Does Hegel's Critique of Kant's Moral Theory Apply to Discourse Ethics?[1]

Gordon Finlayson

And because we are born with the capacity to persuade
each other … not only have we escaped the life of beasts
but by coming together have founded cities, and made laws
and there is scarcely any institution devised by us which
speech has not helped us to establish.

Isokrates

Several years ago Jürgen Habermas wrote a short answer to the question: "Does Hegel's Critique of Kant apply to Discourse Ethics?" The gist of his short answer is, "no". Insofar as Hegel's criticisms of the formalism and abstract universalism of the moral law never even applied to Kant's moral theory in the first place, they also fail to apply to discourse ethics. Insofar as Hegel's criticisms of the rigorism of the moral law and of Kant's conception of autonomy do hit the mark, discourse ethics successfully draws their sting by reconceiving Kant's moral standpoint along the following lines. **1.** Kant wrongly undertakes to establish the moral law as a "fact of reason": discourse ethics derives the moral standpoint from two premises — one formal, a rationally reconstructed logic of argumentation, and one material, namely our intuitions about how to justify utterances. **2.** Kant wrongly contends that we must be able to think of ourselves as both intelligible characters, inhabiting a noumenal world, and as empirical characters inhabiting the world of appearances: discourse ethics allows that in everyday contexts of action and in the context of moral discourse we have one character that has real needs and interests. **3.** Kant is also mistaken in arguing that moral autonomy requires human beings to abstract away from their needs and interests and to will universalizable maxims for the sake of their universal form: discourse ethics understands moral autonomy to consist in the free adoption of a standpoint from which conflicts of interest can be impartially regulated, by giving special weight to the satisfaction of universalizable interests. **4.** Kant misconceives the categorical imperative as an objective test of universalizability that is applied by individual wills in isolation: discourse ethics reconceives the moral universalism as an ideal of intersubjective agreement of participants in discourse. On the differences between the principles of discourse ethics and Kant's categorical imperative Habermas is wont to cite McCarthy's summary of his — Habermas' — position: "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 the 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" (MCCA 67).

Contra Habermas, I shall argue that some of Hegel's objections to Kant do apply to discourse ethics. Habermas' denial fails to appreciate the way in which Hegel's criticism applies to Kant's ethics, and so ignores the ways in which Hegel's objections apply *mutatis mutandis* to discourse ethics. I show this in the final section of four. The first three sections are indispensable

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

preliminaries that, respectively, introduce the principles and presuppositions of discourse ethics (**1.**), report Habermas' reasons for denying that Hegel's criticisms of Kant's moral standpoint apply to discourse ethics (**2.**), and expound my understanding of Hegel's objections to Kant emphasising where these differ from Habermas' interpretation of them (**3.**).

1. Presuppositions and Principles of Discourse Ethics

The argument I will offer targets the formulation and the function of the principles of discourse ethics not their derivation. Indirectly it bears also on the presuppositions of these principles as outlined in Habermas' *Theory of Communicative Action*, that is, on discourse meta-ethics. I shall begin by saying something introductory about Habermas' by no means uncontroversial theory of language use, if only in order to explain what Habermas means by "discourse".

Habermas begins from the assumption that, "The process of reaching agreement inhabits human speech as its telos" (TKH 1 387:TCA1 287). According to Habermas there are always two dimensions to reaching-agreement: the propositional content of an utterance, that picks out what we reach agreement about, and the performative use of an utterance that picks out the way in which one subject reaches agreement with another. Habermas analyses this performative dimension of speech as follows: it is a feature of every speech act that it necessarily, albeit implicitly, raise three different claims to validity, with respect to the truth of the proposition, the rightness of the utterance, and the truthfulness of the speaker.[2] Habermas contends that the interpreter of an utterance is always in principle free to take up a "yes" or "no" stance to the validity-claims raised by an utterance, and thus to accept or reject it. When a validity-claim is challenged with respect to its truth or its rightness, participants enter the "context of discourse", a reflective mode of communication in which the speaker will attempt to justify the disputed validity claim by adducing grounds or reasons for its rightness (or truth).

It is worth emphasising Habermas' conclusion here, that in order to understand the meaning of any utterance I must be able to accept the reasons which would 'redeem' its validity-claim in discourse. Meaningful utterances must satisfy certain reciprocal conditions of acceptability. The important point is that, according to Habermas, the conditions of the acceptability of validity claims to truth and those of validity claims to rightness are more or less isonomic. In other words, truth claims can be justified in theoretical discourse in much the same way that normative claims can be in practical discourse. Habermas establishes the cognitivism of discourse ethics on the basis of the analogy between truth and normative rightness. This means that Habermas is not a cognitivist in the usual sense, insofar as he does not think that normative statements are capable of being true or false, in the same way that descriptive statements are. But they are like such statements insofar as they can be right or wrong and stand in need of justification: "for normative statements a claim to validity is only analogous to a truth claim" (**MCCA** 76, 68 & 56). Habermas thus salvages the cognitivist intuition that normative statements can be wrong or right, but not at the price of moral realism.[3]

"Discourse", in this sense, is not a synonym for language or speech. "Discourse" is nothing else than the attempt in speech to reach a reasoned consensus on the basis of mutually acceptable reasons over disputed validity claims.[4] A reasoned consensus is a consensus that would

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

be reached by participants, under ideal conditions, that is, if they argued long enough and well-enough and were constrained, by nothing external to the discourse situation, only by the unforced force of the better argument. "*Practical* discourse" refers to the process of moral argumentation, the attempt in speech to reach a rationally motivated consensus over a disputed validity claim to normative rightness. If such disputes are successfully resolved then a consensus is reached that flows back into contexts of action and allows the participants confidently to base their actions on justified and mutually acceptable norms of behaviour.

