"Discourse Ethics" (excerpt) by Jürgen Habermas

I will not repeat the analysis of action oriented to reaching understanding that I have presented elsewhere,[6] but I would like to review one fundamental idea. I call interactions *communicative* when the participants coordinate their plans of action consensually, with the agreement reached at any point being evaluated in terms of the intersubjective recognition of validity claims. In cases where agreement is reached through explicit linguistic processes, the actors make three different claims to validity in their speech acts as they come to an agreement with one another about something. Those claims are claims to truth, claims to rightness, and claims to truthfulness, according to whether the speaker refers to something in the objective world (as the totality of existing states of affairs), to something in the shared social world (as the totality of the legitimately regulated interpersonal relationships of a social group), or to something in his own subjective world (as the totality of experiences to which one has privileged access). Further, I distinguish between communicative and strategic action. Whereas in strategic action one actor seeks to *influence* the behavior of another by means of the threat of sanctions or the prospect of gratification in order to *cause* the interaction to continue as the first actor desires, in communicative action one actor seeks *rationally to motivate* another by relying on the illocutionary binding/bonding effect (*Bindungseffekt*) of the offer contained in his speech act.

The fact that a speaker can rationally motivate a hearer to accept such an offer is due not to the validity of what he says but to the speaker's guarantee that he will, if necessary, make efforts to redeem the claim that the hearer has accepted. It is this guarantee that effects the coordination between speaker and hearer. In the case of claims to truth or rightness, the speaker can redeem his guarantee discursively, that is, by adducing reasons; in the case of claims to truthfulness he does so through consistent behavior. (A person can convince someone that he means what he says only through his actions, not by giving reasons.) As soon as the hearer accepts the guarantee offered by the speaker, obligations are assumed that have consequences for the interaction, obligations that are contained in the meaning of what was said. In the case of orders and directives, for instance, the obligations to act hold primarily for the hearer, in the case of promises and announcements, they hold for the speaker, in the case of agreements

[6] Jürgen Habermas, "Social Action, Purposive Activity, and Communication," in *The Theory of Communicative Action*, vol. 1 (Boston, 1984), pp. 273 ff.

and contracts, they are symmetrical, holding for both parties, and in the case of substantive normative recommendations and warnings, they hold asymmetrically for both parties.

Unlike these regulative speech acts, the meaning of a constative speech act gives rise to obligations only insofar as the speaker and the hearer agree to base their actions on situational definitions that do not contradict the propositions they accept as true at any given point. Obligations to act flow directly from the meaning of an expressive speech act in that the speaker specifies what it is that his behavior does not contradict and will not contradict in the future. Owing to the fact that communication oriented to reaching understanding has a validity basis, a speaker can persuade a hearer to accept a speech-act offer by guaranteeing that he will redeem a criticizable validity claim. In so doing, he creates a binding/bonding effect between speaker and hearer that makes the continuation of their interaction possible.

The two *discursively redeemable* claims to validity that are of particular interest to us, claims to propositional truth and claims to normative rightness, play their roles as coordinators of action in different ways. A number of asymmetries between them suggest that they occupy different "positions" in the communicative practice of everyday life.

On the face of it, *assertoric statements* used in *constative speech acts* appear to be related to *facts* as *normative statements* used in *regulative speech acts* are related to *legitimately ordered interpersonal relations*. The *truth* of propositions seems to signify the *existence* of states of affairs in much the same way as the *rightness* of actions signifies the *observance* of norms. If we look at the matter more closely, however, we notice some interesting differences. The relation of speech acts to norms is not the same as the relation of speech acts to facts. Let us look at the case of moral norms that can be formulated in terms of universal "ought" sentences or commandments:

(a) One ought not to kill anybody.
(a') It is commanded not to kill anybody.

We make reference to norms of action of the above kind in regulative speech acts, and we do so in a variety of ways: by giving orders, making contracts, opening meetings, issuing warnings, granting exceptions, giving advice, etc. A moral norm, however, lays claim to meaning and validity regardless of whether it is promulgated or made use of in a specific way. A norm may be formulated in a statement like (a), but this act of formulating it, i.e., of writing a sentence, *need not* itself be conceived of as a speech act, that is, as something other than the impersonal expression of the norm. Statements such as (a) are commands that we can address *secondarily* in one way or another through speech acts. This has no equivalent in the domain of facts. There are no assertoric propositions that have an existence

independent of speech acts, as norms do. If such assertoric statements are to have pragmatic meaning at all, they must be used in a speech act. Unlike sentences (a) and (a'), descriptive statements such as "Iron is magnetic" or "It is the case that iron is magnetic" cannot be expressed or used independently of the illocutionary role of a certain type of speech act if they are to retain their assertoric power.

