifeworld. The solidarity of groups integrated through values and the competences of socialized individuals also serve as resources for action oriented toward
reaching understanding, although in a different way than cultural traditions. The process of reaching an understanding between world and lifeworld. The lifeworld, then, offers both an intuitively preunderstood context for an action situation and resources for the interpretive process in which participants in communication engage as they strive to meet the need for agreement in the action situation. Yet these participants in communicative action must reach understanding about something in the world if they hope to carry out their action plans on a consensual basis, on the basis of some jointly defined action situation. 

The depiction of facts, however, is only one among several functions of the process of reaching understanding through speech. Speech acts serve not only to represent (or presuppose) states and events—in which case the speaker makes reference to something in the objective world. They also serve to produce (or renew) interpersonal relationships—in which case the speaker makes reference to something in the social world of legitimately ordered interactions. And they serve to express lived experience, that is, they serve the process of self-representation—in which case the speaker makes reference to something in the subjective world to which he has privileged access. It is this reference system of precisely three worlds that communicative actors make the basis of their efforts to reach understanding. Thus, agreement in the communicative practice of everyday life rests simultaneously on intersubjectively shared propositional knowledge, on normative accord, and on mutual trust.

Relations to the world and claims to validity A measure of whether or not participants in communication reach agreement is the yes or no position taken by the hearer whereby he accepts or rejects the claim to validity that has been raised by the speaker. In the attitude oriented toward reaching understanding, the speaker raises with every intelligible utterance the claim that the utterance in question is true (or that the existential presup-

positions of the propositional content hold true), that the speech act is right in terms of a given normative context (or that the normative context that it satisfies is itself legitimate), and that the speaker's manifest intentions are meant in the way they are expressed.

When someone rejects what is offered in an intelligible speech act, he denies the validity of an utterance in at least one of three respects: truth, rightness, or truthfulness. His "no" signals that the utterance has failed to fulfill at least one of its three functions (the representation of states of affairs, the maintenance of an interpersonal relationship, or the manifestation of lived experience) because the utterance is not in accordance with either the world of existing states of affairs, our world of legitimately ordered interpersonal relations, or each participant's own world of subjective lived experience. These aspects are not clearly distinguished in normal everyday communication. Yet in cases of disagreement or persistent problematization, competent speakers can differentiate between the aforementioned three relations to the world, thematizing individual validity claims and focusing on something that confronts them, whether it be something objective, something normative, or something subjective.

World perspectives Having explicating the above structural properties of action oriented toward reaching understanding, we are in a position to identify the options a competent speaker has. If the above analysis is correct, he essentially has the choice between a cognitive, an interactive, and an expressive mode of language use. To these modes correspond three different classes of speech acts—the constative, the regulative, and the representative—which permit the speaker to concentrate, in terms of a universal validity claim, on issues of truth, justice, or taste (i.e., personal expression). In short, he has a choice among three basic attitudes, each entailing a different perspective on the world. In addition, the decentered understanding of the world enables him to confront external nature not only in an objectivating attitude but also in a norm-conformative or an expressive one, to confront society not only in a norm-conformative attitude but also in an objectivating or an expressive one, and to
confront inner nature not only in an expressive attitude but also in an objectivating or a norm-conformative one.

A **decentered understanding of the world** presupposes that relations to the world, claims to validity, and basic attitudes have become differentiated. This process springs from something else in turn, the differentiation between lifeworld and world. Every consciously enacted process of communication recapitulates, as it were, this differentiation, which has been laboriously acquired in the ontogenesis of the capacity for speech and action. The spheres of things about which we can reach a fallible agreement with the lifeworld with its absolute certainties and intuitive presence. As this differentiation progresses, the demarcation becomes ever sharper. On one side we have the horizon of unquestioned, intersubjectively shared, nontematized certitudes that participants in communication have “at their backs.” On the other side, participants in communication face the communicative contents constituted within a world: objects that they perceive and manipulate, norms that they observe or violate, and lived experiences to which they have privileged access and which they can express. To the extent to which participants in communication can conceive of what they reach agreement on as something in a world, something detached from the lifeworld background from which it emerged, what is explicitly known comes to be distinguished from what is implicitly certain. To that extent the content they communicate takes on the character of knowledge linked with a potential for reasoning, knowledge that claims validity and can be criticized, that is, knowledge that can be argued about on the basis of reasons.²²

For the matter under discussion it is important to distinguish between world perspectives and speaker perspectives. On the one hand, participants in communication must have the competence to adopt, when necessary, an objectivating attitude to a given state of affairs, a norm-conformative attitude to legitimately ordered interpersonal relations, and an expressive attitude to their own lived experience; they must also be able to vary these attitudes in relation to each of the three worlds. If, on the other hand, they want to reach a shared understanding about something in the objective, social, or subjective world, they must also be able to take the attitudes connected with the communicative roles of the first, second, or third person.

The decentered understanding of the world is thus characterized by a complex structure of perspectives. It combines two things: first, perspectives that are grounded in the formal three-world reference system and linked with the different attitudes toward the world, and second, perspectives that are built into the speech situation itself and linked to the communicative roles. The grammatical correlates of these world and speaker perspectives are the three basic modes of language use on the one hand and the system of personal pronouns on the other.

A crucial point in my argument is that the development of this complex structure of perspectives also provides us with the key to the desired justification of moral stages in terms of a logic of development. Before discussing the relevant literature in the next two sections, I will explain the basic idea that will govern my analysis.

First, I am convinced that the ontogenesis of speaker and world perspectives that leads to a decentered understanding of the world can be explained only in connection with the development of the corresponding structures of interaction. If, like Piaget, we start from action, i.e., from the active confrontation and interaction between an individual who learns constructively and his environment, it makes sense to assume that the complex system of perspectives sketched above develops from two roots: the observer perspective, acquired by the child as a result of his perceptual-manipulative contact with the physical environment, and the reciprocal I-thou perspectives that the child adopts as a result of symbolically mediated contacts with reference persons (in the framework of interactive socialization processes). The observer perspective is later consolidated as the objectivating attitude toward external nature (or the world of existing states of affairs), while the I-thou perspective is perpetuated in the form of first- and second-person attitudes linked to the communicative roles of speaker and hearer. This
stabilization occurs by virtue of a reshaping and differentiation of the original perspectives. The observer perspective is integrated with the system of world perspectives, whereas the I-thou perspective is rounded out into a system of speaker perspectives. The development of structures of interaction can serve as a guide in reconstructing these processes.

Second, I will argue that the completion of a system of speaker perspectives takes place in two major developmental steps. From a structural point of view, the preconventional stage of interaction can be understood as an implementation, in types of action, of the I-thou perspectives learned through experience in the roles of speaker and hearer. Next, the introduction of an observer perspective into the realm of interaction and its linkage with the I-thou perspective permits a reorganization of action coordination at a higher level. The complete system of speaker perspectives is the result of these two transformations. It is only after the transition to the conventional stage of interaction has been completed that the communicative roles of the first, second, and third persons become fully integrated.

The completion of the system of world perspectives takes a different route. In reconstructing this process we can refer back to the observation that the conventional stage of interaction is characterized by the rise of two new contrasting types of action: strategic action and norm-governed interaction. Owing to the integration of the observer perspective into the sphere of interaction, the child learns to perceive interactions, and his own participation in them, as occurrences in an objective world. This makes possible the development of a purely success-oriented type of action as an extrapolation of conflict behavior guided by self-interest. At the same time that strategic action is being acquired through practice, its opposite, nonstrategic action, comes into view. Once the perception of social interaction is differentiated in this way, the growing child cannot avoid the necessity of also reorganizing types of nonstrategic action that had been left behind in his development, so to speak, to bring them into line with the conventional level. What this means is that a social world of norm-guided interactions open to thematization comes to be set off against the background of the lifeworld.

Third, I will argue that the introduction of an observer perspective into the domain of interactions also provides the impetus for constituting a social world and for judging actions according to whether or not they conform to or violate socially recognized norms. For its members, a social world consists of norms defining which interactions belong to the totality of justified interpersonal relations and which do not. Actors who accept the validity of such a set of norms are members of the same social world. The concept of a social world is linked with the norm-conformative attitude, that is, the perspective from which a speaker relates to accepted norms.23

The basic sociocognitive concepts of the social world and of norm-governed interaction thus evolve within the framework of a decentered understanding of the world, which in turn stems from a differentiation of speaker and world perspectives. These very complex presuppositions of Kohlberg's social perspectives should finally provide the key to deriving stages of moral judgment from stages of interaction.

What follows has the limited purpose of making a plausible case for the foregoing hypotheses about the ontogenesis of speaker and world perspectives, on the basis of existing empirical studies. At best, a hypothetical reconstruction of this kind can serve as a guide to further research. Admittedly, my hypotheses do require distinctions not easy to operationalize, distinctions between (a) communicative roles and speaker perspectives, (b) implementations of these speaker perspectives in different types of interactions, and (c) the perspective structure of an understanding of the world that permits a choice between basic attitudes to the objective, social, and subjective worlds. I am aware of the difficulty that results from the fact that I have to bring these distinctions from the outside to bear on material derived from previous research.

III The Integration of Participant and Observer Perspectives and the Restructuring of Preconventional Types of Action

In section 1 below I will interpret R. Selman's stages of perspective taking in terms of the step-by-step construction of a
system of speaker perspectives that are fully reversible. Section 2 describes four different types of interaction in which I-thou perspectives are embodied. I then go on to assess the significance of the introduction of an observer perspective into the sphere of interaction, with particular reference to the restructur- ing of interest-guided conflict behavior as strategic action. Section 3 will reconstruct the transformation of authority-governed action and interest-governed cooperative behavior into normatively regulated action in order to demonstrate that the complex perspective structure of action oriented to reaching understanding cannot develop in any other way.

Selman characterizes three stages of perspective taking in terms of conceptions of persons and relationships. He summarizes his stages as follows:

**Level 1, differentiated and subjective perspective taking (about ages 5 to 9)**

**Concepts of persons: differentiated** At Level 1, the key conceptual advance is the clear differentiation of physical and psychological characteristics of persons. As a result, intentional and unintentional acts are differentiated and a new awareness is generated that each person has a unique subjective covert psychological life. Thought, opinion, or feeling states within an individual, however, are seen as unitary, not mixed.

**Concepts of relations: subjective** The subjective perspectives of self and other are clearly differentiated and recognized as potentially different. However, another’s subjective state is still thought to be legible by simple physical observation. Relating of perspectives is conceived of in one-way, unilateral terms, in terms of the perspective of and impact on one actor. For example, in this simple one-way conception of relating of perspectives and interpersonal causality, a gift makes someone happy. Where there is any understanding of two-way reciprocity, it is limited to the physical—the hit child hits back. Individuals are seen to respond to action with like action.