What makes the theory of discourse attractive to the modern ethicist is that the conception of argumentation which it thematises already harbours a stock of normative rules and commitments that can serve as a premise of a normative moral theory that is not already ethically weighted. Habermas claims that, if one formally reconstructs the pragmatic conditions of discourse, one can identify a set of ideal conditions that every competent speaker, who believes herself to be engaged in argumentation, must suppose to be satisfied. These rules of discourse formalise the intuitive know-how of participants in discourse. Analysis of these rules show that, amongst other things argumentation in principle excludes no-one, renders no assertion immune from question and criticism and prohibits the use of all coercion except the unforced force of the better argument. These rules have the status of idealising presuppositions that are necessarily invoked by all participants in discourse.[5] The thought here is familiar from Kant and elsewhere, that if you will the end you will the means to this end. As soon as you are engaged in discourse or argumentation you implicitly accept both the ideal aim of discourse — to reach a rationally motivated consensus — and the means of achieving this aim — the rules of discourse. It is important to note that the aim of discourse and its necessary presuppositions are ideal, and are invoked as ideals by participants in discourse. Of course real dialogue situations are limited in time and depend on the participants' finite capacity for reasoning, and thus can only approximate these ideals. Still, participants in real discourses must really, if only implicitly, take themselves to be aiming to reach a consensus that would be reached by participants under ideal conditions. Indeed, even if their actual conduct falls short of this aim, and they flout the rules of discourse from within, say they exclude some participant, refuse to listen to their interlocutors, or threaten them with force, then, insofar as they believe themselves to be engaged in a process of rational argumentation they are committing a "performative contradiction" by violating the very rules they implicitly enjoin.

For present purposes we need only bear in mind the main idea that there is a normative core to the conception of communicative action and to the reflective form which it takes in discourse. "The ideas of justice and solidarity are already *implicit* in the idealising presuppositions of communicative action, above all in the reciprocal recognition of persons capable of orienting their actions to validity claims" (**JA** 50). One consequence of this is that Habermas' modest conception of normative moral theory is partly premised, as he readily concedes, on the "outrageously strong" empirical claim that a "universal core of moral intuition" is germane to all forms of life in which action is co-ordinated by communication (**AS** 201).

Taken alone, even this strong generalisation cannot ground a normative moral theory, for fundamental moral principles do not follow directly from the presuppositions of argumentation. If

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

they did, he suggests, it might be clear why these moral principles were binding within practical discourse, but it would remain wholly obscure why they should be binding on actions outside the discourse situation. Hence he adduces a much richer second premise about the level of cognitive competence that is possessed by participants in discourse. Participants in discourses must know intuitively "what it means to discuss hypothetically whether norms of action should be adopted" (**MCCA** 92 & 198). What this second premise amounts to is the ability to abstract from the contingency of one's own perspective — from one's needs, values and interests — and to empathize with others and their needs, values and interests.

Habermas contends that from these two premises the principle of universalisation (U) can be derived by "material implication". His own work on discourse ethics contains no more than a promissory note regarding the derivation of (U), though others have stepped into the breach.[6] Since this is not my concern here, I will assume that (U) can be derived as Habermas claims. (U) is a moral principle that links rationally motivated consensus (concerning the validity of norms) with agreement about the existence of universal interests that would be satisfied by them. (U) states that a norm is valid if and only if

> all affected can freely accept the consequences and the side effects the general observance of a controversial norm can be anticipated to have for the satisfaction of the everyone's interests (MCCA 65).

The idea captured by (U) is quite straightforward. (U) projects the ideal of a discourse in which everyone affected by the implementation of a norm would have equal say in its adoption. Moreover each participant can consent to a norm only if she has reason to. What constitutes a reason to freely accept a norm is something I shall discuss in section **4.** Here, it suffices to note what looks like Habermas' position: that everyone has reason to consent to a norm if the norm's general implementation would satisfy their interest.

What are the alternatives? One alternative is that I simply agree — i.e. blindly volunteer — to be bound by all norms that meet a certain condition, namely that they look like satisfying everyone's interest. This would introduce a large dose of decisionism into the process of reaching agreement. Why do I agree to be bound by norms that meet this condition rather than some other? Another alternative that must be discounted is that I agree for moral reasons to be bound by those norms that look like they are in everyone's interest. In this case discourse ethics would be circular. It would presuppose the moral standpoint, not explain it.

The most natural and plausible reading of (U), then, links the free acceptance of the consequences of a norm in each case with the satisfaction of one's interests; in other words one's interest in a norm gives everyone a *pro tanto* reason to accept it. In discourse such reasons are tested from all perspectives and those that rest on interests that are particular to some participants are "ultimately discarded as not being susceptible to consensus" (**MCCA** 103). (U) thus posits a kind of ideal end point of moral inquiry in which agreement would settle on only those norms that were equally in the interests of all. It represents the standpoint of impartial judgment from which everyone's interests would be weighed.[7]

A propos impartiality, the impartiality aimed at by discourse ethics, unlike Kant's ethics,

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

does not require participants to abstract away from their own interests, and to base their actions on a "purely moral interest" or "reverence for the law". Nor is it gained by taking up a neutral, third person perspective on one's own interests, as recommended by Adam Smith, in his *Theory of The Moral Sentiments.*[8] In these cases the attainment of impartiality gives rise to the problem of dissociation between the self that considers its interests and needs and the self that has those needs.[9] Habermas argues that the intersubjective ideal of impartiality embodied in principle (U) does not fall prey to the critique of the disinterested moral self. Here, impartiality is reached by a process in which every participant in discourse empathetically attempts to occupy the standpoint of all others affected by a norm, and imaginatively to identify with their needs and interests from within their standpoint. The idea is not to adopt an external perspective of a spectator on one's interests; not to dissociate oneself from one's interests, but through the empathetic occupation of the standpoint of all others, to associate oneself with theirs.[10] The impartiality here is not what is demanded of neutral arbiter, but what is gained through the mutual adoption, adjustment and integration of multilateral perspectives. Through this process one is able to check whether or not one can universalize from the existence of an interest in one's own case to its existence in all other cases, too.

> Only under this social-cognitive presupposition can each person give equal weight to the interests of the others when it comes to judging whether a general practice could be accepted by each member on good grounds, *in the same way that I have accepted it.*[11]

2. Habermas on Hegel's Kant Critique

I am going to look at two of Hegel's objections to Kant that Habermas thinks do not apply to discourse ethics: the objection to formalism and the abstract universalism of Kant's moral theory. According to Habermas the objection to formalism runs:

> Since the moral principle of the categorical imperative requires that the moral agent abstract from the concrete content of duties and maxims, its application necessarily leads to tautological judgments (**MCCA** 195).

