Joseph Heath

The problem of foundationalism in Habermas's discourse ethics

Although Jürgen Habermas claims to share with Karl-Otto Apel 'the essentials' of the discourse ethics program, he has recently tried to distance himself from what he takes to be a 'lingering (yet unacknowledged) foundationalism' pervading the 'architectonic disposition' of Apel's thought. Specifically, he objects to Apel's use of the presuppositions of argumentation as a set of supernorms from which concrete moral obligations can be derived. Against this, he argues that norms can be justified only through real practical discourses, and that the philosopher can contribute to the justification of a norm only as a virtual participant in such a discourse. This is because there can be plenty of good reasons given for or against a particular norm, but never an 'ultimate justification', or all-purpose knock-down argument.

The implication here is that Habermas is a contextualist when it comes to questions of justification. Unfortunately, he does not carry this through consistently. In what follows, I will argue that Habermas's own theory of moral argumentation exhibits a form of residual foundationalism. But unlike Apel's critics, I am not interested in claiming that Habermas's philosophical arguments are foundationalist. Rather, I want to suggest that his conception of practical discourse assumes that participants in such a discourse can employ only foundationalist modes of justification. In other words, Habermas's rules of discourse illegitimately impose a foundationalist schema onto the practice of moral argumentation.
My strategy for advancing this claim is somewhat roundabout. I would like to begin by arguing that Habermas's key principle of argumentation, 'U', enshrines a version of the naturalistic fallacy. Specifically, this bridge principle attempts to license an illegitimate intermodal transfer of validity from sincerity to rightness. Next, I would like to suggest that the inclusion of such a problematic inference rule is motivated by Habermas's desire to separate out a class of privileged judgments to serve as the 'grounds' for problematic rightness claims. This is not in itself foundationalist, since Habermas maintains that these privileged judgments can in turn be revised in specialized sub-discourses. Instead, the foundationalism shows up at the level of 'architectonic disposition' (once again). Consistent anti-foundationalism undermines the impulse to create a structured hierarchy of argumentative claims, and thus eliminates the need to worry about bridge principles.

I Habermas on the naturalistic fallacy

Unlike many critics of positivist social science, Habermas does not reject the basic intuition behind the ban on 'is-ought' inferences. In this respect, his position is somewhat unusual. The more standard line (e.g. that of Peter Winch, Charles Taylor, or Alasdair MacIntyre) maintains that the descriptive 'component' of terms used to characterize social phenomena cannot be separated from the evaluative. In Taylor's terminology, we understand the point of a social practice only insofar as we see why it is important, why its performance matters to those involved. The fusion of interpretive horizons requires that we bring our own 'strong evaluations' into dialogue with those of the observed practice. The 'is-ought' distinction is therefore quite misleading, because any insightful description of an action will have to situate it within a shared moral space.

In contrast, Habermas argues that if the descriptive and evaluative seem inseparable, it is only because insufficient attention has been paid to the performative contexts in which speakers use meaningful expressions. In every speech act, he argues, speakers raise three distinct claims to validity: truth for a proposition or an existential presupposition, rightness for an action referred to or the norm governing it, and sincerity for the inner state expressed. In so doing, speakers take up an attitude toward something in the objective, social and inner worlds. While not semantically separable, each of these claims can be distinguished pragmatically. For instance, challenging the speech act in different ways can lead to an argument stressing a single dimension of validity. And different speech-act 'modes' stress one or another claim, e.g. constatives stress the truth of an assertion, regulatives the rightness of an action, expresses the sincerity of an utterance.

The reason positivism is excluded under this framework is that all communicative action raises implicit or explicit claims to normative validity. In order to grasp the meaning of a speech-act, the hearer must have some sense of how such a claim could be redeemed by the speaker; in order to construct a rational interpretation, the hearer must take a position on the various claims, deciding whether they in fact could be redeemed. In the case of the natural sciences, evaluation under the aspect of rightness is manifest only at the level of the community of inquiry (affecting choice of research program, etc.). But the social sciences have as their object of study a domain that is linguistically constituted. The social-scientific observer must assume a participant perspective in order to access the 'meanings' expressed, but in so doing must take a position on the normative aspect of the speech-acts in the object domain itself. This 'double hermeneutic' of the social sciences means that social action cannot be understood without being normatively evaluated.

The tripartite validity-claim scheme, far from undercutting the logical gap between 'is' and 'ought' statements, actually presupposes it. In fact, under Habermas's treatment Hume's insight becomes one of the defining moments of modernity. Following Weber, Habermas argues that the process of cultural rationalization involves the growing differentiation of the object domains referred to under these three aspects of validity. The key epistemic gain in this transformation – disenchantment – involves the 'desocialization of nature and the denaturalization of society,' i.e. the recognition that truth-claims referring to existing states of affairs do not imply rightness-claims about the way things ought to be, and vice versa.

This formulation allows us to articulate far more precisely the intuition that informs Hume's point (which is misleadingly stated in terms of the grammatical form of sentences). Given certain propositional contents, there are three prohibited inferences: there can be no transfer of validity from truth to rightness, sincerity to rightness, or rightness to truth. This perspective provides us with a more comprehensive typology of fallacious inference. For instance, MacIntyre rightly points out Hume's inconsistency in criticizing what later came to be called a 'naturalistic' fallacy, while at the same time elaborating an explicitly naturalistic ethical theory. But if we distinguish between truth and sincerity, it is clear that Hume was critical only of attempts to derive normative claims from the order of nature, i.e. derive rightness from truth. His own moral sentiment theory relied upon a fallacious
inference from sincerity to rightness (e.g. 'I feel that action A is wrong, therefore norm N that legitimates it is wrong').

