

Philosophy & Social Criticism

<http://psc.sagepub.com>

Review essay: Communicative action and rational choice

Cristina Lafont

Philosophy Social Criticism 2005; 31; 253

DOI: 10.1177/0191453705050610

The online version of this article can be found at:

<http://psc.sagepub.com>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Philosophy & Social Criticism* can be found at:

Email Alerts: <http://psc.sagepub.com/cgi/alerts>

Subscriptions: <http://psc.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>

Cristina Lafont

Review essay

Communicative action and rational choice

Joseph Heath, *Communicative Action and Rational Choice*
(Cambridge, MA: MIT Press, 2001), 363 pp.

This book offers an excellent analysis of Habermas's theory of communicative action. It has two distinct but complementary focuses. In the first part, the conception of communicative rationality at the basis of Habermas's theory of action is confronted with the conception of instrumental rationality that is predominant in the social sciences: rational choice theory. The main focus of this analysis is to evaluate the plausibility of one central claim of Habermas's theory, namely, that communicative rationality is irreducible to instrumental rationality and thus constitutes a necessary element of a general theory of rationality. Although Heath's analysis confirms the correctness of Habermas's claim, it does so on the basis of a command of decision and game theory, the sophistication of which goes well beyond anything that Habermas himself (and most of his commentators) have so far actually provided. This is undoubtedly one of the most important achievements of the book. For it offers the so far missing argumentative step from the intuitiveness of Habermas's irreducibility claim to a careful demonstration of what it is exactly about linguistic communication that cannot be modeled instrumentally. Moreover, this careful demonstration includes a very interesting and equally sophisticated analysis of the implications of Habermas's theory for the philosophy of language, which so far has received too little attention by most of Habermas's commentators.

PHILOSOPHY & SOCIAL CRITICISM • vol 31 no 2 • pp. 253–263

Copyright © 2005 SAGE Publications (London, Thousand Oaks, CA and New Delhi)

www.sagepublications.com DOI: 10.1177/0191453705050610

PSC

In the second part, Habermas's conception of communicative rationality is analyzed not from the sociological point of view of its implications for a theory of action but from the philosophical point of view of its implications for a moral theory. In particular, the sociological insights obtained in the first part of the book are put to work to criticize the universalist aspirations of Habermas's discourse ethics. Here the results are far more controversial than those of the first part of the book. On the one hand, Heath's analysis of discourse ethics is by no means purely exegetical and in fact offers an alternative approach to moral theory that it is worth considering on its own merits. Both features of this analysis are surely very interesting. On the other hand, however, it is hard to avoid the impression that in one important respect this analysis of discourse ethics fails to take into account the core aspirations that make the theory worth pursuing in the first place. To the extent that Heath's analysis remains somewhat external to those aspirations it seems that its criticisms, if they turned out to be true (which is, of course, an open question), could at most have a negative impact on discourse ethics. It is hard to see how they could have the positive effect of showing a better way of pursuing Habermas's own project in the way that the analysis of the first part of the book clearly does.

One of the main argumentative intentions of the first part of the book is to reverse the order of explanation favored by Habermas in his defense of communicative rationality. Although Heath agrees with Habermas's central claim that speech-acts cannot be modeled instrumentally, he disagrees with his conclusion, namely, that the accountability characteristic of social action is therefore due to the fact that language is used as a coordination device. In Heath's opinion the explanation goes in the exact opposite direction. It is because the accountability of speech-acts is essential for securing their intelligibility that they cannot be modeled instrumentally. And this in turn is so because they are a species of norm-governed action, a kind of action that has accountability as its primitive feature. Thus, the basic types of action, according to Heath, are not instrumental and communicative action, as Habermas claims, but instrumental and norm-conformative action. This general argument is built in four steps that correspond to the first four chapters of the book.

In chapter 1 Heath offers a short but illuminating outline of Habermas's theory with a very specific purpose in mind, namely, to spell out the argumentative commitments that Habermas's conception of communicative action assumes and thus that need to be justified in order to defend that conception. These commitments provide the structure for the following three chapters. First, it must be shown that communication cannot be purely instrumental. In the context of Habermas's theory this requires showing that the instrumental model of rationality is unable to provide an adequate pragmatics for a theory of meaning.