**Level 2, self-reflective/second-person and reciprocal perspective taking (about ages 7 to 12)**

**Concepts of persons: self-reflective/second-person** Key conceptual advances at Level 2 are the growing child’s ability to step mentally outside himself or herself and take a self-reflective or second-person perspective on his or her own thoughts and actions and on the realization that others can do so as well. Persons’ thought or feeling states are seen as potentially multiple, for example, curious, frightened, and happy, but still as groupings of mutually isolated and sequential or weighted aspects, for example, mostly curious and happy and a little scared. Both selves and others are thereby understood to be capable of doing things (overt actions) they may not want (intend) to do. And persons are understood to have a dual, layered social orientation: visible appearance, possibly put on for show, and the truer hidden reality.

**Concepts of relations: reciprocal** Differences among perspectives are seen relativistically because of the Level 2 child’s recognition of the uniqueness of each person’s ordered set of values and purposes. A new two-way reciprocity is the hallmark of Level 2 concepts of relations. It is a reciprocity of thoughts and feelings, not merely actions. The child puts himself or herself in another’s shoes and realizes the other will do the same. In strictly mechanical-logical terms, the child now sees the infinite regress possibility of perspective taking (I know that she knows that I know that she knows . . . etc.). The child also recognizes that the outer appearance–inner reality distinction means selves can deceive others as to their inner states, which places accuracy limits on taking another’s inner perspective. In essence, the two-way reciprocity of this level has the practical result of detente, wherein both parties are satisfied, but in relative isolation: one single individuals seeing self and other, but not the relationship system between them.

**Level 3, third-person and mutual perspective taking (about ages 10 to 15)**

**Concepts of persons: third-person** Persons are seen by the young adolescent thinking at Level 3 as systems of attitudes and values fairly consistent over the long haul, as opposed to randomly changeable assortments of states as at Level 2. The critical conceptual advance is toward ability to take a true third-person perspective, to step outside not only one’s own immediate perspective, but outside the self as a system a totality. There are generated notions of what we might call an “observing ego,” such that adolescents do (and perceive other persons to) simultaneously see themselves as both actors and objects, simultaneously acting and reflecting upon the effects of action on themselves, reflecting upon the self in interaction with the self.

**Concepts of relations: mutual** The third-person perspective permits more than the taking of another’s perspective on the self; the truly third-person perspective on relations which is characteristic of Level 3 simultaneously includes and coordinates the perspectives of self and other(s), and thus the system or situation and all parties are seen from the third-person or generalized other perspective. Whereas at
Level 2, the logic of infinite regress, chaining back and forth, was indeed apparent, its implications were not. At Level 3, the limitations and ultimate futility of attempts to understand interactions on the basis of the infinite regress model become apparent and the third-person perspective of this level allows the adolescent to abstractly step outside an interpersonal interaction and simultaneously and mutually coordinate and consider the perspectives (and their interactions) of self and other(s). Subjects thinking at this level see the need to coordinate reciprocal perspectives, and believe social satisfaction, understanding, or resolution must be mutual and coordinated to be genuine and effective. Relations are viewed more as ongoing systems in which thoughts and experiences are mutually shared.

The process of language acquisition comes to an end between the ages of five and nine. The incomplete perspective taking that typifies Selman’s level 1 already rests on a firm foundation of linguistically mediated intersubjectivity. If, following G. H. Mead, we start from the assumption that the growing child acquires an understanding of identical meanings, i.e., intersubjectively valid conventions of meaning as he repeatedly assumes the perspectives and attitudes of a reference person in an interactional context, then the development of action perspectives studied by Selman continues on from the now completed development of perspective taking in the domain of speech perspectives. Having learned to speak, the child already knows how to address an utterance to a hearer with communicative intent. Conversely, he also knows what it is to understand another person’s utterance from the perspective of the person to whom it is addressed. The child has mastered a reciprocal I-thou relation between speaker and hearer when he is able to distinguish saying from doing. At that point the child differentiates between acts of seeking understanding with a hearer—that is, speech acts and their equivalents—and acts that have an impact on physical or social objects. Thus our point of departure is a situation where reciprocal speaker-hearer relations have been established at the level of communication but not yet at the level of action. The child understands what alter means when he states, demands, announces, or wants something; ego also knows how alter comprehends ego’s utterances. This reciprocity between speaker and hearer perspectives, which relates to what is being said, is not yet a reciprocity of action orientations; in any event, it does not automatically affect the actor’s structure of expectations, that is, the perspectives from which actors make and pursue their plans for action. The coordination of action plans necessitates a meshing of action perspectives beyond the reciprocity of speaker perspectives. In what follows I will interpret Selman’s stages in these terms.

At level 1 Selman postulates that the child distinguishes between the interpretive and action perspectives of the various participants in interaction. But in judging the actions of others, he is unable to simultaneously maintain his own point of view and step into the other’s shoes, which is why he is also unable to judge his own actions from the standpoint of others. The child is beginning to differentiate between the outer world and the inner worlds of privileged access. However, the sharply delineated basic sociocognitive concepts of the world of normativity that Kohlberg posits for the conventional stage of social perspectives are lacking. At Selman’s level 1, the child correctly uses sentences expressing statements, requests, wishes, and intentions. No clear meaning is as yet attached to normative sentences; imperatives are not yet dichotomized into those which represent the speaker’s subjective claim to power and those which represent a normative, i.e., nonpersonal claim to validity.

The first step toward bringing about the coordination of the action plans of various interacting participants on the basis of a shared definition of the situation is to extend the reciprocal relationship between speaker and hearer to the relationship between actors who interpret a shared action situation in terms of their diverse plans and from different perspectives. It is no coincidence that Selman identifies this stage of perspective taking as the second-person perspective. With the passage to level 2, the young person learns to make a reversible connection between the action perspectives of speaker and hearer. He can now assume the action perspective of another, and he knows that the other can also assume his (ego’s) action perspective. Ego and alter can take each other’s attitudes toward their own action orientations. In this way the first- and second-person communicative roles are extended to the coordination of actions.
The perspective structure built into the performative attitude of a speaker now determines not only linguistic understanding but also interaction as such. The I-thou perspectives of speaker and hearer thereby take on the function of coordinating action. With the transition to level 3, this structure of perspectives again changes as the observer perspective is introduced into the field of interaction. Granted, children have long since been able to make correct use of third-person pronouns in reaching understanding about other persons, their utterances, their possessions, and so forth. They have also already learned to take an objectivating attitude toward things and events that can be perceived and manipulated. Now, however, the young adolescent learns to turn from the observer perspective back to an interpersonal relationship, in which he engages in a performative attitude along with another participant in interaction. The performative attitude is coupled with the neutral attitude of a person who is present but remains uninvolved, in other words, the attitude of a person who witnesses an interactive event in the role of a listener or viewer. This makes it possible to objectify the reciprocity of action orientations attained at earlier levels and to become aware of that reciprocity in its systemic aspect.

The system of action perspectives is now complete. This completion signifies the actualization of the system of speaker perspectives that exists in *me* in the grammar of personal pronouns. This in turn makes possible a new way to organize conversation. The new structure consists of the ability to view the reciprocal interlocking of action orientations in the first and second persons from the perspective of the third person. Once interaction has been restructured in this way, participants can not only take one another's action perspectives but also exchange the participant perspective for the observer perspective and transform the one into the other. At level 3 of perspective taking, the constitution of the social world, for which level 2 represented a preparation, is brought to completion. To show how this occurs, I will first have to identify the types of interaction that are restructured in the passage from level 2 to level 3 to become strategic and norm-guided action.

Selman's theory was initially developed on the basis of material from clinical interviews conducted after the experimenter had shown two filmed stories. One of these short films is about a girl named Holly. The dilemma Holly gets into stems from the conflict between a promise she has made to her father and her relationship with a girlfriend who needs her help. Selman's theory was initially developed on the basis of material from clinical interviews conducted after the experimenter had shown two filmed stories. One of these short films is about a girl named Holly. The dilemma Holly gets into stems from the conflict between a promise she has made to her father and her relationship with a girlfriend who needs her help. Selman's theory was initially developed on the basis of material from clinical interviews conducted after the experimenter had shown two filmed stories. One of these short films is about a girl named Holly. The dilemma Holly gets into stems from the conflict between a promise she has made to her father and her relationship with a girlfriend who needs her help. Selman's theory was initially developed on the basis of material from clinical interviews conducted after the experimenter had shown two filmed stories. One of these short films is about a girl named Holly. The dilemma Holly gets into stems from the conflict between a promise she has made to her father and her relationship with a girlfriend who needs her help.

James Youniss has made a comparative structural analysis of the social relations between children and adults on the one hand and between children in the same age group on the other. He characterizes these relations in terms of different forms of reciprocity. A nonsymmetrical form of reciprocity, complementarity between behavioral expectations of different kinds, tends to obtain whenever authority is unequal, as in the family. By contrast, a symmetry between behavioral expectations of the same kind obtains in egalitarian friendships. With respect to the coordination of actions, a consequence of authority-governed complementarity is that one person controls the other's contribution to the interaction; interest-governed reciprocity, in contrast, means that the participants exercise mutual control over their contributions to the interaction. It appears that authority-governed complementarity 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. In both types of action we find implemented the I-thou perspectives that speaker and hearer assume vis-a-vis one another. According to Selman, children at this level also possess analogously structured concepts of behavioral expectation, authority, motives for action, and the ability to act. This sociocognitive inventory enables them to differentiate between the outer world and the inner world of a person, to impute intentions and need dispositions, and to distinguish intentional from unintentional acts. Children thereby also acquire the ability to control interactions by deception if necessary.
In cooperative relationships the participants renounce the use of deception. 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. For ego's and alter's behavior to be competitive, the impact they have on one another must be reciprocal. This sort of competition also occurs, it is true, within the institutional framework of the family, where there is an objective differential in authority between the generations. In that case, however, the child behaves toward members of the older generation as though the relationship were symmetrical.