Here are a few comparable fallacious arguments:

1. A causes me pain, therefore N is wrong.
2. A frustrates my needs and desires, therefore N is wrong.
3. A interferes with my interests, therefore N is wrong.

Now in each of these cases the sincerity claim may be refutable: e.g. the speaker may be experiencing phantom pain, may have inauthentic needs and desires, or may have miscalculated her interests. But even if we grant the sincerity claim, we still cannot get to the desired conclusion. This is obviously so where argument 1 is made by a child being punished, or 2 by a murderer challenging the right to self-defense of his intended victim, or 3 by an employer criticizing collective bargaining procedures. The premise taken all by itself simply does not imply the stated conclusion. In cases where these arguments do appear plausible, it is because of a suppressed premise that supplies the rightness-claim: e.g. a norm that legitimates certain needs and interests, or that prohibits certain outcomes. Argument 1, for instance, works if we supplement it with a norm such as 'causing suffering is wrong'. Because we assume that this norm holds in all but exceptional cases, it is easy to forget about it. It is this habit of mind that makes the following statement sound counter-intuitive: causing pain is not wrong; rather, violating the norm that prohibits causing pain is wrong.

It might seem somewhat superfluous to distinguish this sincerity-rightness inference (which I will call the 'psychological fallacy') from the truth–rightness inference (for which I have been appropriating Moore's term 'naturalistic fallacy'). It could be taken as a mere subdivision of items within a class whose overall membership is itself not in doubt. Yet I think the distinction remains significant, because avoiding these two fallacies involves resisting two very different philosophical temptations. Avoiding the naturalistic fallacy involves rejecting cosmology. Avoiding the psychological fallacy, on the other hand, involves accepting the social world, in Durkheim's phrase, as a reality sui generis. Obviously, under this definition very few modern philosophers are inclined to commit the naturalistic fallacy, while the psychological fallacy remains an ongoing concern.

II Contractualism and the psychological fallacy

The concept of a 'psychological fallacy' is important because it allows us to locate very precisely one of the central problems with contractualist theories of morality. The basic claim of contractualism is that the rightness of a norm or action depends upon the possibility of its being the object of a rational, voluntary agreement among all those affected. But there is a key ambiguity in the use of the term 'rational', since there are two significantly different accounts of how actors can be rationally motivated to accept an agreement. 'Private reason' versions of contractualism use the model of instrumental rationality to maintain that actors can be motivated to agree only by prudential considerations. 'Public reason' versions use the communicative rationality paradigm to maintain that actors can be motivated to agree by arguments, and thus by moral considerations.

These two characterizations of rationality lead to very different conceptions of agreement and its justificatory role. In the instrumental rationality paradigm, the moral agreement is analogous to the formation of a cartel by economic actors. Here the norms of the cartel are justified, or right, because of the agreement itself. The only way to criticize the norm would be to criticize the agreement, and to do this one would have to find an individual who was affected by it, yet whose interests were inadequately represented (through exclusion, manipulation, self-delusion, etc.). The communicative rationality paradigm, on the other hand, considers the moral agreement analogous to the consensus of a community of scientific investigators. But among scientists it is not the agreement that justifies the truth-claim, but the argumentative justification of the claim that brings about the agreement. In the moral context, this means an agreement does not justify a norm, the norm itself must first be justified argumentatively. Criticizing the norm will therefore involve criticizing the reasons given for it. The function of the agreement is to ensure that the reasons are 'objective' in the Kantian sense, i.e. good for all.

These considerations make it clear that the role of the 'contract' is quite different in the private and public versions of contractualism. This is primarily because the private version attempts to establish moral cognitivism, while the public version presupposes it. Scientific investigation begins with a set of statements accepted as true, then attempts to revise these or generate more; in the same way, public contractualism begins with accepted norms and either revises them or develops more. Private contractualism, on the other hand, focuses on motives for agreement that are, in the specifically moral sense, non-cognitive. It does not begin with norms, but with non-moral preferences. This gives it a structure which I call the 'interests-in-norms-out' schema, shown in Figure 1.

If we take this schema, along with the instrumental rationality paradigm, as the two essential components of a private contractualist
theory, then it becomes apparent that very few theorists have been willing to maintain it in its pure form. Even Richard Posner, who quite single-mindedly defended a pure version for some time, appears to have recanted. But similarly, the only theorists I know of who argue the public version consistently are T. M. Scanlon and Samuel Freeman. Most others – I am thinking specifically of John Rawls, David Gauthier and Habermas – maintain something halfway between the two. Given that the private version has some very serious difficulties, it is useful to examine why these thinkers work so hard to retain elements of it.

It seems to me that pure private contractualism offers two principal advantages.

1 **Predictability.** Because the private theory relies upon the instrumental rationality paradigm, it is possible to model the procedure that leads to agreement, i.e. determine hypothetically what each actor would rationally choose. This means that the procedure need never actually be carried out, because (given knowledge of the payoffs) the theorist can determine monologically what the outcome will be. Morality becomes a subdomain of game theory, and any philosopher with the right tools can definitively resolve any moral dilemma. In contrast, the communicative rationality paradigm resists modeling, so that the practical discourse must actually be carried out. This means the philosopher cannot determine in advance how ‘our’ norms will do when confronted with those of another individual or culture. We may think we have good reasons for many of our important moral obligations, but we must wait until the end of the day to see how good they really are. According to Taylor, this forecloses on our desire to create a philosophically ‘shrunken moral universe’, where we begin every argument having already assured ourselves of the outcome by setting up some of our goods as trumps.