Second, to sustain this claim some explanation must be given as to why speech-acts cannot be planned and executed with instrumental intent. This requires showing that a theory of meaning for speech-acts can be given through some form of acceptability-conditional semantics. For if this were shown, then it could be argued that since the meaning of speech-acts is given by the discursive commitments that are generated through their utterance, and since commitments involve a failure of sequential rationality, which is an essential feature of the instrumental conception of rationality, language use cannot be instrumentally rational. Third, some alternative conception of rationality must be provided to explain the specific sense in which linguistic communication can be considered rational. Here the notion of discursive commitment used to generate the semantic theory could be generalized into a discursive conception of rational acceptability (as a non-foundationalist theory of justification). Finally, it would have to be shown that all actions coordinated through linguistic communication are governed by standards of communicative rather than instrumental rationality. The overall strategy of Heath's argument is to grant the validity of the first three claims (although at times for reasons different than Habermas's own) and to reject the last one.

Accordingly, the focus of chapter 2 is the central claim behind Habermas's conception of communicative rationality, namely, that communication cannot be purely instrumental. Here, however, Heath does not follow Habermas's own line of argument. Instead of building on Habermas's theory of speech-acts (with all of its difficulties in providing a convincing account of the problematic distinction between illocutionary and perlocutionary acts, etc.), Heath offers a very convincing analysis of the internal difficulties that the standard conception of instrumental rationality faces whenever it tries to introduce communication into instrumental models of social interaction. As I mentioned before, this analysis is undoubtedly one of the greatest achievements of the book: not so much perhaps from the point of view of its novelty in the context of the literature on rational choice itself, but certainly for its ability to translate and explain some very interesting and revealing problems of the latter into the context of the literature on critical theory. In a nutshell, Heath's argumentative strategy is the following: in order to show that communication cannot be modeled as a form of strategic action, he discusses the two strategies available within game theory to model linguistic communication. One strategy is to explain the meaning (i.e. semantic content) of linguistic utterances in terms of their pragmatic role in strategic interactions. Here Heath convincingly shows that on this basis it is not possible to confer determinate, fixed meanings to linguistic utterances. In a strategic setting, linguistic messages are essentially meaningless. If this is so, game theory cannot explain how

meaningful communication could develop among instrumentally rational agents. The other strategy takes semantic content as primitive (i.e. assumes that linguistic utterances already have fixed meanings) and tries to explain speech-acts as a type of instrumental action. Here the strategy fails because it cannot provide any game-theoretical solutions to the resulting coordination problem. In a communicative setting, agents cannot determine a strategically rational course of action. One way or the other, the conclusion that 'only no communication is compatible with strategically rational behavior' seems to impose itself upon us.

On the basis of these results, Heath can briefly show at the beginning of Chapter 3 that strategic rationality is indeed incompatible with commitment, and then move on to analyze the second claim of Habermas's general argument that we mentioned above, namely, that discursive commitment is essential for linguistic communication. In this context, Heath analyzes Habermas's theory of language by focusing on Habermas's central claim that there is an intrinsic connection between meaning and validity (and thus that a theory of meaning should be constructed along the lines of an acceptability-conditional semantics). Here, again, Heath builds his own argument in favor of Habermas's claim on the basis of Brandom's defense of inferential semantics as the right model for a pragmatic theory of meaning for natural languages. Regardless of the actual feasibility of the latter model, its discussion in this context proves very helpful in showing some of the difficulties of Habermas's own model; in particular, those related to Habermas's thesis that all types of speech-acts are associated with three universal validity-claims. Of the numerous difficulties associated with this thesis, Heath focuses on two. One is the ambiguity in Habermas's explanation of what the exact relationship between these validity-claims and the meaning of speech-acts is. After showing that the explanation initially put forward by Habermas (namely, merely to postulate a connection between validity-claims and the speech-acts *qua* acts) cannot but fail in view of the theory's overall goal, Heath focuses on a second difficulty that arises from following the alternative explanation (namely, to consider all three validity-claims as constitutive of the meaning of the propositional content of speech-acts). Here Heath convincingly argues that postulating a validity-claim called 'normative rightness' that in the context of imperatives is supposed to function analogously to the way 'truth' functions for assertions (i.e. as a property that is preserved through valid inference and thus functions as a designated value in a system of logic) makes it impossible to explain the uniformity of the semantic and logical features of language. Furthermore, as P. Geach convincingly argued long ago, practical discourses just seem to have the same set of inference rules that theoretical discourses do (they take the form of assertions that can be embedded in conditionals, negations, and other logical connectives