It is therefore advisable to distinguish preconventional types of action not in terms of action systems but in terms of the more abstract criterion of forms of reciprocity (table 1). In cases 2 and 4, conflicts are resolved by means of different strategies. Where the child sees himself as dependent (case 2), he will try to resolve the conflict between his own needs and alter's demands by avoiding threatened sanctions. The considerations that will guide his action resemble in their structures the judgments of Kohlberg's first moral stage. Where the child sees power as distributed equally (case 4), he may try to avail himself of the possibilities for deception that exist in symmetrical relations. J. H. Flavell has simulated this case using a coin experiment.35

The psychological study of perspective taking began with this special case, which is one of the four types of interaction. Here is how Flavell set up this famous experiment: two cups are put upside down on a table. They conceal different amounts of money, one nickel and two nickels respectively. Each cup bears a label in plain view indicating the number of nickels hidden under the cup. The experimental subjects are shown that the relationship between the inscription and the actual amount hidden can be varied at will. The subject's task is to secretly distribute the coins in such a way that another person who is called into the room will fail to guess where the greater amount is hidden, because he has been deceived, and will end up with nothing. The experiment is defined in such a way that the subjects accept the framework of elementary competitive behavior; they try to influence the decisions of the other indirectly. The participants proceed on the assumptions that each pursues only his particular interests, monetary or otherwise, that each knows the other's interest, that they are forbidden to reach an understanding directly (which is why they have to infer hypothetically how the other is going to behave), that deception on both sides is necessary or at least permitted, and that the normative claims to validity that might be bound up with the rules of the game may not be questioned within the framework of the game. The point of the game is clear: alter will try to win as much as he can, an ego is to prevent this. If the experimental subjects have the perspective structure of Selman's level 2, they will choose what Flavell calls strategy B: the child assumes that alter is guided by monetary considerations and will therefore think that the two nickels are under the one-nickel cup; his rationale is, Alter thinks I want to fool him by not putting the two nickels under the two-nickel label. This is an experimentally produced example of competitive behavior that entails a reciprocal I-thou perspective (table 1, case 4). It is easy to trace the reorganization of the preconventional stage of interaction in this type of action. Experimental subjects who are able to engage in Selman's level 3 of perspective taking will choose Flavell's strategy C: they will turn the spiral of reflection one more time, taking into account that alter sees through ego's strategy B (and the reciprocity of action perspectives underlying it). An adolescent acquires this insight as soon as he has learned to objectify the reciprocal relations between ego and alter from the observer perspective and view them as a system. In principle, he is now in a position even to

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Preconventional types of action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forms of reciprocity</td>
<td>Action orientations</td>
</tr>
<tr>
<td>Authority-governed complementarity</td>
<td>Cooperation</td>
</tr>
<tr>
<td>Interest-governed symmetry</td>
<td>Conflict</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
recognize the structure of this two-person game, which is that if both partners act rationally, the chance of losing is as great as the chance of winning, no matter what ego decides to do.

Strategy C is characteristic of a type of action that is possible only at the conventional stage of interaction if, as I have suggested, Selman's more complex level 3 of perspective taking is required for this stage. In these terms the restructuring of preconventional competitive behavior into strategic action is marked by a coordination of observer and participant perspectives.

The concept of the acting subject is also affected by this shift. Ego is now in a position to attribute stability over time to alter's pattern of attitudes and preferences. Alter stops being perceived as someone whose actions are determined by shifting needs and interests and is now perceived as a subject who intuitively follows rules of rational choice. Beyond this, however, no structural change in the sociocognitive inventory is required. In all other respects the preconventional inventory is adequate for the strategic actor. It suffices for him to derive behavioral expectations from imputed intentions, to understand motives in terms of orientation to rewards and punishments, and to interpret authority as the power to promise or threaten positive or negative sanctions (table 2).

Unlike elementary competitive behavior (table 1, case 4), the three other types of preconventional action cannot be adapted quite so economically to the conventional stage of interaction.

Up to this point I have been looking at the development of strategic action as it comes to be differentiated from competitive behavior. The hypothesis I favor explains the transition to the conventional stage of interaction in terms of an amalgamation of the observer and I-thou perspectives into a system of action perspectives that can be transformed into one another. This amalgamation is attended by a completion of the system of speaker perspectives, at which point the organization of dialogue reaches a new level. This latter development need not concern us here. Instead, I want to examine at this juncture how the other three preconventional action types (table 1, cases 1 to 3) change when the transition to the conventional stage of interaction occurs.

Once again I will restrict myself to the structural features of this change, leaving aside the question of how the restructuring of action perspectives might be explained dynamically. All I want to do is propose an analytical distinction between two paths of development, one for normatively regulated action and the other for strategic action. Let me characterize the problematic situation that forms the point of departure for this transition:

- Neither the action-orienting authority of reference persons nor the immediate orientation to individual needs is sufficient to meet the demand for coordination.
- Competitive behavior already has a new basis in strategic action; it has thus been disengaged from an immediate orientation to individual needs.
- A polarization of the attitude oriented toward success and the attitude of action oriented toward reaching understanding

<table>
<thead>
<tr>
<th>Types of action</th>
<th>Sociocognitive inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perspective structures</td>
</tr>
<tr>
<td>Preconventional competitive behavior</td>
<td>Reciprocal interconnection of action perspectives (Selman's level 2, Flavell's strategy B)</td>
</tr>
<tr>
<td>Strategic action</td>
<td>Coordination of observer and participant perspectives (Selman's level 3, Flavell's strategy C)</td>
</tr>
</tbody>
</table>
has been established, and this forces the child to choose between two types of action: one involving deception and the other being free of it. The choice has become both compulsory and normal.

In this situation, preconventional modes of coordinating action come under pressure in areas of behavior not dominated by competition. The sociocognitive inventory has to be restructured to make room for a mechanism of nonstrategic (or understanding-oriented) coordination of action. This mechanism must be independent both of authority relations to actual reference persons and of direct links to the actor's own interests. This stage of conventional but nonstrategic action requires basic sociocognitive concepts revolving around the notion of a suprapersonal will. The notion of behavioral expectations that are covered by this supra-personal authority (i.e., the notion of a social role) levels the difference between alien imperatives and one's own intentions, transforming both the notion of authority and the notion of interest.

Selman (1980) and Damon have given accounts of the development of the concepts of friendship, person, group, and authority during middle childhood. These accounts agree in their essentials. As observations by human ethologists of early mother-child interaction show, these basic concepts undergo an extremely complex development that extends back to the first months of life. From this store of early social ties and intersubjective relations, discrete sociocognitive capacities emerge through a process of differentiation that extends to middle childhood. It would appear that these capacities are used only selectively in the realm of competitive behavior. For preconventional competitive behavior can be transformed into strategic action without the introduction of an observer perspective into the sphere of interaction affecting the sociocognitive inventory as a whole. The passage to normatively regulated action, on the other hand, does require a global reconstruction, which Selman traces in four different dimensions. One possible explanation why this global reconstruction is required is that reorganization along this line of development involves the three preconventional types of action that preclude deception and rely instead on consensus. Studies of children's cooperative behavior in dealing with problems of distribution and peer group conflicts at various age levels provide empirical access to rudimentary prototypes of normatively regulated action. As the child advances in age and cognitive maturity, his ability to solve interpersonal problems in his peer group grows continually. This ability is a good indicator of the mechanisms for coordinating action that are available to him at different stages of development.

In what follows I will limit myself to two concepts: suprapersonal authority and action norm. These concepts are constitutive for the strict concept of the social world as the sum total of legitimately ordered interpersonal relationships. At the preconventional level the child views authority and friendship relations as relations of exchange (e.g., exchange of obedience for security or guidance, of demand for reward, of one achievement for another or for a show of confidence). At the conventional stage, however, the notion of exchange no longer fits the now reorganized relations. At this point the child's views of social bonds, authority, and loyalty become dissociated from specific reference persons and contexts. They are transformed into the normative concepts of moral obligation, the legitimacy of rules, the normative validity of authoritative commands, and so on.

Preparation for this step takes place at the second stage of interaction, i.e., in a framework of reciprocal action orientations, as the growing child (A) learns particularistic behavior patterns by interacting with a specific reference person (B). I have proposed a reconstruction of this transition elsewhere. My proposal, however, is concerned only with issues of conceptual analysis.

What the child initially sees behind particular behavioral expectations is only the authority of a concrete person, an imposing person who is the object of emotional cathexis. The task of passing to the conventional stage of interaction consists in reworking the imperative arbitrary will of a dominant figure of this kind into the authority of a suprapersonal will detached from this specific person. As we know, Freud and Mead alike assumed that particular behavior patterns become detached
from the context-bound intentions and speech acts of specific individuals and take on the external form of social norms to the extent that the sanctions associated with them are internalized by taking the attitudes of others, that is, to the extent that they are assimilated into the personality of the growing child and thus made independent of the sanctioning power of concrete reference persons. In the process the imperative significance of expectation changes in such a way that the individual wills of A and B are now subordinated to a combined will delegated to a generalized social-behavioral expectation. In this way there arises for A a higher-level imperative that is a generalized pattern applicable to all members of a social group. Both A and B appeal to it when uttering an imperative or wish.

Whereas Freud inquired into the psychodynamic facets of this process, Mead was interested in the sociocognitive conditions of internalization. He explained why it is that particular behavior patterns can be generalized only when A has learned to take an objectivating attitude toward his own actions and knows how to divorce the reciprocal system of action perspectives governing A and B from the contingent context of their encounter. Only when A in his interaction with B adopts the attitude of an impartial member of their social group toward them both can he become aware of the interchangeability of his and B’s positions. A realizes that what he thought was a special behavior pattern applicable only to this particular child and these particular parents has always been for B the result of an intuitive understanding of the norms that govern relations between children and parents in general. As he learns to internalize concrete expectations, A forms the concept of a social behavior pattern that applies to all group members, a pattern in which the places are not reserved for alter and ego but can in principle be taken by any member of their social group.

The social generalization of behavior patterns also impinges on the meaning of their imperative aspect. Henceforth A will view interactions in which A, B, C, D, . . . express or obey imperatives and wishes as carrying out the collective will of a group to which A and B jointly subordinate their arbitrary wills. Behind the social role is the authority of a generalized group-specific imperative representing the united power of a concrete group that demands and receives loyalty. In this process the forms of reciprocity inherent in social relations also change. When the persons concerned play their social roles knowing that as members of a social group they are entitled to expect certain actions from others in given situations and at the same time obliged to fulfill the justified behavioral expectations of others, they are basing themselves on a symmetrical form of reciprocity even though the contents of the roles are still distributed in a complementary fashion among the different group participants.

The group’s power to punish and reward, which stands behind social roles, loses the character of a higher-stage imperative only when the growing child once again internalizes the power of institutions (which at first confronts him as a fact of life) and anchors it internally as a system of behavioral controls. Only when A has learned to conceive of group sanctions as his own sanctions, which he himself has set up against himself, does he have to presuppose his consent to a norm whose violation he punishes in this way. Unlike socially generalized imperatives, institutions possess a validity that is derived from intersubjective recognition, the approval of all concerned. The affirmative responses underlying this consensus have at first an ambiguous status. To be sure, they no longer simply denote the “yes” of the compliant hearer responding to an imperative. A “yes” of this kind is equivalent to an intentional sentence referring to the action required; it is merely an expression of an arbitrary will uninformed by a norm. On the other hand, “yes” responses at this particular stage do not yet have the character of affirmative responses to criticizable validity claims. If it were otherwise, one would have to assume that the mere acceptance of norms of action is always and everywhere based on some rationally motivated agreement by all concerned. What speaks against this assumption is the repressiveness indicated by the fact that most norms take effect in the form of social control. Conversely, however, to the extent to which it is exercised through group-specific norms, social control is not based on repression alone.