2 **Getting something out of nothing.** Because of the interests-in-norms-out schema, the valid norm produced (i.e. the rightness claim) is not derived from another rightness claim, and so does not presuppose an antecedent morality. Thus *agreement itself* is the sole right-making characteristic. This permits the spontaneous generation of valid norms, allowing us to avoid all the difficult questions of moral ontology. On the other hand, if we shift to the communicative paradigm and say that, just as all true statements are backed by other true statements, all rightness claims are backed by other rightness claims, then we inherit all the problems alleged to accompany non-correspondence theories of truth.

Despite these advantages, very few theorists since Hobbes have accepted private contractualism at face value. This is because many of the ‘solutions’ it generates are morally outrageous. For instance, take a private contractualist scheme with a simple auctioning procedure ensuring efficient solutions to two-party disputes. Participants represent their interests by bidding against one another for the ‘right’ embodied in the norm. The party that bids highest (i.e. anticipates most aggregate satisfaction) gets the right, then (optionally) compensates the other party for the amount which it had bid. This scheme (which is more or less Posner’s) produces a number of unusual arguments, e.g. laws against rape are just, because if rapists and women were to negotiate over the right, women would be willing to pay more for the right not to be raped than rapists would be willing to pay for the right to rape them.

Ignoring for the moment the suggestion that frustrated rapists should be compensated, the argument collides with a deeply held moral intuition. The very idea that women should have to bargain and compromise with rapists, and that both groups’ interests should be considered ‘on par’, seems outrageous. The strength of the rapists’ desires or the ratio of rapists to women cannot possibly be considered a morally relevant factor. We are faced with the inescapable intuition that some types of interests (no matter how sincerely expressed) should be categorically excluded from any moral decision-making procedure. Raw interest contractualism therefore seems to rest upon a version of the psychologistic fallacy. The mere articulation of an interest simply does not speak *for or against* the rightness of a norm, and thus it is not necessary that a valid norm reconcile the interests of all parties.

Rawls takes this to be a clear limitation on contractualism: ‘We cannot take various contingencies as known and individual preferences as given and expect to elucidate the concept of justice (or fairness) by theories of bargaining.' Both he and Gauthier respond by setting up their procedures in such a way that they can screen the interests that are to be plugged into them. They attempt to distinguish, in effect, between morally legitimate and illegitimate interests. This
allows them to avoid the psychologistic fallacy, but at the cost of giving up advantage 2.

Both thinkers suggest that the distinctive contribution of moral theory over and above game theory is that it attempts to characterize a pre-contract state of affairs that is acceptable from the moral point of view. Simply producing an equilibrium outcome does not entail that it is right or just. According to Rawls, ‘The moral assessment of equilibrium situations depends upon the background circumstances that determine them’.17 If the initial bargaining position is determined by a situation in which the interests of one party are shaped by coercion or threats, then a contractualist procedure would simply grant this state of affairs moral legitimacy. According to Gauthier, ‘to accept the natural distribution as the initial bargaining position would be to accept might as making right’.18

Gauthier’s response to this difficulty self-consciously parallels Locke’s response to Hobbes. He moralizes the state of nature by correlating the ‘natural rights’ of participants with duties toward others. He formulates this in what he calls the Lockean proviso: ‘Each person’s endowment includes whatever he acquires without worsening the situation of his fellows.’19 He thus retains the interests-in-norms-out schema, but has participants enter the procedure with only morally legitimate interests, i.e. rights: ‘Rights provide the starting point for, and not the outcome of, agreement. They are what each person brings to the bargaining table, not what she takes from it.’20 No rational agent would engage in cooperative action in instances where this proviso did not hold.

Rawls’s response is essentially the same, but instead of proscribing illegitimate interests, he prescribes a set of legitimate ones. His first move in establishing the characteristics of the ‘original position’ is to wipe the slate clean, denying participants knowledge of their interests. He then introduces interests in the form of primary goods, which he defines as those goods ‘which it is supposed a rational man wants no matter what else he wants’.21 Participants do not make a rational choice on the basis of their ‘real’ preferences, but simply try to maximize the number of units they will eventually receive from this sanitized list of goods. Despite the fact that everyone’s interests conflict with everybody else’s, because they all compete for the same goods they all share the same interests.

But both of these responses avoid the real issue. The notion of morally legitimate interests cannot be represented as merely a question of rational, shared, or universally accepted interests. Just because everyone can agree that it is rational for an individual to pursue a particular interest, it still does not follow that it is right. Scanlon makes this point quite forcefully:

The right-making force of a person’s desires is specified by what might be called a conception of morally legitimate interests. Such a conception is a product of moral argumentation; it is not given, as the notion of individual well-being may be, simply by the idea of what it is rational for an individual to desire. Not everything for which I have a rational desire will be something in which others need concede me to have a legitimate interest which they undertake to weigh in deciding what to do. . . . A framework of moral argument is required to define our legitimate interests and to account for their moral force. This same contractualist framework can also account for the force of other moral notions such as rights, individual responsibility and procedural fairness.22

In order to make good on the claim that the interests allowed into the contract procedure are acceptable from a moral point of view, both Rawls and Gauthier must provide some form of moral argument. Gauthier’s argument is weakest, since it relies crucially upon our sharing the moral intuitions underlying Robert Nozick’s multiple-Robinson-Crusoes-on-desert-islands thought-experiment. Rawls’s position is more robust, because he makes the structure of the original position the product of ‘reflective equilibrium’ with our traditional conception of justice.) But if some form of moral argument is required to establish the legitimacy of the interests accepted in the contract procedure, then logically the contract itself can no longer claim to provide a comprehensive account of moral reasoning. Accepting that there is a form of moral argumentation that validates the contract thought-experiment amounts to conceding that communicative rationality supervenes over the instrumental. And if we must presuppose the validity of some form of moral argumentation to establish the inputs of the contractualist schema, then why should this same form not be available to participants in the procedure?