We can account for this asymmetry by saying that claims to truth reside *only* in speech acts, whereas the locus of normative claims to validity is primarily in norms and only derivatively in speech acts.[7] To use an ontological mode of expression, we might say that this asymmetry is due to the fact that the orders of society, which we either conform to or deviate from, are not constituted *independently of validity,* as are the orders of nature, toward which we can assume an objectivating attitude. The social reality that we address in our regulative speech acts has by its very nature an *intrinsic* link to normative validity claims. Claims to truth, on the other hand, have no such intrinsic link to entities; they are inherently related only to the constative speech acts by which we refer to entities when we use fact-stating locutions to represent states of affairs.

Owing to the fact that normative validity claims are built into the universe of norms, the latter reveals a peculiar kind of objectivity vis-à-vis regulative speech acts, an objectivity that the universe of facts does not possess vis-à-vis constative speech acts. To be sure, "objectivity" in this connection refers only to the independence of "objective spirit," for entities and facts are, of course, independent in a completely different sense than is everything we consider part of the social world when we take a norm-conformative attitude. For example, norms are dependent upon the continual reestablishment of legitimately ordered interpersonal relationships. They would assume a utopian character in the negative sense and lose their very meaning if we did not *complement* them, at least in our minds, with actors who might follow them and actions that might fulfill them. States of affairs, for their part, must be assumed to exist independently of whether we formulate them by means of true propositions or not.

Normative claims to validity, then, *mediate a mutual dependence* of language and the social world that does not exist for the relation of language to the objective world. This interlocking of claims to validity that reside in norms and claims to validity raised in regulative speech acts is also connected with the *ambiguous nature* of *normative validity.* While there is an unequivocal relation between existing states of affairs and true propositions about them, the "existence" or social currency of norms says nothing about whether the norms are valid. We must distinguish between the social fact that a norm is intersubjectively recognized and its

---

[7]   At most we can compare theories, as higher-level systems of propositions, with norms. But it is debatable whether theories can be said to be true or false in the same sense as the descriptions, predictions, and explanations that we derive from them, whereas norms for their part are right or wrong in the same sense as the actions that satisfy or violate them.

worthiness to be recognized. There may be good reasons to consider the validity claim raised in a socially accepted norm to be unjustified. Conversely, a norm whose claim to validity is in fact redeemable does not necessarily meet with actual recognition or approval. Gaining acceptance on the part of a norm is encoded in a twofold fashion because our motives for recognizing normative claims to validity are rooted both in convictions and in sanctions, that is, they derive from a complex mixture of rational insight and force. Typically, rationally motivated assent will be combined with empirical *acquiescence,* effected by weapons or goods, to form a belief in legitimacy whose component parts are difficult to isolate. Such alloys are interesting in that they indicate that a positivistic enactment of norms is not sufficient to secure their *lasting* social acceptance. Enduring acceptance of a norm *also* depends on whether, in a given context of tradition, reasons for obedience can be mobilized, reasons that suffice to make the corresponding validity claim at least appear justified in the eyes of those concerned. Applied to modern societies, this means that there is no mass loyalty without legitimacy.[8] [. . .]

Just as an impartial process of judging is not guaranteed by an empirical check to see that allowance for disagreement has been made, so a norm cannot be considered the expression of the common interest of all who might be affected simply because it seems acceptable to some of them under the condition that it be applied in a nondiscriminatory fashion. The intuition expressed in the idea of the generalizability of maxims intends something more than this, namely, that valid norms must *deserve* recognition by *all* concerned. It is not sufficient, therefore, for *one person* to test whether he can will the adoption of a contested norm after considering the consequences and the side effects that would occur if all persons followed that norm or whether every other person in an identical position could will the adoption of such a norm. In both cases the process of judging is relative to the vantage point and perspective of *some* and not *all* concerned. True impartiality pertains only to that standpoint from which one can generalize precisely those norms that can count on universal assent because they perceptibly embody an interest common to all affected. It is these norms that deserve inter-subjective recognition. Thus the impartiality of judgment is expressed in a principle that constrains all affected to adopt the perspectives of *all others* in the balancing of interests. The principle of universalization is intended to compel the *universal exchange of roles* that G. H. Mead called "ideal role taking" or "universal discourse."[9] Thus every

---

[8]   J. Habermas, "Legitimation Problems in the Modern State," in Jürgen Habermas, *Communication and the Evolution of Society* (Boston, 1979), pp. 178 ff. On the relation between the justification of norms, their being put into effect, and their being accepted, see also W. Kuhlmann, "Ist eine philosophische Letztbegründung von Normen möglich?" in *Funkkolleg Ethik,* Studienbegleitbrief 8 (Weinheim, 1981), p. 32.