This equivocal traditionalist understanding is already based on a conception of the legitimacy of norms of action. Within
the horizon of this conception, social roles can cease to be properties of primary groups and can become generalized components of a system of norms. What emerges is a world of legitimately ordered interpersonal relations. Similarly, the concept of role behavior is transformed into that of norm-guided interaction. Through reference to the legitimate validity of norms, duties become distinct from inclinations, and responsible action from contingent or unintended violations. Table 3 provides an overview of the corresponding changes in the sociocognitive inventory, which I will not dwell on at this point.

IV On Grounding Moral Stages in a Logic of Development

Having proposed a reconstruction of two stages of interaction along the lines suggested by studies of perspective taking, I will now return to my initial query, namely the question whether Kohlberg's social perspectives can be linked with stages of interaction in such a way as to permit a plausible grounding of moral stages in a logic of development. I will begin by looking at the question of how the considerations raised so far bear on the ontogenesis of a decentered understanding of the world that is structurally rooted in action oriented toward reaching understanding (section 1 below). To do this, I will need to introduce discourse as a third stage of interaction. Introducing the hypothetical attitude into the domain of interaction and passing from communicative action to discourse signify, in reference to the social world, a moralization of existing norms. This devaluation of naively accepted institutions makes necessary a transformation of the sociocognitive inventory of the conventional stage into basic concepts that are moral in the immediate sense (section 2). Finally, I will assemble those aspects of a logic of development in terms of which social perspectives can be correlated with specific stages of interaction and the corresponding forms of moral consciousness can be justified as stages (section 3).

Following Selman, I can characterize the preconventional stage of interaction in terms of reciprocity between the action per-
spectives of participants. These, I argued, represent the implementation of speaker perspectives in action types, more specifically, the implementation of the I-thou perspectives previously acquired by the child along with the communicative roles of speaker and hearer. The conventional stage of interaction is characterized by a system of action perspectives that comes into being when the observer perspective is joined to the participant perspective of the previous stage. This insertion of the observer perspective into the field of interaction makes it possible (a) to link the third-person role to the communicative roles of the first and second persons and thus to complete the system of speaker perspectives (which has an effect on the level of the organization of dialogue). The new perspective structure is a necessary precondition (b) for the transformation of interest-governed conflict behavior into strategic action and (c) for the formation of the basic sociocognitive concepts that structure normatively regulated action. As the social world of legirnately ordered interpersonal relations is taking shape, (d) a norm-conformative attitude and its corresponding perspective are generated. They supplement the basic attitudes and world perspectives linked with the inner and outer worlds. The linguistic correlate of this system of world perspectives is the three basic modes of language use that every competent speaker is able to distinguish and combine when he takes a performative attitude. Processes (a) to (d) satisfy the structural preconditions of a communicative action (e) in which individual plans of action are coordinated by means of a mechanism for reaching understanding through communication. Normatively regulated action represents one among several pure types of action oriented toward reaching understanding.42

In connection with the types of action analyzed so far, the differentiated form of communicative action is of interest only in that it has a corresponding form of reflection, i.e., discourse, which constitutes a third stage of interaction, albeit one in which the pressure to act is minimized. Argumentation serves to focus on and test validity claims that are initially raised implicitly in communicative action and are naively carried along with it. Argumentation is characterized by the hypothetical attitude of those who take part in it. From this perspective, things and events become states of affairs that may or may not exist. Similarly, this perspective transforms existing norms, norms that are empirically recognized or socially accepted, into norms that may or may not be valid, that is, worthy of recognition. Whether assertoric statements are true and whether norms (or the corresponding normative statements) are right thus become matters for discussion.

This third stage of interaction sees a further growth in the complexity of the perspective structure. At the conventional stage the reciprocal participant perspectives and the observer perspective, two elements that developed at the preconventional stage but were not yet coordinated, are joined. In a similar fashion, the two systems of speaker perspectives and world perspectives, two systems that had been fully developed at the second stage but not yet coordinated, are joined at the third stage. On the one hand, the system of world perspectives, which has been refracted, as it were, by the hypothetical attitude, is constitutive of the claims to validity that are thematized in argumentation. 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. In discourse, then, the two systems must be put in relationship to one another.

This increasingly complex perspective structure can be intuitively grasped in the following terms. At the conventional stage the characteristic innovation was the actor's ability to view himself in reciprocal relation to others as a participant in a process of action and at the same time to step outside and observe himself as a constituent part of interaction. At that stage the perspectives had to interlock in an interpersonal framework of interaction: the observer perspective achieved specificity and was joined with the third-person communicative role of the disinterested observer. Similarly, in agreement attained through discourse, the actors rely, in the act of consenting, on the complete reversibility of their relations with other participants in the argumentation and at the same time attribute the position they take to the persuasive force of the better argument, no matter how their consensus was reached in actual fact. Here too the perspectives interlock in an interpersonal
framework of communication whose presuppositions are improbable: world perspectives that have been refracted by reflection are linked up with the roles of opponents and proponents who criticize and defend validity claims.

What typifies the development of interaction is not only the growing ability to coordinate perspectives that used to exist in isolation but also the greater degree of integration of previously separate types of interaction. As we have seen, the type I have called role behavior represents successful integration of two forms of reciprocity that developed in different types of action at the first stage of interaction. Complementary and symmetrical relations are synthesized even before a mature concept of normative validity is available. That synthesis occurs in the notion of a higher-level suprapersonal imperative in which the intersubjective authority of a common will is expressed. Yet the price paid for this synthesis is a polarity, with strategic action on one side and normatively regulated action on the other. Only at the third stage of interaction is this split overcome. What happens in argumentation is that the success-orientation of competitors is assimilated into a form of communication in which action oriented toward reaching understanding is continued by other means. In argumentation, proponents and opponents engage in a competition with arguments in order to convince one another, that is, in order to reach a consensus. This dialectical role structure makes forms of disputation available for a cooperative search for truth. Argumentation can exploit the conflict between success-oriented competitors for the purpose of achieving consensus so long as the arguments are not reduced to mere means of influencing one another. In discourse what is called the force of the better argument is wholly unforced. Here convictions change internally via a process of rationally motivated attitude change.

As he passes into the postconventional stage of interaction, the adult rises above the naiveté of everyday life practice. Having entered the quasi-natural social world with the transition to the conventional stage of interaction, he now leaves it behind. As he becomes a participant in discourse, the relevance of his experiential context pales, as do the normativity of existing orders and the objectivity of things and events. On the plane of metacommunication the only perspectives on the lived world left to him are retrospective ones. In the light of hypothetical claims to validity the world of existing states of affairs is theorized, that is, becomes a matter of theory, and the world of legitimately ordered relations is moralized, that is, becomes a matter of morality. This moralization of society—that is, of the normatively integrated structure of relationships that the growing child initially had to appropriate through construction—undermines the normative power of the factual: from the isolated viewpoint of deontological validity, institutions that have lost their quasi-natural character can be turned into so many instances of problematic justice. Problematization of this sort, in turn, arrests the process of communicative action before it is completed. It severs the ties between the social world and the surrounding lifeworld, and it jolts the intuitive certainties that flow into the social world from the lifeworld. Interactions now appear in a different light. When they become subject to judgment from a purely moral point of view, interactions emancipate themselves from parochial conventions but also lose the vigorous historical coloration of a particular form of life. Interactions become strangely abstract when they come under the aegis of principled autonomous action.

As the social world is dissociated from the context of a form of life that used to be its ever present background of certitude and habituation and is put at a distance by participants in discourse who take a hypothetical attitude, the uprooted and now free-floating systems of norms require a different basis. This new basis has to be achieved through a reorganization of the fundamental sociocognitive concepts available at the preceding stage of interaction. The means to the solution of this problem is the very same perspective structure of a fully decentered understanding of the world that created the problem in the first place. Norms of action are now conceived as subject to other norms in turn. They are subordinated to principles, or higher-level norms. The notion of the legitimacy of norms
of action is now divided into the components of mere de facto recognition and worthiness to be recognized. The social currency of existing norms is no longer equivalent to the validity of justified norms. To these differentiations within the concepts of norm and normative validity corresponds a parallel differentiation in the concept of duty. Respect for the law is no longer considered an ethical motive per se. To heteronomy, that is, 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.

With this concept of autonomy, the notion of the capacity for responsible action also changes. Responsibility becomes a special case of accountability, the latter here meaning the orientation of action toward an agreement that is rationally motivated and conceived as universal: to act morally is to act on the basis of insight.

The concept of the capacity to act that develops at the postconventional stage of interaction makes it clear that moral action is a case of normatively regulated action in which the actor is oriented toward reflectively tested claims to validity. Intrinsic to moral action is the claim that the settling of action conflicts is based on justified reasoning alone. Moral action is action guided by moral insight.

This strict concept of morality can evolve only at the postconventional stage. To be sure, even at earlier stages there is an intuitive grasp of the moral linked with a conception of consensual resolution of action conflicts. But at these earlier stages actors are relying on ideas of, shall I say, the good and just life, ideas that make possible a transitive ordering of conflicting needs. Only at the postconventional stage is the social world uncoupled from the stream of cultural givens. This shift makes the autonomous justification of morality an unavoidable problem. The very perspectives that make consensus possible are now at issue. Independently of contingent commonalities of social background, political affiliation, cultural heritage, traditional forms of life, and so on, competent actors can now take a moral point of view, a point of view distanced from the controversy, only if they cannot avoid accepting that point of view even when their value orientations diverge. Consequently, this moral reference point must be derived from the structure in which all participants in interaction always already find themselves insofar as they act communicatively. As discourse ethics shows, a point of reference of this kind is contained in the general pragmatic presuppositions of argumentation as such.

The passage to principled moral judgment is only a first, incomplete step in the adult's dissociation from the traditional world of existing norms. The principles governing our judgments about norms (principles of distributive justice, for instance) are principles in the plural, and they themselves require justification. The moral point of view cannot be found in a first principle, nor can it be located in an ultimate justification that would lie outside the domain of argumentation. Justificatory power resides only in the discursive procedure that redeems normative claims to validity. And this justificatory power stems in the last analysis from the fact that argumentation is rooted in communicative action. The sought-after moral point of view that precedes all controversies originates in a fundamental reciprocity that is built into action oriented toward reaching understanding. This reciprocity first appears in the form of authority-governed complementarity and interest-governed symmetry. Later it manifests itself in the reciprocity of behavioral expectations that are linked together in social roles and in the reciprocity of rights and obligations that are linked together in norms. Finally, it shows up as ideal role taking in discourse and insures that the right to universal access to, and equal opportunity for participation in argumentation is enjoyed freely and equally. At this third stage of interaction, then, an idealized form of reciprocity becomes the defining characteristic of a cooperative search for truth on the part of a potentially unlimited communication community. To that extent morality as grounded by discourse ethics is based on a pattern inherent in mutual understanding in language from the beginning.

Having reviewed the sociocognitive inventory and the perspective structure of the three stages of interaction, I would like to
Moral Consciousness and Communicative Action

return to the sociomoral perspectives from which Kohlberg’s stages of moral judgment are directly derived. With the help of these social perspectives Kohlberg defines the points of view in terms of which a transitive order of contested interests can be established and conflicts settled consensually. These points of view owe their existence to the combination of a given perspective structure with the corresponding idea of the good and just life, as I will demonstrate. As the two right-hand columns of table 4 show, the first of these two components requires no explanation. The second, however, does.