What Scanlon suggests is that interests be introduced into argumentation as entitlements, licensed through appeal to moral principles that are themselves open to discursive testing and revision. In this way, the individuals involved in the agreement have available to them both the forms of argument and the substantive considerations that Rawls and Gauthier reserve for their philosophical metadiscourse. This expanded contract procedure, which might be referred to as a norms-in-norms-out model, suggests itself as the most consistent way of working out the intuitions underlying the idea of public contractualism.
III Problems in the formulation of discourse ethics

These reflections suggest that private contractualism can only avoid its counter-intuitive results by falling back upon a public form of moral argumentation. This does not automatically lead to public contractualism. To go the full distance, one would have to show that argumentation is not simply one social practice among many, but a forum where the constitutive presuppositions of any particular practice are thematized and discursively tested. If every norm makes an implicit claim to discursive redeemability, then the 'transition to argument' move is always available to social actors, because it is embedded in every practice.

Habermas tries to establish this superordinate role for argumentation through a theory of meaning. The move is highly contested, and whether his particular approach will work in the long run is, I think, still an open question. But given that the basic idea—a valid norm must be discursively redeemable—is fairly plausible, there is likely to be another reason for all the resistance that his project has encountered. I believe many are motivated to reject the argument because they find its consequences unacceptable, given the way Habermas characterizes the practice of moral argumentation.

The problem stems from the fact that Habermas views the counter-intuitive consequences of pure private contractualism not as a basic problem with the interests-in-norms-out schema, but as the result of a procedure that is too weak to screen outcomes adequately. He therefore tries to retain the interests-in-component of private contractualism, while at the same time characterizing the contract procedure as a form of moral argument. This leads to problems on two fronts. First, instead of trying to screen interests in the way that Gauthier and Rawls do, he tries to eliminate the counter-intuitive outcomes exclusively through formal-procedural constraints. He claims that practical discourse has built into it a universalizability requirement, which places a much stronger constraint on possible outcomes than strategic compromise. The question is then whether universalizability alone is adequate to this task. The second problem concerns the psychologistic fallacy. Because he characterizes the contract procedure as a type of argumentation, but restricts the input to the interests of participants, he must posit a form of argument that contains only interests in its premises, yet norms in the conclusion.

Whether Habermas is able to resolve these problems depends upon how he formulates 'U', his universalization principle. Unfortunately, his position is very difficult to evaluate, partly because he does not give any examples of how U would be applied, but mainly because he regularly glosses it in one of at least three non-equivalent ways. In order to keep the discussion focused, I will begin by considering a literal interpretation of U, then deal with the variations.

Habermas attempts to introduce a universalization principle into his philosophical ethics in the form of a rule of argumentation. A contested norm, he argues, cannot meet with the consent of participants unless all affected can freely accept the consequences and the side effects that the general observance of a controversial norm can be expected to have for the satisfaction of the interests of every individual.

Unlike Kant's categorical imperative, which could be used directly to justify norms, Habermas intends U to be used as a 'bridge principle' for practical discourse, which would serve much the same function that the principle of induction does in arguments over truth-claims. In 'Wahrheitstheorien', adopting Stephen Toulmin's argument schema, he suggests an analogy which may be rendered as in Figure 2.

![Figure 2](image)

This formulation suggests that individuals come to the discourse with different, potentially conflicting interests, and work out a rule of action that regulates these interests in a way that all can accept. This interpretation is suggested when Habermas glosses U as providing for a 'balancing of interests'. With this schema, an inductive argument would run something like this: 'Given the relevant facts: \( F_1, F_2, F_3 \), and given that theory X explains all these facts better than any other, I conclude that theory X is true.' I assume that a moral argument would be similar: 'Given the interests of all affected: \( I_1, I_2, I_3 \), and given that norm N satisfies all these interests better than any other, I conclude that norm N is right.' In the same way that scientists accept a theory when they see that it accounts for the facts, moral agents accept a norm when they see that it adequately balances the interests of all. Because all
agents accept the norm for the same reason, considerations of this type qualify as public reasons.

But there is an obvious disanology between the two cases. With induction the backing statements for the theory, i.e. the facts, are also truth claims, while for universalization the backing statements for the norm are not truth-claims. Needs and interests can be expressed sincerely, but they have no intrinsic right-making force. Thus to infer the rightness of a norm from a series of sincerity-claims would be a clear instance of the psychologistic fallacy, even if U ensures that they are adequately balanced.  

The same problem shows up in the form of another disanology. Habermas accepts the view that facts are never theory-independent. Thus the facts used to back a theory could be revised, i.e. made false, by theoretical considerations. But the needs and interests that back a norm could not be revised by normative considerations, since revision would require that they be shown insincere. Habermas is careful to point out that his conception of interests differs from the rational-choice theorist’s preferences, on the grounds that ‘the descriptive terms in which each individual perceives his interests must be open to criticism by others’. But criticizing the adequacy of an individual’s need-interpretations is not the same thing as criticizing the rightness of the need-claim. By not restricting his procedure to morally legitimate needs, Habermas makes the backing statements for a norm unrevivable within the framework of a strictly practical discourse, i.e. one dealing exclusively with rightness-claims.

How well an individual perceives his interests, or how authentically she interprets her needs, are topics for therapeutic and ethical discourses, not moral-practical. The outcome of such discourses will certainly be relevant to practical discourse, particularly with regard to norms that grant individuals broad license for self-fulfillment. But this is true in the same way that decisions about truth claims settled in technical discourses will be relevant for determining, among other things, the likely consequences of certain courses of action. None of this changes the fact that the statements used to back a rightness-claim must themselves be rightness-claims, and they must be falsifiable on the basis of other rightness-claims. Inferences licensed by U would therefore qualify prima facie as instances of the psychologistic fallacy.