[9]   G. H. Mead, "Fragments on Ethics," in *Mind, Self, and Society* (Chicago, 1934), pp. 379 ff. See also H. Joas, *G. H. Mead: A Contemporary Reexamination of His Thought* (Cambridge, Mass., 1985), and J. Habermas, *The Theory of Communicative Action,* vol. 2 (Boston, 1987), pp. 92 ff.

valid norm has to fulfill the following condition:

> (U) All affected can accept the consequences and the side effects its *general* observance can be anticipated to have for the satisfaction of *everyone's* interests (and these consequences are preferred to those of known alternative possibilities for regulation).[10]

We should not mistake this principle of universalization (U) for the following principle, which already contains the distinctive idea of an ethics of discourse.

> (D) Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse.*

This principle of discourse ethics (D), to which I will return after offering my justification for (U), already *presupposes* that we *can* justify our choice of a norm. At this point in my argument, that presupposition is what is at issue. I have introduced (U) as a rule of argumentation that makes agreement in practical discourses possible whenever matters of concern to all are open to regulation in the equal interest of everyone. Once this bridging principle has been justified, we will be able to make the transition to discourse ethics. I have formulated (U) in a way that precludes a monological application of the principle. First, (U) regulates only argumentation among a plurality of participants; second, it suggests the perspective of real-life argumentation, in which all affected are admitted as participants. In this respect my universalization principle differs from the one John Rawls proposes.

Rawls wants to ensure impartial consideration of all affected interests by putting the moral judge into a fictitious "original position," where differences of power are eliminated, equal freedoms for all are guaranteed, and the individual is left in a condition of ignorance with regard to the position he might occupy in a future social order. Like Kant, Rawls operationalizes the standpoint of impartiality in such a way that every individual can undertake to justify basic norms on his own. The same holds for the moral philosopher himself. It is only logical, therefore, that Rawls views the substantive parts of his study (e.g., the principle of average utility), not as the *contribution* of a participant in argumentation to a process of discursive will

---

[10] With reference to Bernard Gert's *Moral Rules* (New York, 1976), p. 72, G. Nunner-Winkler has raised the objection that (U) is unable to single out from among the norms that fulfill the stated conditions those that are moral in the narrow sense and to exclude others (e.g., "You ought to smile when you say hello to people"). This objection is met when one proposes to call moral only those norms that are strictly universalizable, i.e., those that are invariable over historical time and across social groups. This usage of the moral theorist does not, of course, coincide with that of the sociologist and the historian, who tend to describe epoch-specific and culture-specific rules as moral rules if they are accepted as such by the members of the group under study.

formation regarding the basic institutions of late capitalist society, but as the outcome of a "theory of justice," which he as an expert is qualified to construct.

If we keep in mind the action-coordinating function that normative validity claims play in the communicative practice of everyday life, we see why the problems to be resolved in moral argumentation cannot be handled monologically but require a cooperative effort. By entering into a process of moral argumentation, the participants continue their communicative action in a reflexive attitude with the aim of restoring a consensus that has been disrupted. Moral argumentation thus serves to settle conflicts of action by consensual means. Conflicts in the domain of norm-guided interactions can be traced directly to some disruption of a normative consensus. Repairing a disrupted consensus can mean one of two things: restoring intersubjective recognition of a validity claim after it has become controversial or assuring intersubjective recognition for a new validity claim that is a substitute for the old one. Agreement of this kind expresses a *common will.* If moral argumentation is to produce this kind of agreement, however, it is not enough for the individual to reflect on whether he can assent to a norm. It is not even enough for each individual to reflect in this way and then to register his vote. What is needed is a "real" process of argumentation in which the individuals concerned cooperate. Only an intersubjective process of reaching understanding can produce an agreement that is reflexive in nature; only it can give the participants the knowledge that they have collectively become convinced of something.