The objection to abstract universalism runs:

> Since the categorical imperative enjoins separating the universal from the particular, a judgement considered valid in terms of that principle, necessarily remains external to individual cases and insensitive to the particular context of a problem ... (ibid).

Habermas rejects Hegel's attempted knock-down objection that the formalism of the moral law implies emptiness by pointing out that, on Kant's theory and his own, the moral principle is both formal and not empty. The procedure captured by (U) is "not formal in the sense that it abstracts from content. Quite the contrary ... practical discourse depends on contingent

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

content being fed into to it from outside" (**MCCA** 103). It is thus not true that a formal moral principle can have no purchase on the "substantive problems of everyday life"; it bears directly on all those problems that concern norms embodying "universalizable" interests (**MCCA** 204). What is notable here is that Habermas largely accepts Kant's conception of the will as a tester of maxims that Hegel criticises in the *Philosophy of Right*. Indeed he invites the comparison between the way in which content is fed in to the formal testing procedures represented by (U) and the formula of the universal law (**MCCA** 204). (U) relates to norms as the categorical imperative relates to maxims. For Kant maxims embody interests: so, for Habermas, do norms.

In response to the second charge of abstract universalism, Habermas argues that both Kant's theory and his own involve abstractions, but that it is not true that "a moral point of view based on the universalizability of norms necessarily leads to the neglect … of existing … interests … " (**MCCA** 204). The abstractions made by a moral theory that concentrates on the question of justification are unavoidable. However this necessary decontextualisation is counterbalanced by a capacity for nuanced, appropriate and context sensitive judgment, whereby justified moral norms are applied to particular cases.

3. Hegel's Criticisms of Kant's Moral Theory

The objection to formalism that Habermas rejects occurs in the 1802 essay on Natural Law; and again, in the *Phenomenology of Spirit*, Hegel claims that Categorical Imperative is a merely logical test of universalizability which any maxim can be made to pass.

> The criterion of law which Reason possesses within itself, fits every case equally well, and is thus in fact no criterion at all (**3** 319, Miller, p. 259).[12]

Habermas is right, this attempted knock-down argument does not stand up. At least some maxims can be made to fail the test, in particular those which contain what Kant calls a "contradiction in conception" (**AA** IV 424), e.g. "I will break promises when convenient". But if some maxims fail the test of universalizability, then Kant can at least show that the negation of those maxims expresses a strict duty: e.g. "Do not break promises when convenient". In which case it is not true that any maxim can be made to pass the test, and hence not true that the Categorical Imperative is empty.

Hegel has a second, more careful objection to the Categorical Imperative, which is that valid moral principles emerge successfully from the test of universalizability only because Kant presupposes the existence of substantive moral values against which the results of the test of universalizability can be weighed. He adduces Kant's example of the man wondering whether he should keep hold of an unrecorded deposit that has been entrusted to his care. Kant declares that, in answering the question of whether or not the maxim, "I shall keep on a deposit entrusted to me whenever the opportunity presents", can be universalised:

> I become immediately aware that such a principle would destroy itself if made into a law, for it would entail that there would be no deposits (V 27).

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

Hegel's response is that there is no contradiction here. It is just as self-consistent to will a world in which no such deposits are made, because property does not exist, as it is to will a world in which deposits are made and property exists.

> Property, simply as such, does not contradict itself ... non-property, the non-ownership of things, or a common ownership of goods is just as little self-contradictory (**3** 317, Miller, p.258; see also **2** 460).

The contradiction arises, argues Hegel, only because Kant presupposes that the moral world should be a property-owning world where deposits can be made and where depositees can be trusted. The maxim can be made self-consistent, but it conflicts with existing values, beliefs and institutions. But the Categorical Imperative was supposed to provide a critical test, through which we can reflectively endorse those of our moral intuitions which are contained in justifiable maxims. The resultant justified maxims, the permissions or prohibitions yielded by the test, are supposed to be valid *a priori*, regardless of context. They cannot be rejected just if they conflict with our untested intuitions. That would compromise the autonomy of the moral law by making it depend upon the heteronomous content of our beliefs, desires and practices.

Hegel's argument, such as it is, is hindered rather than helped by this example. Kant's deposit example does contain contradiction in conception. This contradiction may only come to light because of the meaning of the concept of "deposit", which is interwoven with a set of background assumptions about property rights and relations of trust. Nonetheless a contradiction in conception arises because, when I attempt to universalise my appropriation of the deposit I have at the same time to will the existence of world in which relations of trust obtain between the givers and receivers of deposits, and the existence of a world in which everyone would appropriate deposits if they could, and in which, therefore, such relations of trust would not obtain.[13]

But although Hegel does not choose a convincing example to illustrate his point, he still has a point. The problem is not that any principle which is formal in Kant's sense is empty, but that, so long as the principle is just a test of the universalizable form of the maxim, the results of the application of the test will be insufficiently determinate. The charge that formalism implies emptiness, was the bogus one that any maxim could be made to pass the test; the charge that formalism implies "indeterminacy" is that too many maxims pass the test, so that, alone, it is not sufficient to determine their moral worth. The objection needs to be fleshed out with an example of one of many possible 'rogue' maxims, i.e. a maxim which produces counterintuitive, not to say absurd, results when subjected to the test of universalization: "Always open doors for other people". This is a plausible example of a maxim. It is at least as plausible as any of the examples that Kant himself discusses. Yet, given the fact that two people cannot open the same door for each other, the maxim clearly fails the test of universalization. It would, however, be absurd to conclude that it was therefore morally impermissible always to open doors for other people, or that one had a strict duty not to do so.