On the face of it, it is difficult to understand how the normative component of the social perspectives, namely the conception of justice, emerges from the sociocognitive inventory of the corresponding stages of interaction.

In trying to explain this process, one has to take into account the fact that the normatively integrated fabric of social relations is moral in and of itself, as Durkheim has shown. The basic moral phenomenon is the binding force of norms, which can be violated by acting subjects. All basic concepts that are constitutive of normatively regulated action, then, already have a moral dimension, which is merely being actualized and fully employed when people judge conflicts and violations of norms. With the formation of the social world and the transition to norm-guided interaction, all social relations take on an implicitly ethical character. Golden rules and obedience to the law are ethical imperatives that merely sue, as it were, for what is already implicit in social roles and norms prior to any actual moral conflict: the complementarity of behavioral expectations and the symmetry of rights and duties.

The conformity to role expectations and norms, which assures consensus, flows naturally from the sociocognitive inventory only because at the conventional stage the social world is still embedded in the lifeworld and reinforced by its certainties. At this point, morality (Moralität) and the ethics (Sittlichkeit) of an unquestioned, habitual, particular form of life have not yet parted ways; morality has not become autonomous as morality. Duties are interwoven with habitual concrete life practice in such a way that they derive self-evidence from background certitudes. At this stage, issues of justice are posed within the framework of questions of the good life, questions which have always already been answered. This is also true of religious and classical-philosophical ethics that take ethical life as their theme. They too understand and justify the moral not in its own terms but within the horizon of a larger soteriological or cosmological whole.

As we have seen, this syndrome disintegrates when a hypothetical attitude is introduced. Before the reflective gaze of a participant in discourse the social world dissolves into so many conventions in need of justification. The empirical store of traditional norms is split into social facts and norms. The latter have lost their backing in the certainties of the lifeworld and must now be justified in the light of principles. Thus the orientation to principles of justice and ultimately to the procedure of norm-justifying discourse is the outcome of the inevitable moralization of a social world become problematic. Such are the ideas of justice that, at the postconventional stage, take the place of conformity to roles and norms.

At the preconventional stage we cannot speak of conceptions of justice in the same sense as at later stages of interaction. Here no social world in my sense of the term has yet been constituted. The sociocognitive concepts available to the child lack a clear-cut dimension of deontological validity. For perspectives with socially binding force, the child must look to an inventory that interprets reciprocally interlocking action perspectives in terms of authority relations or external influence. Hence, preconventional notions of bonds and loyalties are based either on the complementarity of command and obedience or on the symmetry of compensation. These two types of reciprocity represent the natural embryonic form of justice conceptions inherent in the structure of action as such. Only at the conventional stage, however, are conceptions of justice conceived as conceptions of justice. And only at the postconventional stage is the truth about the world of preconventional conceptions revealed, namely that the idea of justice can be gleaned only from the idealized form of reciprocity that underlies discourse.

For now these few remarks will have to suffice to give plausibility to my thesis that there are structural relationships be-
<table>
<thead>
<tr>
<th>Types of action</th>
<th>Perspective structure</th>
<th>Structure of behavioral expectations</th>
<th>Concept of authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preconventional</td>
<td>Reciprocal interlocking of action perspectives</td>
<td>Particular behavior pattern</td>
<td>Authority of reference persons: externally sanctioned will</td>
</tr>
<tr>
<td>Interaction controlled by authority</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperation based on self-interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td>Coordination of observer and participant perspectives</td>
<td>Group-wide generalized of behavior patterns: social roles</td>
<td>Internalized authority of supraindividual will (Wille): loyalty</td>
</tr>
<tr>
<td>Role behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normatively governed interaction</td>
<td></td>
<td>Group-wide generalization of roles: system of norms</td>
<td>Internalized authority of an impersonal collective will (Wille): legitimacy</td>
</tr>
<tr>
<td>Postconventional</td>
<td>Integration of speaker and world perspectives</td>
<td>Rules for testing norms: principles</td>
<td>Ideal vs. social validity</td>
</tr>
<tr>
<td>Discourse</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept of motivation</th>
<th>Social perspectives</th>
<th>Justice concept</th>
<th>State of moral judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loyalty to reference persons; orientation toward rewards and punishments</td>
<td></td>
<td></td>
<td>Complementarity of order and obedience</td>
</tr>
<tr>
<td></td>
<td>Egocentric perspective</td>
<td>Symmetry of compensation</td>
<td>2</td>
</tr>
<tr>
<td>Duty vs. inclination</td>
<td>Primary-group perspective</td>
<td>Conformity to roles</td>
<td>3</td>
</tr>
<tr>
<td>Perspective of a collectivity (the system's point of view)</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Orientation to the existing system of norms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principled perspective (prior to society)</td>
<td>Orientation toward principles of justice</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Procedural perspective (ideal role taking)</td>
<td>Orientation toward procedures for justifying norms</td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>
Moral Consciousness and Communicative Action

between moral stages and social perspectives on the one hand and stages of interaction on the other, relationships that justify the parallels I drew in table 4. These parallels can sustain the burden of a justification in terms of a logic of development, however, only if what I have thus far merely tacitly anticipated by using the term "stages" can in fact be demonstrated for stages of interaction, namely that the proposed hierarchy of action types reflects a logic of development. I have tried to make this anticipatory theoretical characterization clear in my presentation of how stages of interaction are introduced and how the transitions between them can be reconstructed. I began by showing that the I-thou perspectives and the observer perspective serve as building blocks for an increasingly complex perspective structure that culminates in the decentered understanding of the world displayed by subjects who act with an orientation toward reaching understanding. Viewed in terms of a progressively decentered understanding of the world, the stages of interaction express a development that is directed and cumulative. Second, I distinguished stages of interaction in terms of different achievements of coordination. At the preconventional stage the action perspectives of different participants are reciprocally related to one another. At the conventional stage the observer perspective is added to these participant perspectives. In the end the separately formed systems of speaker perspectives and world perspectives that have been developed on the basis of these earlier achievements are integrated. The existence of these breaks supports the view that successive perspective structures are in fact discrete totalities. Third, we saw that in normatively regulated action the opposition between authority-governed complementarity and interest-governed symmetry characteristic of preconventional action types is overcome, just as the opposition between a consensual orientation and a success orientation, which emerges in the relationship between normatively regulated and strategic action, is overcome in argumentation. This seems to confirm the assumption that higher-level cognitive structures replace the lower ones while preserving them in reorganized form. This dialectical sublation of structures that have been superseded—-a relationship that is difficult to analyze—would need to be demonstrated in detail in the restructuring of the sociocognitive inventory.

We can, however, identify a few trends in specific dimensions. For instance, it is possible to derive the more complex structures of behavioral expectations from the relatively simple ones through generalization and self-application: the generalized social expectation of reciprocally linked behavioral expectations gives rise to norms, and the generalized self-application of norms gives rise to principles by which other norms can be normatively assessed. Similarly, the more complex concepts of normative validity and autonomy emerge from the simpler concepts of an imperative will (Willkür) and personal loyalty, or pleasure-pain orientation. What happens in each of these cases is that the central semantic component of the more elementary concept is decontextualized and thus thrown into sharper relief, which allows the higher-level concept to stylize the superseded concept as a counterconcept. From the perspective of the next-higher stage, for instance, the exercise of authority by reference persons becomes mere arbitrary will, which is then explicitly contrasted with legitimate expressions of will. To cite another example, personal loyalties or pleasure-pain orientations become mere inclinations sharply set off from duties. Correspondingly, the legitimacy of action norms is viewed at the next stage as their mere social acceptance, which is contrasted with ideal validity, while action based on concrete duties is now contrasted with autonomy as something merely heteronomous.

A similar process of dichotomization and devaluation takes place with the transition from a concept of externally imposed punishment to the concepts of shame and guilt and with the transition from the concept of natural identity to the concepts of role identity and ego identity. These comments are programmatic in nature. One would need a more precise concept of developmental logic to carry out this kind of analysis rigorously and to show how the sociocognitive inventory of the elementary stage is subjected to the reconstructive operations of self-application (reflexivity), generalization, and idealizing abstraction.

To review what has been said above, it is clear that placing moral development within the framework of a theory of com-
municative action has advantages for interpretation, both for
a clearer understanding of the connections between moral
judgment and social cognition and for grounding moral stages
in a logic of development.

First, we saw that the same perspective structures are var-
iously embodied in a whole range of types of interaction. A
completely decentered understanding of the world develops
only in the domains of behavior unaffected by competition.
This decentered understanding becomes reflective with
the transition from conventional action to discourse. The contin­
uation of communicative action with argumentative means
marks a stage of interaction that necessitates our going beyond
Selman’s stage of perspective taking. The integration of world
perspectives and speaker perspectives achieved in argumenta­
tion represents the interface between social cognition and
postconventional morality.

These clarifications proved helpful in the attempt to ground
moral stages in a logic of development. Kohlberg’s social per-
spectives are intended to have this function. As we saw, they
can be correlated with stages of interaction that are ordered
hierarchically according to perspective structures and basic
concepts. This allows us to see how notions of justice are de-
vided from the forms of reciprocity available at the various
stages of interaction. With the transition from normatively reg-
ulated action to practical discourse, the basic concepts of prin-
cipled morality spring directly from the reorganization of the
available sociocognitive inventory, a reorganization that occurs
with the necessity of developmental logic. This step marks the
moralization of the social world, with forms of reciprocity that
are built into social interaction and become increasingly ab-
stract forming the naturalistic core, so to speak, of moral
consciousness.

Whether the interpretive advantages I have tried to dem-
onstrate here will prove fruitful in terms of research strategies
as well is a question to be answered at another level. For the
time being, I will use the reconstruction proposed here only to
illuminate some of the difficulties Kohlberg’s theory has had
to contend with in recent years.44

V Anomalies and Problems: A Contribution to Theory
Construction

At present the debates surrounding Kohlberg’s approach re-
volve primarily around four problems. First, given that it has
not yet been possible to prove experimentally the existence of
a hypothetical stage 6 of moral judgment, the question arises
whether and in what sense, if at all, we can speak of natural
stages at the level of postconventional morality. Second, the
cases of regression that occur in the postadolescent period, that
is, in the third decade of life, raise doubts about whether the
normative point of reference of moral development has been correctly
chosen and especially about whether the morally mature adult’s
capacity for judgment and action can be adequately defined in
terms of cognitivist and formalist theories. Third, the question
of accommodating relativists or value skeptics as a group in Kohl-
berg’s stage model remains problematic. Fourth, the question
of whether structuralist theory can be combined with the find-
ings of ego psychology in a way that would do justice to the
psychodynamic aspects of the formation of judgments remains an
open one.