for drawing inferences). And if this is the case, then there seems to be nothing distinctive about practical reasoning other than its topic.

At this point Heath's general argument reaches its climax. In the remainder of the book he will just draw consequences from the argumentative steps taken so far. First of all, in the next chapter he will show why the last argumentative inference that Habermas draws from the three basic commitments discussed so far does not follow. This, in turn, would give indirect support to Heath's own alternative proposal in the sociological context of theories of action. Second, the arguments provided so far would also support the central claim behind the alternative interpretation of discourse ethics that he develops in the second part of the book, namely, that practical discourse is governed by a deflationary truth predicate.

But let us focus first on chapter 4. Here Heath draws out the consequences of assuming that a successful theory of meaning within the framework of an acceptability-conditional semantics would have to be developed along the lines of Brandom's inferential semantics. If such a theory of meaning were feasible, the thesis that linguistic commitment is essential for communication would have been proved. However, this would concern only *assertional* commitments. For, according to this inferentialist view, assertion is the only content-conferring category of linguistic action. Thus, this strategy would be of no help in sustaining Habermas's further claim that all actions coordinated through linguistic communication are governed by standards of communicative rather than instrumental rationality because of the extra-discursive *normative* commitments that they involve.

But no matter whether this specific claim is correct or not, in this context the reader would expect a more in-depth discussion of the *full* consequences of such a change of strategy. On the one hand, it seems that for Heath's own argumentative purposes, a brief comparison of Habermas's and Brandom's theories of language may be sufficient. For in order to problematize the role that the validity-claim 'rightness' is supposed to play in Habermas's approach without at the same time giving up on the pragmatic connection between meaning and validity Brandom's model is clearly useful. On the other hand, though, there is more to Habermas's theory of language than just a defense of a sociological claim about the role of language as a coordination mechanism. Habermas's claim about the internal connection between meaning and validity and his interpretation of acceptability-conditional semantics in strict parallelism with truth-conditional semantics are also his attempt to secure a strong universalist conception of objectivity for both theoretical and practical discourses. In this context, one misses in Heath's discussion an in-depth analysis of the extent to which the radically perspectival conception of objectivity that follows from the strong

inferentialism of Brandom's approach can actually be an adequate means for Habermasian goals. This is certainly a very complex topic that is still under much discussion and whose outcome is uncertain. Thus, it could turn out that, against the suspicions of many commentators (including Habermas), Brandom's model actually contains the needed resources to avoid the relativistic consequences of his strong inferentialism. In fact, Brandom's appropriation of the innovative tools provided by theories of direct reference in his anaphoric conception of reference is explicitly designed to avoid the problem of incommensurability that has haunted theories of meaning and reference ever since the linguistic turn. But, whatever Heath's stand on these issues may be, it is quite surprising not to find any analysis of them precisely in the context of a discussion of what the most appropriate theory of meaning for natural languages may be.

The question of the compatibility between Heath's proposals and Habermas's aspirations that arises in chapter 4 becomes more and more pressing in the second part of the book. As mentioned before, the focus of Heath's discussion here is Habermas's discourse ethics. This discussion follows two main lines of argument. The first of them focuses on the alternative strategy for understanding practical discourses and the validity-claim 'normative rightness' that Heath had briefly mentioned at the end of chapter 4. The aim here is to show that adopting the view that practical discourses are governed by a deflationary truth predicate can solve the problems discussed in the prior chapter. This is a very interesting methodological proposal that offers more promising means to reach the same Habermasian ends. The second line of argument, however, is of a very different nature. The aim here is to question those ends themselves, namely, Habermas's attempt to introduce a universalization principle governing moral discourse.