One response to the problem of 'rogue' maxims has been to introduce a scope-restriction on what can count as a candidate maxim thereby ensuring that only 'morally relevant' maxims are

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

available for testing by the moral will. This is the response that is usually offered in Kant's defence. Onora O'Neill draws a distinction between "underlying intentions", e.g. to be hospitable to one's guests, and "ancillary intentions", e.g. offering them a cup of tea; she reserves the term maxim for the former. In the same spirit Otfried Höffe draws a distinction between maxims, which are general (subjective) principles of the will and capture the "fundamental normative pattern" of actions, and rules or precepts of action, which reflect more or less arbitrary decisions about how to order one's life, such as to rise early in the morning.[14] I suspect that any scope restrictions on candidate maxims that are sufficiently determinate to rule out examples like "always open doors for other people" as trivial or irrelevant must ultimately refer to their content not their form. But the moral will is supposed to abstract from considerations of content. So this defence of Kant is vulnerable to the objection that, under the guise of redescribing the function of maxims in shaping a life, it smuggles in normative considerations to determine candidature, considerations which are supposed to result from the reflective testing of maxims, not to be fed into it.

In spite of its unpromising formulation in his early works Hegel does have here the lineaments of a good argument against the indeterminacy of Kant's moral standpoint. It is this argument that Hegel, in the *Elements of the Philosophy of Right*, directs against its proper target, not at the categorical imperative itself, but at Kant's whole conception of the will as a tester of maxims. His claim is that the "indeterminacy" of the moral standpoint results from the attempt to settle the question of the validity of moral laws formally, i.e. prior to and independent of their relation to a possible content. Note that Hegel here does not attack a crude caricature of Kant's moral theory as the early Hegel, following Friedrich Schiller, was wont to do. Maxims are not the content onto which moral form is stamped. Maxims are not identical with the 'materials' of moral psychology, however these materials — sensations, feelings, emotions, needs, wants and interests — may rate on the scale of refinement and complexity. Rather, maxims are first order principles of the will that contain and form this material and serve as candidate moral norms that are available for uptake into the moral will. The Categorical Imperative is a second order principle of the will that reflectively selects maxims on the basis of their universalizability and, more importantly, rejects those that are not universalizable. It is this quite sophisticated picture of the will as a tester of maxims against which the later Hegel directs his fire.

It would be wrong, however, to suggest that the later Hegel rejects Kant's conception of the will in all respects. On the contrary, to a great extent he shares Kant's moral psychology.[15] Hegel agrees with Kant that the moral will is autonomous (*PR* §133).[16] He agrees that the autonomous will is one which gives itself a law or adopts a maxim (*PR* §135). He agrees that the law the will gives to itself or the maxim it adopts is universal, not one which merely ministers to particular inclinations (*PR* §137R). He even agrees that the law or maxim be adopted in virtue of its universality, i.e. that the moral agent performs "duty for its own sake" (*PR* §133). And yet he claims that Kant reduces the moral standpoint to an "empty formalism" when he insists that the maxim be adopted *only* for the sake of its universal form and not *also* for the sake of any desires or interests the maxim may advance.

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

> However essential it may be to emphasize the pure and unconditional self-determination of the will as the root of duty — for knowledge of the will first gained a firm foundation and point of departure in the philosophy of Kant, through the thought of its infinite autonomy — to cling on to a merely moral point of view without making the transition to the concept of ethics reduces this gain to an *empty formalism* ... **From this standpoint, no immanent doctrine of duties is possible**. One may indeed bring in material from outside and thereby arrive at particular duties, but it is impossible to make the transition to the determination of particular duties from the above determination of duty as *absence of contradiction*, as *formal correspondence with itself*, which is no different from the specification of *abstract indeterminacy*; **and even if such a particular content for action is taken into consideration, there is no criterion within that principle for deciding whether or not this content is a duty** ... A contradiction must be a contradiction with something, that is, with a content which is already fundamentally present as an established principle. Only to a principle does an action stand in a relation of agreement or contradiction. **But if a duty is to be willed merely as a duty and not because of its content, it is a *formal identity* which necessarily excludes every content and determination** (7, 253, **PR** §135R; my emphasis in bold).[17]

Hegel's argument is that Kant fails to show how the moral will can give itself contentful moral principles whilst remaining truly self-determining and free. Hegel's initial reproach, that Kant can give no "immanent doctrine of duties" cuts deeper than the familiar charge that Kantian morality is deficient in substantial, determinately action-guiding duties. Such an argument could be easily deflected by pointing out that in the *Groundwork* and the second *Critique*, Kant's principal aim is *to justify the moral law*, not to provide a doctrine of determinate duties; this is a task he undertakes later in the second part of the *Metaphysics of Morals* (the 'Doctrine of Virtue').[18] Hegel's deeper point is that, because Kant conceives the autonomy of the moral will as an *a priori* determination of itself, that is, a determination that abstracts from the ends it adopts, he introduces an hiatus between the moral will, with its principle of maxim selection — the Categorical Imperative — and the empirical will, the bearer of the candidate maxims.

The hiatus can be brought into view by considering the reflective structure of the will that Hegel attributes to Kant. I morally will (on the basis of the Categorical Imperative) that I empirically will an end, in adopting a maxim, say, not to make a deceiving promises. The categorical "ought" has its source in the moral will. But it is addressed to the human, empirical will — the will which has the interests of a being that is both rational and sensible. According to Hegel, Kant simply assumes here that there is a partial identity between the moral will and the empirical will, since the former is wholly rational and the latter both rational and sensible. Hegel insists that there is *only a formal identity*, that Kant's moral will is characterised by "formal correspondence with itself" and "abstract indeterminacy" (135R). For the moral will incorporates maxims on the basis of their universal form alone. Thus the moral will remains, ultimately,

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

discontinuous with the "content" or the interest that is contained in the maxim. But the having of a content — an interest — is what distinguishes the empirical from the moral will. So the moral will must be discontinuous with the empirical will: they remain, ultimately, different agencies. The content of the maxim itself exerts no constraints on the adoption of the maxim, only the form of the content. What this means more concretely is that my desire to keep my promises, and my wanting not to let the person to whom I promise down, are not considerations that are morally relevant to the adoption of my maxim. Insofar as these desires form part of the content of the maxim, this content remains a moment of heteronomy within the will or, as Hegel writes elsewhere, "the last undigested lump in the stomach" (**20**, 369). Hegel concludes that this discontinuity obscures the relation of the *a priori* principle to its possible content: "there is no criterion within that principle for deciding whether this content is a duty".