The second major problem concerns the ability of U to adequately constrain the set of outcomes. Habermas suggests that U is more restrictive than strategic compromise, but unfortunately bases his argument on the rejection of a relatively unsophisticated bargaining model. When compared with Gauthier’s version, U actually turns out to impose weaker constraints than bargaining. To see this, it is enough to note that U can be derived literally from the four ‘conditions on rational bargaining’ that Gauthier uses to establish his principle of minimax relative concession. Thus Gauthier’s bargaining outcomes will always satisfy U. But unlike the minimax relative concession principle, U does not provide any criteria for selecting among Pareto non-comparable welfare outcomes, and so must be supplemented by substantive norms in order to solve any questions of distributive justice. This suggests that without some constraint on the interests admitted into the argument, U would not help eliminate any of the counter-intuitive outcomes that trouble private contractualist models.

Because Gauthier requires the equal rationality of bargainers, ‘no one can expect any other rational person to be willing to make a concession if he would not be willing to make a similar concession’. He further stipulates that no actor be willing to entertain a concession unless required to do so. A rational joint strategy thus requires that each participant not accept the concession of another unless she would also be willing to accept it were she in that person’s position. Thus ‘rational persons will voluntarily accept an agreement only insofar as they perceive it to be equally advantageous to each’. And the bargaining outcome is acceptable to agents not because it is in their private interest, but rather because each can see that it is in the interests of all, i.e. it specifies what ‘all can will in common’. Given this characterization of bargaining, it is difficult to see what the specific contribution of argumentation is supposed to be in Habermas’s contractualism, if all that is required is an impartial balancing of interests.

This is where Habermas’s non-equivalent glosses on U become significant. Perhaps due to the problems associated with accepting interests as procedural input, Habermas often presents U in a way that downplays the significance of individual interests. Rather than seeking to balance a variety of interests, participants in practical discourse try to discover a common, or universal, interest. Under this interpretation, rather than reconciling a variety of interests, practical discourse would have the task of identifying a single interest that everyone shared. Since the need for normative regulation arises out of conflicts of interest, this universal interest would presumably not be immediately available. It could only be discovered by having agents move up from their particular interests to higher and higher levels of abstraction, until they found an interest that was shared. Moral argumentation could therefore only secure agreement on the most absolutely general of norms.

This formulation has many disadvantages: it is hard to see how one
Foundationism in Habermas’s discourse ethics

It is very important to note the conspicuous lack of such an argument in Habermas’s work. He usually begins by claiming that norms embody a general will, then shifts suddenly to the claim that they represent a general interest. When he does attempt to justify this transition, he argues that it is simply analytic. In Moral Consciousness and Communicative Action he says that U can be derived from the rules of argumentation along with knowledge of ‘what it means to justify a norm’.36 And in The Theory of Communicative Action he claims that it ‘follows with conceptual necessity from the meaning of normative validity claims’ that normative claims are decidable on the grounds that they express ‘an interest common to all those affected’.37 This appeal to analyticity is clearly not persuasive, and it seems quite at odds with the Durkheimian sense in which he usually uses the term ‘norm’.

This brings me to my central thesis, which is that Habermas tries to retain the interests-in norms-out schema because of a residual foundationalism in his argumentation schema. Here I am considering foundationalism as a particular strategy for explaining how it is possible to justify a truth- or rightness-claim.38 This can be described best by considering how the foundationalist responds to skeptical doubts about the possibility of justification. The well-known skeptical argument, Agrippa’s trilemma, states that when any proposition advanced as knowledge is challenged, the proponent is faced with only three options:

1 to refuse to respond, i.e. to make an undefended assumption
2 to repeat a claim made earlier in the argument, i.e. to reason in a circle
3 to keep trying to think of something new to say, i.e. to embark on an infinite regress.39

The central claim is that we face an unavoidable regress of justification, since any attempt to justify a proposition by appealing to another proposition leads nowhere, because the proposition appealed to as grounds will in turn have to be justified. Thus the chain of justification either goes on for ever, or must be cut short in one of two unsatisfactory ways.

The foundationalist responds to this by dividing all of our knowledge into two classes: a privileged and a problematic. Statements of the privileged class are indubitable, or highly certain, but in either case not in need of justification. The task of epistemology is then to determine the inferential relations through which the problematic class can be safely derived from the privileged. The most popular form of foundationalism is empiricism, which privileges the class of statements

can get this interpretation from the actual text of U, so there is a substantial addition to the burden of proof; the mere sharing of an interest does not make it morally legitimate, and so the psychologistic fallacy persists; the term ‘universal’ interest is incorrect, since the scope of U is restricted to ‘all those affected’ by the norm; the connection between practical discourse and communicative action is severed, since the institutions and social norms of the society could not be justified as such in this type of discourse; abstracting conflicts of interest does not do anything to resolve them, so the problem of mediating disputes persists; and in the context of cultural pluralism, there is no reason to expect that a common interest can be attained through abstraction.

In other places, Habermas tries to cash out the idea of a common interest as that which is ‘equally good for all’, or in the ‘equal interest of all’. The idea is presumably that a common interest is one that satisfies everyone’s particular interests to an equal extent. Here the difference between the actual text of U and the gloss is more significant, because there is no mention of ‘equality’ in the U, and there is no obvious way that ‘equality’ could be derived from it. And since the concept of ‘equality’ is notoriously difficult to specify, it seems unlikely that any particular conception could be shown to be latent in the pragmatic presuppositions of argumentation. Furthermore, if one grants that there can be morally illegitimate interests, then treating all interests equally creates more problems than it solves.