With regard to the first line of argument, I am too convinced of its general plausibility to find any interesting critical comments to make on it. I can only recommend a careful reading of chapter 5 to anyone interested in the metaethics of Habermas's moral theory. In fact, in a recent response to some of Hilary Putnam's similar criticisms of discourse ethics, Habermas accepts the interpretation of practical discourses in terms of a deflationary view of truth,¹ although he also insists on the need to articulate the different sense of validity that empirical and moral statements express. This change of attitude is certainly motivated by Habermas's retraction of his discourse theory of truth in his book *Truth and Justification*.² Unfortunately, these developments of Habermas's approach have taken place too recently to be discussed in Heath's book. But, although their full consequences have not yet become entirely clear, they certainly bring Habermas's approach closer to his diagnosis.

The second line of argument is developed in chapters 6 and 7. The main target of Heath's criticism here is Habermas's principle of universalization (U). But, as mentioned before, the nature of the criticism is qualitatively different from any other in the book. For the general strategy here is not to discuss the difficulties of Habermas's account of the principle in order to offer a better, alternative account. To the contrary, as he states very clearly right at the beginning of chapter 6, the very attempt to introduce a universalization principle governing moral discourse 'should, in my view, be retracted' (p. 220). Of course, this claim immediately raises the question of what would be left of discourse ethics, if the principle of universalization were not just reformulated in one way or another, but rejected altogether. Let us see what the answer to this question may be in light of Heath's general argument.

In what looks like a clear understatement, Heath recognizes that Habermas's principle of universalization 'occupies a central role in his discourse ethics' (ibid.). But the use of the indefinite article in this context clearly suggests that, in Heath's view, this principle is not the very core of discourse ethics. Thus, there must be some further goal toward which the principle of universalization is supposed to be a means. Heath's own summary of the chapter's conclusion in this context provides an initial clue as to what that goal may be. He remarks: 'I conclude by arguing that Habermas's concern to show that agreement is possible in principle leads him to devalue a more robust mechanism for overcoming disagreement, namely, bargaining and compromise' (ibid.).

Regardless of its correctness, this claim provides an indication of the scope and direction of Heath's analysis and criticism of discourse ethics in the coming chapters. From the sociological point of view that has structured the discussion of Habermas's theory of communicative action so far, discourse ethics appear to be a further means for the sociological goal of explaining *social integration*, that is, for finding 'mechanisms for overcoming disagreement'. Accordingly, moral theory answers to social theory. Thus if moral consensus turns out to be a less robust mechanism for overcoming disagreement than bargaining and compromise, so much the worse for moral consensus. From a purely exegetical point of view, this interpretation of the order of explanation that structures Habermas's general approach seems problematic. But regardless of what the exact internal connection between Habermas's own goals and aspirations may in fact be, it is clear that if discourse ethics has any chance to stand on its own *as a moral theory*, it will have to be in virtue of its capacity to explain *moral discourses*. From the internal perspective of moral philosophy, it may seem somewhat beside the point to claim that beyond moral discourses there are other, more effective

mechanisms of social integration one can focus on. Even if this were true, it would by no means replace the task of providing a successful explanation of the consistency and tenability of the moral intuitions that guide moral discourses. This certainly seems to be a meaningful task that can be pursued for its own sake.