The nature of these problems will become clearer if we can
clarify some important issues: first, the degrees of freedom the
adolescent attains when he makes the transition from norma-
tively regulated action to discourse and achieves detachment
from the social world of quasi-natural embeddedness (section
1), 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 (section
2), third, the escape route that the adolescent takes when he
distances himself from the devalued traditional world of norms
but stops there without taking the further step of reorganizing
the sociocognitive inventory of the conventional stage as a
whole (section 3), and fourth, the discrepancies between moral
judgment and moral action that result from a failure to sepa-
rate the attitude oriented toward success from the attitude
oriented to reaching understanding (section 4).
In recent decades Kohlberg has repeatedly revised his scoring schema [for assigning subjects' responses to a stage level—trans.]. I would not necessarily say that the most recent scoring represents an improvement. In the coding of responses, theories in the Piagetian tradition require a theoretically based hermeneutic interpretation, a type of interpretation that is certainly not susceptible of being operationalized in a foolproof way, that is, in such a way as to neutralize highly complex preunderstandings. Be that as it may, the present methods of scoring interview material have forced Kohlberg to delete stage 6 because longitudinal studies in the United States, Israel, and Turkey no longer provide evidence for its existence. Today Kohlberg is reluctant to answer the question of whether stage 6 is a psychologically distinct natural stage or a philosophical construct. A revision (if it were not based solely on problems of measurement) would necessarily affect the status of stage 6 along with that of stage 6. As soon as we give up the attempt to differentiate stages at the postconventional level, we face the question of whether principled moral judgments represent a natural stage in the same sense as judgments assigned to the preconventional and conventional levels.

From the standpoint of discourse ethics, I have already tacitly suggested a different interpretation of stages 5 and 6 by distinguishing the orientation toward general principles from the orientation toward procedures for the justification of possible principles (table 4). In this interpretation there is no differentiation of stages according to the kind of principles involved (utilitarian versus natural-right versus Kantian). The relevant distinction is conceived solely in terms of two stages of reflection. At stage 5, principles are viewed as being ultimate and beyond the need for justification. At stage 6 they are not only handled more flexibly but also explicitly made relative to procedures of justification. Differentiating stages of reflection in this way is intimately tied up with the larger framework of a specific normative theory and has to prove its mettle there. One must be able to show that a person relying not on the self-evident nature of universal principles but on the legitimating power of procedures for justification is in fact better equipped to oppose skeptical objections and thus also better able to judge consistently. On the other hand, there are ethical positions that reject proceduralism and insist that a procedure for moral justification in no way differs from, and is unable to achieve more than, a universal moral principle. As long as this philosophical controversy is not settled, the fundamental assumptions of discourse ethics should be defended in the arena where they clash directly with other ethical views rather than understood naturalistically as propositions about natural stages of moral consciousness. Discourse ethics itself certainly offers no grounds for a (reifying?) interpretation that claims that stages of reflection have the same status as natural intrapsychic stages of development.

If there is no empirical evidence to suggest that we are dealing with more than one postconventional stage, Kohlberg's description of stage 5 also becomes problematic. We may at least suspect that the ideas of the social contract and the greatest good for the greatest number are confined to traditions that hold sway primarily in England and America and that they represent a particular culturally specific substantive manifestation of principled moral judgment.

Taking up certain misgivings of John Gibbs, Thomas McCarthy points out that the relation between a psychologist knowledgeable about moral theory and his experimental subject changes in a way that is methodologically significant as the subject reaches the postconventional level and takes a hypothetical attitude to his social world:

The suggestion I should like to advance is that Kohlberg's account places the higher-stage moral subject, at least in point of competence, at the same reflective or discursive level as the moral psychologist. The subject's thought is now marked by the decentralization, differentiation and reflexivity which are the conditions of entrance into the moral theorist's sphere of argumentation. Thus the asymmetry between the prereflective and the reflective, between theories-in-action and explanation, which underlies the model of reconstruction begins to break down. The subject is now in a position to argue with the theorist about questions of morality.
In the same essay McCarthy draws a useful parallel between sociomoral and cognitive development:

Piaget views the underlying functioning of intelligence as unknown to the individual at lower stages of cognition. At superior levels, however, the subject may reflect on previously tacit thought operations and the implicit cognitive achievements of earlier stages, that is, he or she may engage in epistemological reflection. And this places the subject, at least in point of competence, at the same discursive level as the cognitive psychologist. Here, too, asymmetry between the subject's prereflective know-how and the investigator's reflective know-that begins to break down. The subject is now in a position to argue with the theorist about the structure and conditions of knowledge.

At the level of formal operations the adult has reflectively appropriated the intuitive knowledge he used in successful problem solving. This means that he has acquired the ability to continue constructive learning processes by means of reconstruction. In principle, he has now broadened his competence to include the reconstructive sciences.

This acquisition has an implication for the methodology of the reconstructive sciences. A psychologist trying to test his hypotheses about the stage of formal operations is dependent on his experimental subjects, whom he must treat as partners equal in principle in the business of scientific reconstruction. His own theory will convince him that at this stage the asymmetry that existed on previous levels between prereflective mental functioning and the attempt to grasp that functioning reflectively disappears. To the extent to which the reconstructive scientist views himself as standing within the open horizon of a research process whose results he cannot foresee, he must accord that same standpoint to experimental subjects who have reached the highest stage of competence.

The same holds for respondents who handle moral dilemmas from the standpoint of a postconventional participant in discourse. Insofar as they essentially share the perspective of the moral psychologist who interviews them, their moral judgments no longer have the form of utterances that are naively generated with the help of an intuitive understanding of rules. Postconventional experimental subjects are drawn into the business of moral psychology—the reconstruction of moral intuitions underlying everyday life—in that their moral reasoning no longer mirrors, that is, prerefexively expresses, a pretheoretical knowledge but rather explicates potentially theoretical knowledge. Principled moral judgments are not possible without the first step in the reconstruction of underlying moral intuitions. Thus principled moral judgments already represent moral-theoretical judgments in nuce. As postconventional thought leaves the world of traditional norms behind, it operates in the same arena in which moral theorists debate their issues. This debate is fuelled by historical experience, and for the time being it is decided on the basis of philosophical arguments and not by developmental paths identified by psychology.

A second cluster of problems has sparked a wide-ranging discussion in the past few years. This cluster is difficult to disentangle. Studies by Norma Haan and Carol Gilligan marked the beginning of this particular debate. Its immediate occasion was the suspicion that in certain critical instances the stage level assigned in accordance with Kohlberg's schema might deviate too far from the intuitive understanding of a morally sensitive scorer. The two instances in question are, first, female respondents whose utterances have to be scored at 3 despite a presumption of greater moral maturity on their part and, second, experimental subjects classified as relativistic value skeptics at 41/2 (see section 3 below) despite the fact that their utterances seem more mature than the usual postconventional judgments. Gilligan and Murphy make the point that Kohlberg's criteria would put more than half of the population of the United States at some level below the postconventional in terms of their moral consciousness. In addition, they show that in a sample of 26 subjects most of the subjects who were at first classified as postconventional in terms of the revised scoring procedure later regressed to relativistic positions (stage 41/2). Although Kohlberg disputes the facts on which his critics rely—the disproportionate numbers of female subjects at lower
stages and instances of regressions that the theory cannot explain\(^\text{52}\)—the controversy has drawn attention to problems that, in the language of the philosophical tradition, pertain to the relation of morality to ethical life (Sittlichkeit).

Gilligan and Murphy envisage a postconventional path of development leading from Kohlberg's stages 5 and 6 (the postconventional formal stage) to a stage they call “contextual relativism” (the postconventional contextual stage). This notion is based on W. B. Perry's work on the overcoming of absolutist thought in late adolescence\(^\text{53}\) and K. Riegel's hypotheses about postformal operations.\(^\text{54}\) At the postconventional contextual stage, we are told, the adult who has become morally mature through conflicts and experiences learns to overcome the abstractions of a strict deontological morality of justice along Kantian lines, a morality that absolutizes the aspect of normative rightness. This relativistic ethics of responsibility deals with real moral dilemmas, not merely hypothetical ones, it takes the complexity of lived situations into account, it joins justice with caring and with responsibility for those under one's care, and it presupposes a more inclusive concept of a mature personality that goes beyond the abstract notion of autonomy:

While the logical concepts of equality and reciprocity can support a principled morality of universal rights and respect, experiences of moral conflict and choice seem to point rather to special obligations and responsibility for consequences that can be anticipated and understood only within a more contextual frame of reference. The balancing of these two points of view appeared to us to be the key to understanding adult moral development. In our view, this would require a restructuring of moral thought which would include but supersede the principled understanding of Kohlberg's highest stages.\(^\text{55}\)

The position of an ethics of responsibility is thereby distinguished from that of value skepticism in the transitional stage 4½. While both are relativistic, only contextual relativism is based on ethical formalism while at the same time superseding it.

From the standpoint of discourse ethics, things look somewhat different. Gilligan and Murphy do zero in on the problems resulting from a successful transition to principled morality. Principled morality, as we have seen, emerges out of a peculiar abstractive achievement that robs the social world, with its legitimately ordered interpersonal relations, of its natural stability and compels it to justify itself. Initially the social world owes its unshakable facticity to its embeddedness in naively habituated concrete forms of life, which form an unquestioned prerelatively given background against which subjects act. By contrast, communicative actors have an explicit knowledge of the given institutional orders to which their speech acts refer. At the conventional stage, however, this explicit knowledge is so intimately tied up with the implicit background certainties of a particular form of life that the intersubjectively accepted norms are accorded absolute validity. As the social world becomes increasingly moralized in the hypothetical attitude of a participant in discourse and thus begins to stand over against the totality of the lifeworld, the erstwhile fusion of validity and what is merely accepted in society is dissolved. At the same time the unity of the practice of everyday communication splits into two parts: norms and values. The first part of the domain of the practical, which consists of norms, is susceptible to the requirement of moral justification in terms of its deontological validity; the second part, which consists of particular value configurations belonging to collective and individual modes of life, is not.

Cultural values embodied and fused in the totalities of life forms and life histories permeate the fabric of the communicative practice of everyday life through which the individual's life is shaped and his identity secured. It is impossible for the individual as an acting subject to distance himself from this life practice as he can distance himself from the institutions of his social world. Cultural values too transcend actual courses of action. They congeal into historical and biographical syndromes of value orientations, enabling individuals to distinguish the reproduction of mere life from ideas of the good life. But ideas of the good life are not notions that simply occur to individuals as abstract imperatives; they shape the identity of groups and individuals in such a way that they form an integral part of culture and personality. A person who questions the forms of life in which his identity has been shaped
questions his very existence. The distancing produced by life crises of that kind is of another sort than the distance of a norm-testing participant in discourse from the facticity of existing institutions.