IV Foundationism and the psychologistic fallacy

The arguments presented above cannot be conclusive, mainly because Habermas has not clearly specified how U is to be interpreted and applied. Nevertheless, they reinforce the impression that there is no good way to establish the validity of a norm through a procedure that operates only on the interests of agents. But these difficulties invite us to step back for a moment and ask the more general question: why is Habermas discussing interests in the first place? Looking over the history of contractualist thought, it is apparent that theorists seldom start from the claim that morality involves the reconciliation of interests; instead, they begin by equating all rational action with instrumentally rational action, from which it then follows that interests are the sole motivating grounds for agreement. But if one rejects the instrumental rationality paradigm, there is no longer any obvious reason why the interests of participants should be considered morally relevant. At very least an argument is needed to establish the connection.
that report direct sensory experience. All other true statements are justified by adding a chain of reasons that eventually terminates with one or more of these experiential statements. These are not mere assumptions, they are statements made true by some non-epistemic relation, such as a causal relation to the actual object of stimulus. They are the 'brute facts'.

Now I think Habermas's conception of how norms are justified adopts a very similar form of foundationalism, using interests as a privileged class of statements in moral argumentation. The task of moral theory is then to determine how the problematic class of norms can be derived from this privileged class. The fact that interests are criticizable is not at issue here. The empiricist is attracted to the 'brute-facts-in-truth-claims-out' schema because she is nervous about the consequences of 'truth-claims-in-truth-claims-out'. I believe Habermas's use of the interests-in-norms-out schema is similarly motivated. Unlike Scanlon, he is nervous about the implications of a norms-in-norms-out model.

However, the waters are muddied by the fact that Habermas responds to Agrippa's trilemma at the wrong level.40 The skeptic claims that morality is fundamentally decisionistic because norms cannot be justified without an infinite regress. A participant in practical discourse who is pressed to justify a rightness-claim will eventually run out of reasons. Every moral argument will then wind up with an appeal to accepted practice – Richard Rorty's 'WE do not do that'. But as far as public contractualism is concerned, these doubts are not, strictly speaking, a problem for the philosopher. Agrippa's trilemma is a challenge to the participants in practical discourse because they are the ones trying to justify norms. Because Apel tries to use discourse principles to justify norms directly, it makes sense for him to mount a transcendental-pragmatic defense of his rules of discourse in response to the Agrippian trilemma.41 But Habermas tries to justify U in a similar way, thereby missing the force of the skeptic's doubts. Agrippa's trilemma does not challenge the way in which chains of justification are constructed, i.e. it does not challenge the rules of inference involved. It simply points out that the chain, once constructed, either goes on for ever or ends in an unsatisfactory manner. U's status as a pragmatic presupposition would only have to be defended against the trilemma challenge if it were a proposition that could be appealed to in order to justify a norm. But U cannot be called upon by participants as a reason for accepting a norm, it is merely a rule that licenses the use of certain propositions to justify others.42 As the 'Wahrheitstheorien' argumentation schema makes clear, we do not justify scientific theories by appealing to the principle of induction, we appeal to evidence. Similarly, we justify norms by appealing to other rightness-claims – and this is where the skeptic's doubts acquire purchase.

Thus Habermas's most important response to the moral skeptic is found not in his pragmatic presupposition argument, but in his argumentation schema. He cuts short the regress by separating out a class of statements that are not backed by other rightness-claims, and thus can serve as the ground for the problematic class of moral judgments. Interests, within practical discourse, come to play the same role that the empiricist's 'brute facts' were supposed to in scientific discourse. Any participant, pressed to justify a rightness-claim, need only point to the interests at stake. And these interests clearly cannot be doubted in any morally relevant sense. Through the illegitimate intermodal transfer of validity enshrined in U, the regress of justification is cut short, the foundations are established.

This is unnecessary, since there is another way of dealing with the regress problem. The claim that an argument is justifiable on the basis of good reasons does not suggest that these reasons can all be given simultaneously. Within any process of justification, there will be certain framework propositions, or propositions that 'stand fast'. All arguments exchanged within that common horizon will presuppose the truth of these propositions, and so any particular chain of justification will eventually terminate with one of them. The foundationalist's view is characterized by the belief that the same class of propositions stands fast in every context, and that membership of this class can be determined in advance on the basis of some particular epistemic relation between it and other propositions. One version of the contextualist view, on the other hand, maintains that while certain propositions are always taken for granted, which ones are presupposed depends entirely on the context. This draws attention to the important fact that just because justification cannot proceed without assumptions, it does not follow that what is assumed cannot in turn be justified. What we take to be assumptions in one context can be questioned and revised or justified in another.

The skeptical challenge simply points out that there is no statement that cannot, in principle, be challenged and discussed. The regress of justification amounts to the claim that 'in discussing one lot of assumptions further assumptions have to be made, so that it is impossible to conceive of a situation which is well-founded in all its aspects'.43 With the robust conception of the background supplied by Habermas's understanding of the lifeworld, this just means that all of our moral and scientific knowledge is knowledge-for-all-practical-purposes. And the fact that justification always occurs in a context in
which some statements are taken for granted does not mean that claims to truth and rightness cannot have a context-transcendent import.

If we take this view, then it does not make much sense to place too much emphasis on bridge principles like induction. As Wittgenstein pointed out, there is no hard and fast distinction between facts and theories; much of it depends on context. 'The earth rotates around the sun' can be either a fact or a theory, depending on what kind of conversation we are having. U could reasonably be reconstructed by referring to participants' entitlements instead of their interests, but I am not sure what the point would be. In moral argument we move easily between references to morally legitimate values, imperatives, guidelines, roles, expectations, entitlements, demands, etc. Nothing important seems to be gained by picking out a particular class and assigning it a privileged role.