Let me review my reconsiderations of Hegel's Kant-critique. **1.** Hegel's most plausible objection to the categorical imperative is that it captures too many maxims. The results of the testing process are thus not sufficiently determinate to capture our moral intuitions successfully. **2.** Hegel challenges Kant to show how, on his picture, the moral will can acquire any determinate content, if a maxim must be adopted in virtue of its universalizable form alone, and not also in virtue of the interest it contains. This way of putting the point makes Hegel's criticism of "abstract universalism" concerning the rigid separation of universal and particular, into an aspect of the criticism of formalism. But this is not a problem. Rather, Habermas is wrong to suggest that Kant's artificial separation of universals and particulars forms a separate problem, one that arises only in the application of valid moral norms to particular situations. Kant's difficulty is to show how there can be valid, contentful moral norms. Of course there *are* such norms. But Kant cannot show how this is possible. If I'm right, Hegel's answer to this question will involve giving a plausible account of how a universalizable maxim can be adopted in virtue of its form *and* in virtue of its content. And to show that, Hegel will have to give some account of how particular interests acquire universal form. This is the one of the tasks Hegel assigns to the philosophy of objective spirit.

4. A Critique of the Formalism of Discourse Ethics

Do Hegel's criticisms of Kant apply *mutatis mutandis* to discourse ethics, and if so how? We can begin to answer this question by asking whether the results of the test of universalizability in principle (U) would capture enough of our moral intuitions. We know that (U) rules out any norm the general observance of which would not be likely to satisfy "everyone's interests"; that is, (U) rules out all norms that do not embody, to use Habermas' term, a "generalizable" or "universalizable interest". But what counts as "in everyone's interest", or to put it differently, what are the conditions of the universalizability of interests? Looking closely there is a worrying ambiguity in Habermas' formulation and subsequent explanations of (U). A norm is not valid unless:

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

> all affected can *freely* accept the consequences and the side effects its *general* observance can be anticipated to have for the satisfaction of everyone's interests (**MCCA** 65).

The ambiguity is captured by a different translation of the principle later in the English translation of the same work.

> For a norm to be valid, the consequences and side effects of its general observance for the satisfaction of *each person's particular interests* must be acceptable to all (**MCCA** 197).

The text should read "for the interests of each" and not "for each person's particular interests". Habermas' formulation of (U) does not specify that the interests, the hypothetical satisfaction of which constitutes the validity of a norm, be particular.[19] Nonetheless this interpretation of (U) is open. (U) could mean that assent is conferred on a norm only if everyone has *an* interest in its general observance, but not necessarily the same interest, which means that people can assent to a norm for different reasons. I shall call this the unofficial interpretation. Alternatively (U) could mean that assent is conferred on a norm only if everyone has *one and the same interest* in its general observance, which would imply that a rationally motivated consensus about a norm can be reached only if everyone can freely accept it for *one and the same reason.* I shall call this the official version.

Given what Habermas argues elsewhere, discourse ethics must rule out the unofficial versions of (U). The second translation of (U) cited above turns out to contain an egregious misunderstanding of the idea of an ethics of discourse. Why is this? After all, it is not implausible to claim that a valid norm must satisfy *an interest* of everyone affected by its general observance, though not necessarily the same one. But such an interpretation of (U) implies that valid norms can satisfy different interests for different people. Now Habermas accepts, unlike Kant, that we have reason to consent to norms because they satisfy our interests. So it would follow from the unofficial interpretation of (U) that different people can consent to a universally valid norm for different reasons. For example, we can imagine a small self-sufficient farming co-operative consisting of vegetarians who also happen to be atheists, and religious believers who happen not to be vegetarian, all agreeing that it is wrong to eat pork. The norm that one ought not to eat pork commands universal assent in spite of the fact that the reasons for the norm are not themselves universally recognised as valid, but are rather relative to some other context — vegetarianism and religion respectively. Indeed, assent is universal in spite of the fact that, since this context is not generally shared, neither group can be persuaded by the reasons advanced by the other party. Yet, on the unofficial interpretation of (U) the norm would be valid, since it would pass the test of universalization.

The unofficial version validates a norm on the basis of a contingent overlap of particular interests, on condition that all participants in discourse can judge that everyone affected by the norm has some interest in its general observance. However, (U) was supposed to function as a criterion that would enable participants in discourse to distinguish sharply between rationally

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

motivated consensus about norms and merely *de facto* consensus about precepts that strategically promote the merely particular interests of some group. Clearly, on the unofficial interpretation, (U) cannot fulfil this function. It can distinguish only between *de facto* universal consensus and *de facto* dissensus.

Worse still, if (U) permits the validity of norms to rest on a merely contingent overlap of interests, rather than on the universal acceptability of good reasons, a mainstay of discourse meta-ethics — the analogy between normative rightness and truth — breaks down. Newtonian and Einsteinian physicists might agree that light consists of particles rather than waves, but they would do so for very different reasons. Even if, as a consequence, they might agree, say, that light travels in straight lines, this agreement would be coincidence and not convergence. It would not be "rationally motivated", amenable to consensus only on the basis of good reasons. Justifying reasons as well as the conclusions they warrant must also converge, if we are to speak of convergence at all. If normative rightness is analogous with truth, the same must hold of a rationally motivated consensus about norms.

What about the official version, according to which a norm is valid if and only if its general observance can be anticipated to satisfy one and the same interest everybody shares?[20] There is much evidence that Habermas must have the official version in mind. *To begin with,* he frequently equates "generalizable" or "universalizable interests" with "the common interest" (**MCCA** 65 & **JA** 13), with the "common will" (**MCCA** 67), with shared needs (**LC** 107) or with what all can want: "the interest is common because the constraint-free consensus permits only what all can want ..." (**LC** 110).[21] He also slips from talk of "generalizable interests" to talk of the "general interest" (**MCCA** 104), from interests which are possibly common to all, to those which actually are. [22] The implication is that an interest is universalizable only if all have — that is, only if everyone has — or can take themselves to have, an identical interest in the existence of a behavioral norm.