Thus the formation of the moral point of view goes hand in hand with a differentiation within the sphere of the practical: moral questions, which can in principle be decided rationally in terms of criteria of justice or the universalizability of interests are now distinguished from evaluative questions, which fall into the general category of issues of the good life and are accessible to rational discussion only within the horizon of a concrete historical form of life or an individual life style. The concrete ethical life of a naively habituated lifeworld is characterized by the fusion of moral and evaluative issues. Only in a rationalized lifeworld do moral issues become independent of issues of the good life. Only then do they have to be dealt with autonomously as issues of justice, at least initially. The word "initially" points to the problem dealt with under the rubric of an "ethics of responsibility."

For the increase in rationality brought about by isolating questions of justice has its price. Questions of the good life have the advantage of being answerable within the horizon of lifeworld certainties. They are posed as contextual and hence concrete questions from the outset. The answers to these questions retain the action-motivating potential of the forms of life that are presupposed in the contexts. In the framework of concrete ethical life within which conventional morality operates, moral judgments derive both their concreteness and their action-motivating potential from the intrinsic connection to ideas of the good life and institutionalized ethical life. At this level, problematization has not gone so deep as to jeopardize the advantages of an existing form of ethical life. This, however, is exactly what does occur with the transition to postconventional morality, when the social world becomes moralized and thus divorced from its background in the lifeworld. This abstractive achievement has a double effect: under a strict deontological point of view, moral questions are taken out of their contexts in such a way that moral solutions retain only the rationally motivating force of insights.

Moral issues are never raised for their own sake; people raise them seeking a guide for action. For this reason the demotivated solutions that postconventional morality finds for decontextualized issues must be reinserted into practical life. If it is to become effective in practice, morality has to make up for the loss of concrete ethical life that it incurred when it pursued a cognitive advantage. Demotivated solutions to decontextualized issues can achieve practical efficacy only if two resulting problems are solved: the abstraction from contexts of action and the separation of rationally motivated insights from empirical attitudes must both be undone. Every cognitivist morality will confront the actor with questions both of the situation-specific application and of the motivational anchoring of moral insights. And these two problems can be solved only when moral judgment is supplemented by something else: hermeneutic effort and the internalization of authority.

The notion of a “stage” of contextual relativism rests on a misconception of the basic problem of how ethical life and morality are to be mediated. Carol Gilligan fails to make an adequate distinction between the cognitive problem of application and the motivational problem of the anchoring of moral insights. Accordingly, she tends to make the distinction between conventional formalism and postconventional contextualism in terms of the distinction between hypothetical and actual situations. She ignores the fact that the question of whether what I ought to do is the same as what I would do concerns only the motivational side of the problem of mediation. The other side of the problem is cognitive in nature: In a given situation, how is one to interpret a universal command, which says merely what one ought to do, in such a way that one can then act in accordance with the command within the context of the situation?

Second, Gilligan fails to see that the two problems arise only after morality has been abstracted from ethical life and the basic moral-philosophical question of the justifiability of norms has been answered in terms of a cognitivist ethics. The question of the context-specific application of universal norms should not be confused with the question of their justification. Since moral norms do not contain their own rules of application,
acting on the basis of moral insight requires the additional competence of hermeneutic prudence, or in Kantian terminology, reflective judgment. But this in no way puts into question the prior decision in favor of a universalistic position. 57

Third, Gilligan's contextual relativism is designed to offset certain deficiencies that emerge at the postconventional level of moral judgment when the two resulting problems mentioned above are not dealt with. I speak of moral rigorism, one such deficiency, when hermeneutic sensitivity to the problem of application is lacking and when abstract moral insights are mechanically applied to concrete situations in line with the adage Fiat justitia, paret mundus. Max Weber's dichotomy between an ethics of conviction and an ethics of responsibility is based largely on this popular critique of Kant. I speak of intellectualization, another such deficiency, when moral abstractions serve as defenses. Gilligan tends to misconstrue these deficiencies as characteristic of a normal stage of postconventional formalism.

Finally, Gilligan sets up parallels between postconventional formalism and justice on the one hand and postconventional contextualism and caring and responsibility for a specific circle of people on the other, hypothesizing that the two orientations are unequally distributed between the two sexes.

Yet strictly speaking, the moral point of view is constituted only with the transition from the second to the third stage of interaction. This moral point of view comes about when the social world is moralized from the hypothetical attitude of a participant in argumentation and split off from the world of life. This deontological abstraction separates issues of justice from issues of the good life. Moral questions are thereby dissociated from their contexts and moral answers are dissociated from empirical motives. These dissociations make contextual application and a specific kind of motivational anchoring of moral insights necessary. If one keeps these facts in mind, the solution to these problems requires a mediation between morality and ethical life that goes beyond what can be accomplished by moral judgments as defined by deontological ethics. That is why it does not make sense to try to supplement or revise the stages of moral judgment. The two problems discussed above are on a different level than the capacity for moral judgment. They require a different order of achievement, to wit, contextual sensitivity and prudence on the one hand and autonomy and self-governance on the other. The critical contributions to the discussion touched off by Gilligan's work can be summarized under these headings. 58

The cognitive problem of application
Those who seek to supplement Kohlberg's moral stages, whether by adding another postconventional stage (Carol Gilligan) or by introducing a parallel stage hierarchy (Norma Haan), fail to distinguish sufficiently between moral and evaluative issues, between issues of justice and issues of the good life. In terms of the conduct of an individual life, this corresponds to the distinction between self-determination and self-realization. 59 Typically, questions of preferences as to forms of life or life goals (ego ideals) and questions of the evaluation of personality types and modes of action arise only after moral issues, narrowly understood, have been resolved. 60 Furthermore, a definition of the moral point of view in terms of discourse ethics rules out the possibility of competing points of view with a status equal to that of justice or normative rightness. Since valid norms cannot but embody generalizable interests, it follows that the principle of the general welfare (Frankena's principle of beneficence, for example) 61 or the principle of care and responsibility—insofar as these expressions designate moral principles—are already contained in the meaning of the term normative validity.

The discourse-ethical conception of the moral principle also rules out any narrowing of moral judgment through considerations of an ethics of conviction. Again, consideration of the consequences and indirect effects which are expected to follow from the general application of a contested norm to specific contexts does not need to be supplemented by an ethics of responsibility. Interpreted from the perspective of discourse ethics, practical reason does indeed require practical prudence in the application of rules. But use of this capacity does not restrict practical reason to the parameters of a specific culture or historical period. Learning processes governed by the univer-
salistic substance of the norm being applied are possible even in the dimension of application.

Ideal role taking has come to signify a procedural type of justification. The cognitive operations it requires are demanding. Those operations in turn are internally linked with motives and emotional dispositions and attitudes like empathy. Where sociocultural distance is a factor, concern for the fate of one's neighbor—who more often than not is anything but close by—is a necessary emotional prerequisite for the cognitive operations expected of participants in discourse. Similar connections among cognition, empathy, and agape can be shown to hold for the hermeneutic activity of applying universal norms in a context-sensitive manner. This integration of cognitive operations and emotional dispositions and attitudes in justifying and applying norms characterizes the mature capacity for moral judgment. It is only when we conceptualize maturity in this way that we can see moral rigorism for what it is: an impairment of the faculty of judgment. This concept of maturity, however, should not be applied externally to postconventional thought in the form of an opposition between an ethics of love and an ethics of law and justice. Rather, it should flow from an adequate description of the highest stage of morality itself.62

The motivational problem of anchorage
Those who would supplement Kohlberg's moral stages in one of the ways noted above fail to distinguish clearly between moral development and ego development. What corresponds in the personality system to moral judgment are behavioral controls, or superego structures. At higher stages these are formed only in a process of distancing oneself from conflict with the social world understood as a matrix of relations in the social environment integrated through norms; superego structures can be analyzed in terms of the basic sociocognitive concepts of normatively regulated action. The formation of ego identity, on the other hand, takes place in the more complex contexts of communicative action, more specifically, in the interplay between an individual and the structures of the objective, social, and subjective worlds that gradually become differentiated from the contexts of the lifeworld.63

The postconventional disengagement of morality from ethical life signifies a loss of congruence between fundamental moral conceptions and what is taken for granted as part of a culture, or the certainties of the lifeworld in general. Moral insights and culturally habituated empirical motives are no longer one and the same. The resulting gap between moral judgments and moral actions needs to be compensated for by a system of internal behavior controls that is triggered by principled moral judgments (convictions that form the basis for motivations) and that makes self-governance possible. This system must function autonomously. It must be independent of the external pressure of an existing recognized legitimate order, no matter how small that pressure may be. These conditions are satisfied only by the complete internalization of a few highly abstract and universal principles that, as discourse ethics shows, follow logically from the procedure of norm justification. One way of testing postconventional superego structures is by checking responses to questions of the form "What should I do?" against responses to questions of the form "What would I do?" Responses to the latter Kohlberg calls responsibility judgments. They are indicators of the respondent's intention to act, or his confidence that he will act, in accordance with his moral judgments. Such responsibility judgments are at the same cognitive level as moral judgments. Even if we can interpret them as expressing a conviction (Gesinnung), as judgments they cannot in any way guarantee a correspondence between judgments and actions. We may be able to derive the kind of motivational anchoring without which a postconventional morality cannot be translated into action from the structure of our capacity to act, that is, from the sociocognitive inventory as restructured at the postconventional stage. But whether the psychodynamic processes will in fact meet the requirements of that structure is not something we can learn from answers to questions of the form "Why me?" Only actual practice can tell us that.64

Even if the passage to the postconventional level of moral judgment has been successful, an inadequate motivational anchoring can restrict one's ability to act autonomously. One
especially striking manifestation of such a discrepancy between judgment and action is intellectualization, which uses an elaborate process of making moral judgments about action conflicts as a defense against latent instinctual conflicts.

Let me turn now to the complications Kohlberg faced when he dealt with the group of moral judgments that forced him to introduce the intermediate stage 4½. At issue are relativistic statements that tend to be made from a strategic rather than a moral point of view. Initially Kohlberg and his coworkers were tempted to stress the affinities of relativistic statements with the instrumental hedonism of stage 2. They could not classify these judgments as preconventional, however, because the general level of argumentation among respondents of this type was too high. The hypothetical attitude with which they judged the social world without moralizing it also spoke for an affinity of their statements with judgments on the postconventional level. For these reasons Kohlberg has placed these relativistic judgments between the conventional and the postconventional levels, assigning them to a transitional stage of their own, a stage that calls less for structural description than for psychodynamic explanation and is the expression of an unresolved crisis of adolescence. This interpretation leaves something to be desired, since it cannot explain the possibility of this level of judgment becoming stabilized. That such stabilization is indeed possible is indicated, among other things, by the fact that a philosophical version of the value skepticism of stage 4½ has been developed and defended as a position to be taken seriously by thinkers in a line that starts with Weber and extends to Popper.

Common to subjectivist approaches in ethics is a value skepticism grounded in empiricist assumptions. Such value skepticism calls into question the rationalistic assumptions underlying Kohlberg’s theory of moral development. Modern value skepticism disputes the contention that moral issues can be settled on the basis of valid (i.e., intersubjectively binding) reasons.