V Getting by without U

Habermas resists any suggestion that the justification of moral rules could be essentially context-dependent,⁴⁴ even though his characterization of practical discourse points in that direction. The tension between the two impulses shows up in the problems associated with his formulation of U. I have tried to make a persuasive case for the claim that Habermas's reasons for introducing U are not defensible. In conclusion, I would like to argue that U can be discarded and a fully constructivist conception of practical discourse adopted, without compromising the basic ambitions of his project.

It is probably best to understand the formulation of U as a surreptitious attempt to retain advantages 1 and 2 of private contractualism without buying into the instrumental rationality paradigm. For obvious reasons, Habermas does not want to abandon advantage 1 (predictability) entirely. It is difficult, pace Taylor, to give up on the idea that the moral universe is not at least slightly 'preshrunken'. Surely we do not have to go through a practical discourse to see that some norms are unjustifiable. And it seems wrong to suggest that in principle any proposed norm could turn out to be justified, given the right context and a clever argument. But public contractualism, despite abandoning the claim to offer a definitive resolution to every moral problem, can still exclude certain outcomes on the basis of purely philosophical considerations.

Consider the following as a basic principle of discourse ethics. A valid norm must fulfill the condition: 'No one could reasonably reject it as a basis for informed, unforced general agreement.'⁴⁵ The very idea that a norm must be justified by good reasons precludes a morality based on what Habermas calls 'power-claims', i.e. threats, coercion, violence. It also rules out deception, and any straightforward appeal to tradition, revelation, or other form of authority (although not in cases where there is a good reason why a particular authority should be trusted). Any norm that depended crucially upon these types of conditions could reasonably be rejected.

The anticipation of an agreement also imposes constraints on possible outcomes. If, as Habermas suggests, rightness involves an idealization of rational acceptability, then the association of a validity claim with a norm implies that the norm would be acceptable under ideal conditions. Although inferring the rightness of a norm from a projected ideal consensus would be committing the fallacy of affirming the consequent, the agreement can still serve as what Robert Alexy calls a 'negative hypothetical criterion'.⁴⁶ Through the modus tollens, predicting a dissensus allows one to infer that the norm at issue is unjust. Used regulatively in this sense, the argumentation procedure incorporates the second formulation of the categorical imperative rather than the first. Treating an individual as a means only implies disregarding that person's potential acceptance or rejection of the norm licensing the action. Any norm that exhibited significant disregard for the morally legitimate needs of some individuals could safely be predicted to fail, and any that denied the right of those individuals to veto the arrangements could be excluded in principle.

If the discourse-theoretic analysis of moral argumentation were reformulated in such a way as to incorporate only these weaker constraints, the result would be a more consistently procedural conception of morality. As it stands, U goes too far in the direction of specifying the content of the arguments that can be employed in defense of a norm. At the danger of closing on a cryptic note, I would like to suggest that purely formal constraints could be used to restructure the structure of moral argumentation in a way that more closely resembled Habermas's analysis of legislative procedure in Faktizität und Geltung.⁴⁷ Rather than providing a transcendental justification for a principle, the analysis would establish a system of participant rights. Norms whose cognitive content or practical consequences contradicted these rights could be excluded in principle. The scope of properly philosophical claims about morality would then parallel that of the supreme court in his model of constitutional adjudication.⁴⁸ While I cannot go into the details of this alternative here, it may suffice to note that Habermas's legal-theoretic instantiation of practical discourse results in a very strong conception of democratic freedom. This suggests that the same formal structure applied to moral
arguementation could eliminate the problems associated with U., without undercutting his basic theoretical ambitions.

Northwestern University, Evanston, IL, USA

Notes

I would like to thank Thomas McCarthy and William Rehg for a number of very helpful ideas and criticisms.

1 Jürgen Habermas, Justification and Application (Cambridge, MA: MIT Press, 1993), p. 79; hereafter cited as JA.

2 This is not to be confused with the form of contextualism Habermas opposes, which is contextualism about truth- and rightness-claims. One can hold that truth and rightness make claims to context-transcendent validity, while still acknowledging that any particular argument given as a justification will rely upon a shared context of taken-for-granted beliefs, norms, etc.


4 ibid., Vol. 1, p. 48.

5 Habermas sets out the valid and invalid inferences in a helpful chart; ibid., Vol. 1, p. 445.


9 Thomas McCarthy makes this point well in 'Philosophy and Social Practice: Avoiding the Ethnocentric Predicament', in Axel Honneth et al. (eds) Philosophical Interventions in the Unfinished Project of Enlightenment (Cambridge, MA: MIT Press, 1992). Habermas has not always taken this position. In 'A Reply' in Axel Honneth and Hans Joas (eds) Communicative Action (Cambridge, MA: MIT Press, 1991), he writes: 'I do not understand the discourse theory of truth to mean that the consensus achieved discursively is a criterion of truth (as was the case in some of my earlier statements); rather, it should explain via the discursive redemption of validity claims the meaning of each element of unconditionality which we intuitively link with the concept of truth' (p. 233). Thus his talk about the consensus theory of truth is quite misleading. Rather than consensus-under-ideal-conditions implying truth, in this new version truth implies consensus-under-ideal-conditions. This means that the consensus functions regulatively, i.e. lack of consensus presents a prima facie challenge to the truth-claim. Assuming that what holds for truth holds for rightness, statements like the following must be considered outdated: 'Practical discourse is an exacting form of argumentative decision-making. Like Rawls' original position, it is a warrant of the rightness (or fairness) of any conceivable normative agreement that it is reached under these conditions'; see 'Morality and Ethical Life' in Moral Consciousness and Communicative Action (Cambridge, MA: MIT Press, 1991), p. 198; hereafter cited as MCCCA.

10 For discussion, see Freeman, 'Reason and Agreement', pp. 123-4.


12 Rawls says: 'What these individuals will do is then derived by strictly deductive reasoning from these assumptions about their beliefs and interests, their situation and the options open to them'; A Theory of Justice (Cambridge, MA: Harvard University Press, 1971), p. 119.