Secondly, Habermas equates the moral standpoint with impartial judgement. Universalizable interests give *impartial* reasons to assent to norms, i.e reasons that everybody else can have, too. "True impartiality pertains only to that standpoint from which one can universalize precisely those norms that can count on universal assent because they perceptibly embody an interest common to all affected" (**MCCA** 65, 198).[23] *Thirdly* and decisively, Habermas claims that (U) allows participants in discourse, on the basis of the distinction between universalizable and particular interests, to draw a "razor sharp" distinction between "evaluative" and "normative" statements, between values and norms, between questions of the good life and questions of justice (**MCCA** 104 & 204). Values embody particular interests, norms embody universalizable interests. The discursive process is one in which participants necessarily abstract from all interests that are unique to them: "particular values are ultimately discarded as being not susceptible to consensus" (**MCCA** 103). Particular interests and the values they underwrite do not command the same authority that post-conventional moral agents confer on universalizable interests. Any consensus they underwrite will be relative to some conception of the good life (**JA** 9).

The problem is that, whilst the unofficial version of (U) was too weak to fulfil the

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

function assigned to it, the official version is now too strong. The problem is not that such interests exist — interests which are universalizable in the sense that we can confidently expect them to be common to all. Everyone has a shared interest, say, in breathing unpolluted air or in being treated justly. It does not really make sense for me to a have an interest in *just my* being able to breath unpolluted air, or in justice only *in my own case*. To have an interest in air or justice is, given the peculiar nature of each, to have an interest in clean air and justice for everybody. Nor does the problem lie in the claim that a principle linking rationally motivated consensus to the existence of one and the same universalizable interest is a sufficient condition of the justifiability of moral norms. For, it is plausible to claim that whenever everyone freely consents to a norm, because it satisfies an identical interest which each can take themselves and everyone else to have, then that norm is valid. The problem lies in the claim that (U) contains a necessary condition of justifiability, that a norm cannot be justified unless it meets the condition set by (U).

For one thing, as many commentators have remarked, the number of justifiable moral norms that could meet such a stringent condition would be too few.[24] Habermas apparently does not see this as a problem. There just are not many norms that can be justified in practical discourse. But this surely is a problem for a modest conception of moral theory that claims merely to clarify and explicate our intuitions about morality. For surely even Habermas does not believe that justifiable moral norms are as scarce as all that. If he did, one would think that moral theory would play very minor role in his theory of practical reason, not, as it does, occupy centre stage. For another, the official version upsets the analogy with truth and thus threatens a central tenet of discourse meta-ethics. According to *The Theory of Communicative Action* (**TCA1** 297-8), we cannot reach agreement in theoretical discourse over a validity claim to the truth of a proposition, unless I can *accept or recognise* the reasons you adduce for it, not unless I *share those reasons or have them myself.* Suppose you believe it is midday because you hear the clock strike, and I who cannot hear the clock, look at my watch. I do not and cannot *share or have your reason* for believing it is midday, but we could still reach a rationally motivated consensus that it is, because I can recognise or accept your reason. There are many routes to the truth. We do not have to take the same one. Why then, should participants in practical discourse have to assent to a norm on the basis of their all having the same interest, and thus the same reason to agree to it? The official version of (U) makes the conditions of the possibility of reaching consensus about normative claims in practical discourse stronger than the conditions of the possibility of reaching consensus about truth claims in theoretical discourse.

These objections to Habermas' too demanding conception of the justifiability of norms are not directly equivalent with Hegel's objection to the "abstract indeterminacy" of Kant's categorical imperative and of Kant's formal conception of the will. For, while Kant's problems arose from he fact that the categorical imperative justified too many maxims, requiring the introduction of *ad hoc* scope restrictions, principle (U) justifies too few. But Kant and Habermas are vulnerable to similar criticisms directed towards the same problematic area — the way in which content is given to the formal moral principle.

To be more precise, Habermas' problems stem from the way in which discourse ethics

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

tries to respond to the problem of the identity of the moral and the empirical will. He does this by making universalizability not just a rational requirement on each individual's will, but a constraint exerted also by the content of collective willing. He comes very close to Hegel's demand that a duty is willed "not just as a duty but also because of its content" (**PR** §135R). In the process, however, Habermas falls foul of Hegel's criticism of "abstract universalism" — the rigid separation of universal and particular. There is a slight difference here. For Kant, in Hegel's eyes, is guilty of separating universal form from particular content, whereas the procedural principle (U) separates universalizable content from particular content.

The result, however, is similarly incapacitating. So long as Habermas stands by his "razor-sharp" distinction between values and norms, so long as the official interpretation of (U) entails that interests be exclusively *either* particular *or* universalizable, no actions based on particular interests can find normative justification. This rules out too may *prima facie* candidates. My interest in avoiding *my* pain, or your interest in caring especially for *your* loved ones are universalizable in the sense that everyone has an interest in avoiding *their own* pain or caring for *their* loved ones. But these are at least numerically different interests, since they contain non-identical pronominal referents in each case. No-one else need have an interest in my avoiding pain *to me*, or in caring for *my* children, or in being the one who chooses *my* mother's birthday present. These interests are particular in the sense that they either have objects that are different in each case or they belong to different people in each case, or both.[25] The reasons which these particular interests give us are agent-relative; nonetheless they — the interests and the reasons — are still universalizable. It seems strangely disabling for an ethics of discourse that is avowedly universalist and deontological to deny that these agent-relative reasons given to us by our particular interests can justify moral norms even though they are clearly universalizable, when this claim speaks against the very intuitions that it sets out to explain and clarify.