Instead, it conducts metaethical investigations designed to explain how the rationalistic illusions of everyday moral intuitions are rooted in our language. Psychology is certainly not the right forum for the dispute between the cognitivists and the skeptics. The cognitivists will have to prevail on the basis of philosophical arguments—this is the premise of the theory of the development of moral consciousness. But psychology must be able to explain why value skepticism, for all its incongruence with the logic of moral development, seems to be a natural stage in that development. Kohlberg should not rest content with inserting a transitional stage into his overall scheme that can be explained only on a psychodynamic basis. By opting for classification as a solution to the problem, he commits himself to indicating the place of the transitional stage in the logic of development and thus to giving a structural description of stage 4½, as he has done with the other stages. The description he currently offers does not satisfy this demand. It reads as follows:

**Level B/C, transitional level**

This level is postconventional but not yet principled. Content of transition: At Stage 4½, choice is personal and subjective. It is based on emotions, conscience is seen as arbitrary and relative, as are ideas such as “duty” and “morally right.”

Transitional social perspective: At this stage, the perspective is that of an individual standing outside of his own society and considering himself an individual making decisions without a generalized commitment or contract with society. One can pick and choose obligations, which are defined by particular societies, but one has no principles for such choice.

My own explanation for the troubling phenomenon of a transitional stage is that the group of respondents in question have only partially completed the transition to the postconventional level. If the integration of speaker perspectives with world perspectives has not been fully achieved and does not take in the social world and the norm-conformative attitude corresponding to it, there will fail to occur the coordination of the success-oriented attitude of the subject who acts strategically with the attitude oriented toward reaching understanding.
of the person who continues to communicate through argument—a coordination that discourse ethics presupposes. Moreover, it will fail to occur in precisely those cases in which problematic normative claims to validity are thematized. In such cases the sociocognitive inventory of the conventional stage of interaction can be said to have been only partly reorganized in the sense that while the adolescent has learned how to reason theoretically, he stops short of moral argumentation. In another context I formulated this hypothesis in the following way: by acquiring the ability to think hypothetically about moral-practical issues, the adolescent fulfills the necessary and sufficient condition for dissociating himself from the conventional mode of thought.58

But taking this step does not predetermine his decision between two alternative developmental paths. There are different ways in which the adolescent can use this newly acquired detachment from a world of conventions that have lost the naive force of social acceptance by being hypothetically relegated to a horizon of possibilities and have thereby become reflexively devalued. One alternative is that on his new level of reflection he can try to preserve something from that lost world of factually accepted conventions, namely what it means for norms and prescriptive statements to have validity. If he does that, he must reconstruct the basic concepts of morality without giving up the ethical perspective. He must relativize the de facto social currency of existing norms in terms of a normative validity that satisfies the criteria of rational justification. Maintaining a sense of normative validity reconstructed in this way is a necessary condition for the transition to the postconventional mode of thought. This is one path the adolescent can take.

Alternatively, the adolescent can dissociate himself from the conventional mode of thought without making the transition to the postconventional one. In this case he views the collapse of the world of conventions as a debunking of false cognitive claims with which conventional norms and prescriptive statements have hitherto been linked. The basic moral concepts in their cognitively devalued conventional form then appear retrospectively as requiring explanation. The adolescent’s task is to come to terms with the dissonance between his moral intuitions, which continue to determine his unreflective everyday knowledge and actions, and his (presumed) insight into the illusory nature of this conventional moral consciousness, which reflection has discredited but which has not ceased to function in daily life. For the second developmental path, metaethical explanation of moral illusions takes the place of a postconventional renewal of ethical consciousness. The more successfully theoretical skepticism has been reconciled with the intuitions that go unchallenged in practice, the more easily such explanation can handle the dissonances. In this regard Weber’s ethical skepticism, for example, has greater efficacy than Stevenson’s emotivism. The former leaves untouched the existential nature of value bonds, whereas the latter explains moral intuitions away by reducing them to emotional dispositions and attitudes. From the viewpoint of Kohlberg’s theory these metaethical versions must submit to being classified in terms of a logic of development and being subordinated to cognitivist ethics.

The last problem is one that Kohlberg’s theory shares with any approach that distinguishes competence from performance. Such theoretical approaches face specific measurement problems because competence can be captured only in its concrete manifestations, that is, only in performance. Only insofar as these measurement problems have been solved can we isolate factors determining performance from theoretically postulated competences. It may be helpful to distinguish factors determining performance that, as stimulators or accelerators, must supplement or can accompany an acquired competence from the braking and inhibiting factors that act as filters.

To consider moral judgment as an indicator of competence and moral action or behavior as an indicator of performance is of course a crude simplification. On the other hand, the motivational anchoring of the capacity for postconventional judgment in homologous superego structures does represent
an example of supplementary performance-determining factors without which moral judgments at this level could not become effective in practice.\textsuperscript{60} As a rule, discrepancies between judgment and action can be accounted for in terms of the selective effect of inhibiting factors. A number of interesting studies point in this direction.\textsuperscript{70} Among the performance-determining factors that act as inhibitors are some that explain motivational deficits. Of these the defense mechanisms first systematically studied by Anna Freud are of particular interest because they interfere with a process of motive formation that is structurally necessary. Accordingly, they can be analyzed from a structural perspective.

Identification and projection are the two fundamental mechanisms of defense against conflict. The individual acquires them in early childhood. Only later, at the conventional stage of interaction, do they seem to develop into the familiar system of defense mechanisms.\textsuperscript{71} Defense mechanisms differ in terms of the ways in which they undermine the differentiation between action oriented toward success and action oriented toward reaching understanding that emerges at this level. Generally, the way defense works is that barriers to communication are set up in the psyche, separating the (unconscious) strategic aspect of action (which serves the gratification of unconscious desires) from the manifest intentional action that aims at reaching understanding. This explains why the subject can deceive himself about the fact that he is objectively violating the shared presuppositions of action oriented toward reaching understanding. Unconsciously motivated actions can be explained as a latent reversal of the differentiation between strategic and communicative action, a dedifferentiation that is hidden from the actor and others. The self-deceptive effect of the defense can be interpreted as an intrapsychic disturbance of communication. This interpretation makes use of the conception of systematically distorted communication that can manifest itself in similar ways on two different levels: the interpersonal and the intrapsychic. But this concept requires independent discussion in the framework of communication theory.\textsuperscript{79}


22. This oversimplified opposition disregards the differences between those parts of the lifeworld that have never been dissociated from intuitive background knowledge and thematized, and those that have been focused upon at least once only to sink back into the lifeworld, there to be reabsorbed once more. The latter sort of taken-for-grantedness is a second-order immediacy. I owe this insight to U. Matthiesen.

23. A corresponding hypothesis concerning the construction of an inner world and how it is delimited from the objective and social worlds need not concern us here except insofar as the subjective world and its thematizable experiences represent a further basic attitude and perspective, rounding out the system of world perspectives.

24. I am ignoring stage 0, where the child does not differentiate in a way that would be relevant for us. I am also ignoring stage 4. Stage 4 already presupposes the concept of an action norm. I will argue below that this concept derives reconstruction in terms of perspective taking alone and requires instead sociocognitive concepts of a different provenance. Selman is unable to differentiate stages 3 and 4 solely in terms of perspective taking.


26. Age indicators vary with the situation being investigated. Placed in their natural environment, children in Western societies today turn out to acquire the corresponding competences earlier.

27. The link between possessive-pronoun use and action perspectives is discussed by K. Böhme, Children’s Understanding and Awareness of German Possessive Pronouns (Nijmegen, 1983), pp. 156ff.


31. “Holly is an 8-year-old girl who likes to climb trees. She is the best tree climber in the neighborhood. One day while climbing down from a tall tree she falls off the
wish in the expectation that A for his part will obey B's imperative that q. By taking over B's expectation of how A will behave, A acquires the concept of a pattern of behavior. This concept conditionally joins the particular interlocking and complementary behavioral expectations of A and B.

41. On what follows, see Habermas (1987), vol. 2, pp. 31ff.


46. Kohlberg has emphasized that the construction of stage 6 resulted from material obtained from a small elite sample (it included statements by Martin Luther King). Such elite figures do not establish stage 6 as a natural stage of development. See C. Gilligan and J. M. Murphy, "Kohlberg's Stages of Moral Development," ms. (Cambridge, 1980).


53. W. B. Perry, Forms of Intellectual and Ethical Development in the College Years (New York, 1968).


56. The general problem of applying norms to situations of action is already posed at the conventional stage of moral judgment and interaction. In this context the focus is on the particular aggravation this problem undergoes when the links are severed through which norms and situations of action, as parts of one and the same problematic form of life, refer to each other through their prior coordination. See H. G. Gadamer, Truth and Method (New York, 1975).


60. This is the case with the decisions about abortion studied by Gilligan. Possible repercussions for one's relation to friend or husband, for the occupational careers of man and woman, for family life can be considered important only when abortion itself has been accepted as morally licit. The same goes for problems of divorce and adultery. This is confirmed by the two cases that Gilligan and Murphy (1981) refer to. Only when sexual infidelity is morally unobjectionable can the problem arise of under what conditions concealment of the facts is less harmful to or more considerate of the party directly or indirectly concerned than immediate disclosure.


62. In terms of his moral theory the young Hegel was still a Kantian when he worked out the historical dichotomy between a Christian ethics of love and a Jewish ethics of law and justice.


64. To that extent Kohlberg and Candee (1983) assign too great a burden of proof to "responsibility judgments."


66. See pp. 36–41 in this volume.


Moral Consciousness and Communicative Action


Morality and Ethical Life:
Does Hegel’s Critique of Kant
Apply to Discourse Ethics?

In recent years Karl-Otto Apel and I have begun to reformulate Kant’s ethics by grounding moral norms in communication, a venture to which I refer as “discourse ethics.”¹ In this paper I hope to accomplish two things: first, to sketch the basic idea of discourse ethics and then to examine Hegel’s critique of Kantian moral philosophy. In part I, I will deal with two questions: What is discourse ethics? and What moral intuitions does discourse ethics conceptualize? I will address the complicated matter of how to justify discourse ethics only in passing.

In part II, I will turn to the question of whether Hegel’s critique of Kantian ethics applies to discourse ethics as well. The criticisms Hegel leveled against Kant as a moral philosopher are many. From among them I will single out four which strike me as the most trenchant. These are as follows:

• Hegel’s objection to the formalism of Kantian ethics. Since the moral principle of the categorical imperative requires that the moral agent abstract from the concrete content of duties and maxims, its application necessarily leads to tautological judgments.²

• Hegel’s objection to the abstract universalism of Kantian ethics. Since the categorical imperative enjoins separating the universal from the particular, a judgment considered valid in terms of that principle necessarily remains external to individual cases and insensitive to the particular context of a problem in need of solution.³