As should be obvious, though, this evaluation of the specific nature of Heath's criticisms of discourse ethics does not say anything about the quality or relevance of those criticisms, even within moral philosophy. With regard to this question, I find many of the specific analyses and criticisms of chapter 6 illuminating and surely worth serious consideration by anyone interested in discourse ethics. Others, however, may seem only to reinforce the standard impression that it is a waste of time to talk morality with a Hobbesian.³ In general, here as in the rest of the book, Heath's strength clearly lies in pointing to the difficulties of Habermas's account of moral rightness from the point of view of his theory of action and his theory of language. Whether the difficulties he discusses from the narrower point of view of moral philosophy are equally convincing is much less clear. On the one hand, his discussion is way too short to settle the kind of issues that he discusses (e.g. the principle of universalization and the moral/ethical distinction). On the other hand, his discussion remains too external to the kind of Kantian approach that discourse ethics represents to be convincing for those who would actually like to pursue the Habermasian project within moral philosophy. This is not surprising, though, given that, as we already mentioned, Heath sees discourse ethics as a means for the broader goal of a theory of action. Thus, for his argumentative purposes, it seems sufficient⁴ to show that this further goal can be achieved at a lower cost.

This is what Heath attempts to show in chapter 7. In this context, the issue is no longer (if it ever was) whether and how our moral intuitions can be squared with a plausible account of moral discourses, but what can be a successful (external) explanation of some of its most striking features, such as the assumption that moral questions have single right answers or that convergence in moral judgments is possible. Heath's answer is based on what he calls 'a pragmatist theory of convergence'. According to this view, convergence of belief is necessary in order to construct the cognitive framework needed to interact instrumentally. It is not a metaphysical but a pragmatic necessity in so far as we want to pursue projects that require instrumental reasoning in contexts of interdependent choice. Similarly, convergence of social norms can be explained just by providing a set of pragmatic reasons (such as massive social interaction, the need for cooperation, the high cost of disagreement, etc.). This in turn makes it possible to offer an alternative to Habermas's account as to why moral discourses are structured in the way they are, without appealing to any intrinsic epistemic

feature of the discourse topic. 'Instead it provides an action-theoretic analysis of the reasons that highly regimented discourses should be favored in particular domains. . . . It can simply be shown that there is no pragmatic alternative to the normative regulation of social interaction, and as a result, there is a powerful incentive for agents to pursue moral argumentation until they reach some agreement – if need be, through compromise or bargaining' (pp. 279–80).

Here we find a mixture of claims and strategies that is hard to disentangle. On the one hand, there is the suggestion that such an account in some sense constitutes an alternative to Habermas's own. Obviously, this can only be so, if both accounts are supposed to respond to the same question. However, the sense in which Habermas's discourse ethics tries to provide an account of why moral discourses 'are structured the way they are' has nothing to do with the 'sociological' view of morality that Heath endorses in his own proposal. From the point of view of Habermas's moral theory, this question does not have the external sense of asking why, given our pragmatic needs, it is *useful* to have moral discourses, but rather the internal sense of asking why it is *sound* or justified to have them, whether they are useful or not. No matter how plausible an answer to the former question may be, it just cannot provide an answer to the latter. This discrepancy is even clearer with regard to Heath's ambiguous suggestion that his strategy offers something like a pragmatic answer to the question, why be moral (i.e. why 'there is a powerful incentive for agents to pursue moral argumentation')? As a sociological claim, it can hardly be an alternative to Habermas's account. He surely agrees that practical discourse can be a less costly mechanism for cooperation than strategic action or open violence. But as a claim within moral philosophy, it cannot be an alternative to Habermas's account either, although for entirely different reasons. With regard to the question Why be moral?, discourse ethics does not contain any answer whatsoever to which Heath's answer could be an alternative. As a good Kantian, Habermas does not consider that a pragmatic answer to this question makes sense. Not so much because being moral for instrumental reasons would not be rational, but because it would not be moral. There is no external justification of morality. This is, however, something that Heath himself seems to claim in chapter 8, where he examines and to some extent endorses the transcendental arguments that Habermas uses in defense of discourse ethics. In the context of explaining the reasons for his endorsement in the concluding remarks, he claims that 'there is no ultimate justification required for a higher-order choice disposition that assigns priority to normative reasons for action, just as there is no ultimate justification required for the rules of discourse' (p. 309). But in the end he still insists that:

Normative regulation serves to counteract the general tendency of human affairs to go very badly when left to self-interest, but it also allows agents to eliminate a vast number of unnecessary coordination failures that can arise from the inability of strategic reasoning to focus expectations on a single set of equilibrium beliefs. (I argued that the same set of considerations will motivate agents to structure their moral arguments in such a way as to promote high levels of convergence.) From this point of view, it is Habermas this time who makes 'unnecessary work' for himself, in attempting to show that the transition to discourse and the expectation of convergence is required as a matter of principle, as opposed to being 'simply functionally recommended'. (p. 310)

But regardless of its specific sense as a claim within moral philosophy, this is surely an important claim in the context of a dialogue between Habermas's theory of communicative action and the theory of rational choice. For in the context of evaluating the relative strengths and explanatory scope of both theoretical strategies, showing that moral discourses are indeed 'functionally recommended' would already show the internal limitations of the latter theory and thus the need for the former. And this is a conclusion that is hard to avoid in light of the excellent arguments that Heath provides in his book.

Northwestern University, Evanston, USA

PSC

Notes

- 1 For Putnam's criticisms see M.-L. Raters and M. Willaschek (eds) *Hilary Putnam und die Tradition des Pragmatismus* (Frankfurt: Suhrkamp, 2002), pp. 306–24, esp. pp. 312–16. For Habermas's reaction to them see 'Response to Putnam's "Concluding Remarks"' (unpublished manuscript, here esp. pp. 8–9).
- 2 Habermas's retraction took place in his article 'Richard Rorty's Pragmatic Turn' from 1996, which was first published in *On the Pragmatics of Communication* (Cambridge, MA: MIT Press, 1998), pp. 343–82. But the full discussion of the retraction was developed in his later book *Wahrheit und Rechtfertigung* (Frankfurt: Suhrkamp, 1999); English-language version *Truth and Justification* (Cambridge, MA: MIT Press, forthcoming). In chapter 5, Heath seems to be a little bit over-charitable with regard to Habermas's discourse conception of truth, when he interprets it as a kind of deflationism and rejects as a misunderstanding any interpretation of the theory as an attempt to explain truth in terms of ideal (discursive) justification. To support his rejection he remarks: 'this obviously cannot be Habermas's view, since it is inconsistent with the illocutionary-act analysis of the truth predicate (not to mention Schema T)' (p. 214). As an explanation

of some of the reasons behind Habermas's retraction of his epistemic view of truth as ideal (discursive) justification, Heath's remark can be seen as correct *in retrospect*. However, as an exegetical claim about Habermas's conception of truth *prior to the retraction*, Heath's interpretation is problematic. For it renders Habermas's retraction unintelligible. In a footnote, though, Heath recognizes that Habermas does not accept deflationism and offers a more plausible interpretation of the issue, when he remarks that 'Habermas officially distances himself from deflationism, but his reasons for doing so are, in my view wholly mistaken' (p. 334).

- 3 A good example of this impression is Heath's (rather tautological) argument against Habermas's claim that practical discourse involves a search for generalizable interests: 'According to this suggestion, participants in practical discourse might come to reconceptualize their conflict when they see that underlying their differences is some kind of common interest stemming from their shared humanity. But none of this is helpful. When people are in conflict, they have a general interest in cooperating, and cooperation is almost always achieved through compromise. After all, if they didn't need to compromise, then they wouldn't need to cooperate. All that agents are going to discover at a higher level of abstraction or generalization is that they have an abstract or general interest in compromising' (p. 233).
- 4 One difficulty internal to his strategy seems to remain, though. Heath has already argued in chapter 3 that the only difference between theoretical and practical discourses is their respective topics. In chapter 7, he rejects Habermas's own answer to the question of what the specific topic of moral discourses is (namely, the question of what is equally in everyone's interest). But he does not offer an alternative account of what the topic of moral discourses is, according to his own interpretation. This in turn makes it impossible to evaluate the plausibility of his specific proposals, such as the rejection of the principle of universalization or of the moral/ethical distinction.