13 There is a tendency to hold moral argumentation up to peculiarly high standards, in such a way that the indeterminacy of practical discourse appears to cast a shadow of doubt over the rationality of the procedure. Jean Hampton, for instance, considers it a problem that the argumentation procedure of public contractualism cannot be axiomatized: 'Such indeterminacy makes the contractarian "proof procedure" a messy, ill-defined business, and so reliant on intuitions as to make it appear only a more sophisticated variant of traditional ethical intuitionism.' See 'Two Faces of Contractarian Thought', in Peter Vallentyne (ed.) Contractarianism and Rational Choice (Cambridge: Cambridge University Press, 1991), p. 53. But it seems willfully perverse to demand that moral argumentation be formalizable, when we
have already given up on the idea of formalizing scientific proof procedures. And who would suggest that abandoning the original logical positivist program reduced all modern science to simply a more sophisticated variant of traditional Aristotelian intuitionism?


15 Ernst Tugendhat makes this advantage explicit: 'There is a lower level of *premoral beliefs* which concern the question whether the endorsement of a norm is in the interest of the individual A and whether it is in the interest of individual B, etc. It is now only these premoral empirical beliefs that are being presupposed, and the moral belief that the norm is justified if everybody can agree to it is not presupposed but is the result of the communicative process of justifying to each other a common course of action on the basis of those premoral beliefs.' Christian Gauss lectures (1981), p. 17, cited in MCCA, p. 75.

16 Rawls, Theory of Justice, p. 135.

17 ibid., p. 120.


19 ibid., p. 203.

20 ibid., p. 222.

21 Rawls, Theory of Justice, p. 92.


24 Habermas, 'Discourse Ethics: Notes on a Program of Philosophical Justification', in MCCA, p. 93. Two formulations of U are given in this text, but the English translation makes them seem dissimilar by translating the phrase *jeden Einzelnen* as 'everyone' the first time and 'each individual' the second. I have modified the translation here to read 'every individual'.


26 MCCA, p. 65.

27 A number of critics have focused on this point. Seyla Benhabib argues that U puts an undesirable consequentialist spin on discourse ethics (that is not implied by D); see 'Afterword' to Seyla Benhabib and Fred Dallmayr (eds), The Communicative Ethics Controversy (Cambridge, MA: MIT Press, 1990), p. 343. Charles Taylor has also noted the attempt to reconcile 'Kantian universality and the Benthamite refusal to decide for other people what is right for them'; Sources of the Self, p. 87.

28 MCCA, p. 67. He repeats the same claim in response to McCarthy, JA, p. 90.

29 I say prima facie because other theorists have tried to work around this problem. In his attempt to provide a formal derivation of U, William Rehg suggests that norms must play a role in licensing interests: 'In the language of interests, the question is whether an understandable interest is also a common or general interest. It is precisely moral norms which provide the warrant for such a transformation.' See his 'Discourse and the Moral Point of View: Deriving a Dialogical Principle of Universalization', Inquiry 34(1991): 27–48 at 35. Habermas has endorsed Rehg's argument (JA, p. 32 n, and Faktizität und Geltung [Frankfurt: Suhrkamp, 1992], p. 140 n; hereafter cited as FG), even though it has some difficulties. Rehg accepts that an interest rendered understandable through common values becomes morally relevant only when it is licensed by a norm. A moral norm regulates conflict by assigning certain values, and thus certain needs and interests, *priority over others* (Rehg, ibid., p. 35). But then Rehg goes on to argue that justifying a norm requires showing that it 'coordinates action in accordance with an interest-regulating value that has priority for all' (p. 43). This is clearly circular, since it is the norm itself that establishes the priority of values. Unless one grants that values can establish priorities without normative license, interests become secondary (the formulation collapses into the norms-in-norms-out schema). But in this case, the problem of conflicting values becomes a serious issue. Thomas McCarthy has raised objections to U along these lines in 'Practical Discourse', Ideals and Illusions (Cambridge, MA: MIT Press, 1991) and 'Legitimacy and Diversity' (forthcoming).

30 MCCA, pp. 68–76.

31 Gauthier, Morals by Agreement, p. 144.

32 ibid., p. 144.


34 Thus the outcome 'turns on arguments showing that the interests incorporated in contested norms are universalizable as such', FG, p. 200 (all FG translations are by William Rehg). This gloss dominates Faktizität und Geltung. Kenneth Baynes accepts this interpretation as definitive in The Normative Grounds of Social Criticism (New York: SUNY Press, 1992), p. 149.
35 The role of abstraction is discussed in JA, pp. 90–1.
36 MCCA, p. 86.
40 MCCA, pp. 78–86.
41 Apel lays out the issues nicely in his ‘Normatively Grounding “Critical Theory” through Recourse to the Lifeworld?’, in *Philosophical Interventions in the Unfinished Project of Enlightenment*.
42 In this way it is completely unlike the categorical imperative, although many have tried to apply it in the same way; e.g. see Seyla Benhabib’s ‘Afterword’ to *The Communicative Ethics Controversy*, pp. 345–6.
44 FG, p. 195.
45 This is the portion of Scanlon’s principle which is relevant to the justification of norms; see his ‘Contractualism and Utilitarianism’, p. 110. It is basically the same as Habermas’s D (see MCCA, p. 66), except that Scanlon suggests that it holds only among participants who are moved by the desire to find and agree on principles. If, as Habermas claims, there is a telos of mutual understanding built into communicative action, then there is no need to posit a special desire of this type. For alternate discussion of the parallels between Habermas and Scanlon, see Baynes, *Normative Grounds*, pp. 115–18.