From this viewpoint, the categorical imperative needs to be reformulated as follows: "Rather than ascribing as valid to all others any maxim that I can will to be a universal law, I must submit my maxim to all others for purposes of discursively testing its claim to universality. The emphasis shifts from what each can will without contradiction to be a general law, to what all can will in agreement to be a universal norm."[11] This version of the universality principle does in fact entail the idea of a cooperative process of argumentation. For one thing, nothing better prevents others from perspectivally distorting one's own interests than actual participation. It is in this pragmatic sense that the individual is the last court of appeal for judging what is in his best interest. On the other hand, the descriptive terms in which each individual perceives his interests must be open to criticism by others. Needs and wants are interpreted in the light of cultural values. Since cultural values are always components of intersubjectively shared traditions, the revision of the values used to interpret needs and wants cannot be a matter for individuals to handle monologically.[12]

[11] Thomas McCarthy, *The Critical Theory of Jürgen Habermas* (Cambridge, Mass., 1978), p. 326.

[12] Seyla Benhabib, "The Methodological Illusions of Modern Political Theory: The Case of Rawls and Habermas," *Neue Hefte für Philosophie* 21 (1982), 47 ff.

Discourse ethics, then, stands or falls with two assumptions:

(a) that normative claims to validity have cognitive meaning and can be treated like claims to truth and

(b) that the justification of norms and commands requires that a real discourse be carried out and thus cannot occur in a strictly monological form, i.e., in the form of a hypothetical process of argumentation occurring in the individual mind. [. . .]

The fact that the transcendental-pragmatic strategy of justification depends on the objections of a skeptic is not wholly to its advantage. Such arguments are telling only with an opponent who does his proponent the favor of entering into an argumentation. A skeptic who sees in advance that he will be caught in performative contradictions will reject the game of wits from the outset. The *consistent skeptic* will deprive the transcendental pragmatist of a basis for his argument. He may, for example, take the attitude of an ethnologist vis-à-vis his own culture, shaking his head over philosophical argumentation as though he were witnessing the unintelligible rites of a strange tribe. Nietzsche perfected this way of looking at philosophical matters, and Foucault has now rehabilitated it. When this happens, the debate suddenly changes. If the cognitivist persists in his analysis, he will now be talking only *about* the skeptic, not *with* him. At this point the cognitivist usually throws up his hands, confessing he has no further remedy for this dropout posture of the skeptic. He will say that a willingness to argue and to think about one's actions must really be presupposed if the whole concern of moral theory is not to become pointless. There may be, he will continue, a residue of decisionism that cannot be disproved by argumentation. This, he will grant, is where the volitional moment comes into its own.

It seems to me that the moral theorist ought not to leave it at that. A skeptic who could wrest the topic from the cognitivist's grip merely with his behavior might not have the last word, but he would be performatively right, so to speak—he would assert his position mutely and impressively.

At this juncture in the discussion, if we can still call it that, it helps to keep in mind that through his behavior the skeptic voluntarily terminates his membership in the community of beings who argue—no less and no more. By refusing to argue, for instance, he cannot, even indirectly, deny that he moves in a shared sociocultural form of life, that he grew up in a web of communicative action, and that he reproduces his life in that web. In a word, the skeptic may reject morality, but he cannot reject the ethical substance (*Sittlichkeit*) of the life circumstances in which he spends his waking hours, not unless he is willing to take refuge in suicide or serious mental illness. In other words, he cannot extricate himself from the communicative practice of everyday life in which he is continually forced to take a position by responding yes or no. As long as he is still alive *at all,* a Robinson Crusoe existence through which the skeptic demonstrates mutely and impressively

that he has dropped out of communicative action is inconceivable, even as a thought experiment.

As we have seen, in reaching an understanding about something in the world, subjects engaged in communicative action orient themselves to validity claims, including assertoric and normative validity claims. This is why there is no form of sociocultural life that is not at least implicitly geared to maintaining communicative action by means of argument, be the actual form of argumentation ever so rudimentary and the institutionalization of discursive consensus building ever so inchoate. Once argumentation is conceived as a special form of rule-governed interaction, it reveals itself to be a reflective form of action oriented toward reaching an understanding. Argumentation derives the pragmatic presuppositions we found at the procedural level from the presuppositions of communicative action. The reciprocities undergirding the mutual recognition of competent subjects are already built into action oriented toward reaching an understanding, the action in which argumentation is *rooted*. That is why the radical skeptic's refusal to argue is an empty gesture. No matter how consistent a dropout he may be, he cannot drop out of the communicative practice of everyday life, to the presuppositions of which he remains bound. And these in turn are at least partly identical with the presuppositions of argumentation as such.