If I am right here, then discourse ethics is beset by a dilemma. The unofficial version of (U) is too weak and the official version too strong to do any real work in determining contentful moral norms. The simple answer would be to weaken (U) to somewhere in between, so that it would disallow a consensus based on a contingent overlap of dissimilar interests, but allow a consensus based on universalizable agent-relative interests. In other words, the most pressing task for Habermas' ethics of discourse is a clarification of the opaque notion of "universalizable interests," which would allow that certain particular interests can be universalized and are thus specially reason giving from the moral standpoint. Weakening (U) in this manner, however, would require some far-reaching adjustments elsewhere in the theory of discourse ethics. To begin with Habermas would have to redraw his array of strict distinctions between norms and values, justice and the good-life, and between *Moralität* and *Sittlichkeit*. Interestingly, such a move would push Habermas away from a neo-Kantian position in which norms/justice/morality and values/the good life/ethical-life are located in separate, but complementary spheres, towards a more orthodox Hegelian conception of ethical-life and a dialectical account of the good, as "the unity of the concept of the will and the particular will" (**PR** § 129). It is not my argument that there can be no useful differentiation of norms from values, justice from the good-life, morality from ethical-life. My argument is that Habermas' differentiations are untenable because they rest

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

ultimately on the confused distinction between universalizable and particular interests.

As a matter of fact Habermas is well aware that he needs to close the hiatus between norms and values, justice and the good, morality and ethical life, that has opened up, if I'm right, largely as a result of his analysis rather than of historical and social forces. This hiatus comes clearly and crucially into view when discourse ethics attempts to answer the question: in virtue of what are universalizable interests especially worthy of recognition from the moral standpoint? Either the answer is that communicative subjects already have an moral interest in recognising universalizable interests as specially reason-giving from the moral standpoint, which is circular, or discourse ethics is forced to say, as Habermas concedes, "something relevant about substance as well ... about the hidden link between justice and the common good" (**MCCA** 202). That is, we have to offer some account of why it is good to be moral. But if the domain of ethical life, if ethical questions of the good, consist ultimately in values underwritten by particular interests or subjective preferences, what can be said about this connection that is not merely descriptive, value-laden and culturally parochial? Habermas is right that the link between morality and the good remains hidden, but it is his own taxonomy that is the source of the obscurity.

In fact, Habermas' unfulfilled demand that the interrelation between morality and ethics, justice and the common good be made clear, is nothing less than a Hegelian insight, a call to uncover a suppressed dialectical relation between a Kantian dichotomy of particular and universal interests. Unlike the moral theories of Kant and Habermas, Hegel's philosophy of objective spirit, which culminates in the moment of "ethical life", emphasises the continuity between the moral content — the interests embodied in the candidate maxims — and the moral form — the principle of the will. Roughly speaking, it does this in the form of a narrative, in which rational human subjects reflectively revise, refine and realign the particular desires, interests and ends they pursue in concert with others within the framework of their social and political practices and institutions. The point is that, in this framework, the reflective pursuit of particular interests can advance and sustain more universal interests, in such a way that formal considerations of fairness, reciprocity and universality come to accrue enduring recognition. This summary is no doubt too brief and too vague to be convincing. I have not undertaken to defend it here. But, if I have shown how, pace Habermas, Hegel's criticism of Kant's moral theory can be applied to discourse ethics, I will have gone some way to demonstrating the enduring relevance of Hegel's ethical insight for the criticism of morality.

Gordon Finlayson
University of York

[1] Abbreviations of Habermas' works referred to here are as follows:
BFN = *Between Facts and Norms* (Cambridge: Polity Press, 1997).
CES = *Communication and the Evolution of Society* (London: Heinemann, 1979).
DEA = *Die Einbeziehung des Anderen* (Frankfurt a/M: Suhrkamp, 1996).
ED = *Erläuterung zur Diskursethik* (Frankfurt a/M: Suhrkamp 1990).
EI = *Erkenntis und Interesse*, (Frankfurt a/M: Suhrkamp, 1973).
JA = *Justification and Application*, (Cambridge: Polity Press, 1993).
LC = *Legitimation Crisis* (London: Heinemann: London, 1976).
MCCA = *Morality and Communicative Consciousness* (Cambridge: Polity Press, 1990).

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

MKH = *Moralbewusstsein und Kommunikatives Handeln* (Frankfurt a/M: Suhrkamp, 1986).
OCCM = "On the Cognitive Content of Morality", *Proceedings of the Aristotelian Society*, 1997.
PMT = *Postmetaphysical Thinking* (Cambridge: Polity Press, 1992).
SE = "Sprechakttheoretischer Erläuterungen zum Begriff der kommunikativen Rationalität", in *Zeitschrift für Philosophische Forschung* 50 (1996): 65-91.
TCA = *Theory of Communicative Action*, two vols. (Boston: Beacon Press 1984 & 1987).
TKH = *Theorie des Kommunikativen Handelns* (Frankfurt a/M: Suhrkamp 1982).
VE = *Vorstudien und Ergänzungen zur Theorie des Kommunikativen Handelns* (Frankfurt a/M: Suhrkamp, 1984).
WT = "Wahrheitstheorien" in *Vorstudien und Ergänzungen zur Theorie des Kommunikativen Handelns* (Frankfurt a/M: Suhrkamp, 1984).
WUP = "What are Universal Pragmatics" in *Communication and the Evolution of Society* (London: Heinemann, 1979).
Other abbreviations:
AA = refers to the Prussian Academy Edition of Kant's Complete Works, vol. IV, Berlin, 1902
AS = *Autonomy and Solidarity* ed. P. Dews (London: Verso 1992)+0, p.194.
HCD = *Habermas: Critical Debates*, J. Thompson & D. Held eds. (London: MacMillan, 1982).

[2] Originally Habermas outlines four validity claims, the fourth being that of intelligibility, but he soon pares it down to three. (**WUP** 2) These three validity-claims correspond to the three types of illocutionary act which Habermas' suggested taxonomy of speech acts allows — constatives, regulatives and expressives. (**TKH1** 443: **TCA1** 322) These in turn relate to the three value-spheres which structure the life-world, the scientific-technical, the legal-moral and the aesthetic-expressive. Since our concern is not with the claim to truthfulness, and its related value sphere the aesthetic-expressive, the claims to truth and rightness are my sole concern here.

[3] "I defend a cognitivist position ... namely that there is a universal core of moral intuition ... In the last analysis, they stem from the conditions of symmetry and reciprocal recognition which are unavoidable presuppositions of communicative action. ... Any attempt ... to defend a cognitivist-universalist ethical theory involves the public assertion that in your own society and in others all practical and political questions have a moral core which is susceptible to argument". (**AS** 201)

[4] **TCA1** p.42.

[5] **MCCA** 87-94. **JA** 50, 55-6. For an elaboration of the premises in the derivation of (U) see W. Rehg, "Discourse and the Moral Point of View: Deriving a Dialogical Principle of Universalisation", *Inquiry* 34 (1991): 27-48.

[6] Habermas does not provide the derivation, rather he states that such a derivation is possible. Some of the difficulties posed by the derivation of (U) are unearthed by W. Rehg *op. cit.*. In particular the second premise brings culturally specific and value-laden assumptions into play, assumptions about the moral relevance of interests and needs. But this threatens to blur the strict distinction Habermas wishes to draw between values and norms, between moral questions of justice and ethical questions of the good. See below.

[7] "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". (**MCCA** 65, 198: **JA** 12-13) Elsewhere Habermas

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

claims that the principles of discourse ethics explicate the moral standpoint, i.e. "the point of view from which norms of action can be impartially grounded" and that moral discourses aim at "the impartial evaluation of action conflicts". (**BFN** 97)

[8] "We must view them (his interests and my interests) neither with our own eyes nor with his, but from the place and with the eyes of a third person who has no particular connection with either and who judges with impartiality between us". Adam Smith *Theory of the Moral Sentiments* III 3.3 cited from David Wiggins "Universality, Impartiality, Truth" in *Needs Values Truth: Essays in the Philosophy of Value* (Oxford: Blackwell, 1987), pp. 74.

[9] B. Williams, *Ethics and The Limits of Philosophy* (London: Fontana, 1985).

[10] See **MCCA** 182 & "Individuation through Socialisation. On George Herbert Mead's Theory of Subjectivity", **PMT** 179-188.

[11] "Justice and Solidarity: On the Discussion Concerning 'Stage 6'" in *Philosophical Forum* XXI (1989-90): 39 (my emphasis).

[12] All references to G.W.F.Hegel, *Werke in zwanzig Bänden,* eds. E. Moldenhauer and K. Michel, (Frankfurt a/M: Suhrkamp, 1986). (Volume number in bold, followed by page number.)

[13] See C. Korsgaard "Kant's Formula of the Universal Law", in *Pacific Philosphical Quarterly,* 66 (1965): 31; See also A.W. Wood, *Hegel's Ethical Thought,* (Cambridge: Cambridge University Press, 1990), p.157.

[14] O. O'Neill "Kant after Virtue" in *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), pp. 145-65. O. Höffe *Immanuel Kant* (Albany: SUNY Press, 1994), pp. 149-51; H. Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp.85-94.

[15] On this see Lottenbach and Tenenbaum, "Hegel's Critique of Kant in the Philosophy of Right", *Kant-Studien* (1995): 219-21.

[16] The moral will here refers to what Kant terms "*Wille*" as opposed to "*Willkür*". See John R. Silber, "The Ethical Significance of Kant's Religion" in *Religion Within The Bounds of Pure Reason Alone* (New York: Harper and Row, 1960), pp. xciv-cvi, and H. Allison, *op. cit.* pp.129-36.

[17] I have used the excellent English translation by H.B.Nisbet, ed. A.Wood, *Elements of the Philosophy of Right* (Cambridge: Cambridge University Press, 1991), pp. 162-3.

[18] The task Kant sets himself in these works is that of justifying the moral law. Sally Sedgwick makes this point comprehensively in "On the Relation of Pure Reason to Content: A Reply to Hegel's Critique of Formalism in Kant's Ethics" in *Philosophy and Phenomenological Research* XLIX, 1 (1988).

[19] The translator is not wholly at fault here, since his rendering, although inaccurate, is certainly permitted by an ambiguity in the formulation of (U) in the German:
daß die Folgen und Nebenwirkungen, die sich jeweils aus ihrer *allgemeinen* Befolgung für **die Befriedigung der Interessen eines *jeden* Einzelnen** ... ergeben, von *allen* Betroffenen akzeptiert ... werden können. (**MKH** 75; **ED** 12 my emphasis in bold)

[20] Felmon John Davis thinks discourse meta-ethics requires the official version: "parties must have the same reason (to the same degree) to agree". He does not note that this would have fatal implications for Habermas' overall theory. "Discourse Ethics and Ethical Realism: A Realist Realignment of Discourse Ethics", *European Journal of Philosophy* 2, 2: 125 -43.

[21] Since all those affected have, in principle, at least the chance to participate in the practical deliberation, the "rationality" of the discursively formed will consists in the fact that the reciprocal behavioral expectations raised to a normative status afford validity to a

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press

common interest obtained without deception. (**LC** 110) Cf. also **WT** 173-4 & **CES** 88-90.

[22] See also W. Rehg, *Insight and Solidarity: The Discourse Ethics of Jürgen Habermas* (Berkeley: University of California Press, 1994), p. 39.

[23] Similarly, Principle (D), claims Habermas, explicates the moral standpoint, i.e. "the point of view from which norms of action can be impartially grounded". (**JA** 12-13)

[24] Albrecht Wellmer makes this criticism in *Ethics and Dialogue in the Persistence of Modernity* (Cambridge: Polity Press, 1985), p.154. See also S. Benhabib, *Critique Norm and Utopia: A Study of the Foundations of Critical Theory* (New York: Columbia U.P., 1986); Maeve Cooke, in "Habermas and Consensus" *European Journal of Philosophy* 1, 3: 257-8; and Thomas McCarthy "Practical Discourse: On the Relation of Morality to Politics" in *Ideals and Illusions* (Cambridge: Cambridge University Press, 1991), p.198. Habermas is not unaware of this problem (**MCCA** 305) but does not seem to regard it as a pressing problem for moral philosophy.

[25] An interest can have both a particular object and a particular subject, such as my interest in being the one who loves my children. I have a particular interest in my children's (not in all children's) being loved. And I have an interest in being the one who loves them.

https://doi.org/10.1017/S0263523200001269 Published online by Cambridge University Press