What are 'Universalizable Interests'?[1]

Many of Habermas's critical commentators agree that Discourse Ethics fails as a theory of the validity of moral norms and only succeeds as a theory of the democratic legitimacy of socio-political norms.[2] The reason they give is that the moral principle (U) is too restrictive to count as a necessary condition of the validity of norms. Other more sympathetic commentators want to abandon principle (U) and remodel Discourse Ethics without it.[3] Still others, try to downplay the role of universalizing moral discourse and to make more of Habermas's less demanding, though still somewhat vague, conception of ethical discourse.[4] Against this chorus of critical voices Habermas maintains that his conception of moral discourse and the moral principle (U) are central to Discourse Ethics in general, and to the normative heart of his political theory in particular.

This conflict may have arisen in part because of the obscurity surrounding the central concept of a 'universalizable interest'. Actually Habermas's concept of interest is pretty obscure too. But the obscurity surrounding the concept of interest is not the issue here. For our present purposes we can simply stipulate that an interest is a reason to want.[5] The obscurity that is the problem here arises from ambiguities in the notion of universalizability that is in play. Once we pay due attention to the conditions of the universalizability of interests contained in Habermas's formulation of the moral principle (U), we can distinguish between a weaker and a stronger version of it. I argue that only the weaker version is

defensible. But I also want to show why Habermas is tempted into defending the stronger version.

2. The Meaning and Function of Principle (U)

A recent formulation of (U) states that:

> a norm is valid if and only if the foreseeable consequences and side effects of its general observance for the interests and value-orientations of *each individual* could be freely accepted *jointly* by *all* concerned. (**OCCM** p. 354/**DEA** p. 60)[6]

The most recent formulations of (U), unlike the earlier ones, make clear that the amenability to a consensus of interests is a sufficient as well as a necessary condition of a norm's validity.[7] The trouble is that, thus formulated, (U) is fraught with ambiguity. (U) could be taken to mean that in discourse validity is conferred on a norm if and only if everyone can 'freely and jointly' accept it on the basis of *an* interest in its general observance, though not necessarily the same one. Call this *unofficial* version of (U), $\mathbf{(U)_1}$. Alternatively, however, (U) could mean that validity is conferred on a norm if and only if there is 'free joint acceptance' of it on the basis of *one and the same interest* in its general observance.

This ambiguity is contained in all the various formulations of principle (U) (**MCCA** 65 & 197: **OCCM** 354: **BFN** 108: **DEA** 60). It also infects many of Habermas's descriptions of valid moral norms as 'equally good for all' or 'equally in everyone's interest', as embodying 'universally shared', 'common', 'general' or 'universalizable' interests, and in his claim that moral norms can be

'willed from the perspective of everyone'. It is a systematic ambiguity, not a slip.

It is easy to see why Habermas must reject the unofficial version. We must bear in mind that Habermas thinks that interests provide participants in discourse with reasons to assent to norms. Now, consider the case of a culturally mixed and economically self-sufficient community in which everyone agrees that one should not eat pork, but for different reasons. Some members of a this community do not eat pork because they believe God has proscribed it, the others because they are vegetarians. For argument's sake let us assume that the believers are not vegetarian and the vegetarians are not believers. So everyone in this community can assent to the 'norm' that one should not eat pork, since its general observance satisfies their different, albeit compatible, interests. According to the unofficial version of (U) this norm would be valid. On the official version it would not. Since the two parties have different interests, hence different reasons, to assent to the norm that prohibits eating pork, their consensus is wholly serendipitous; it is not 'rationally motivated'.

The example is contrived, but it shows why Habermas is wedded to the official version. Discourse conforming to (U) is supposed to aim at establishing 'rationally motivated' consensus, not compromise or *de facto* agreement. Compromise and merely *de facto* agreement may rest on a mere overlap of different interests and thus on different reasons. And a discursive consensus is rationally motivated only if everyone can accept the same norm 'for the same reasons' [*aus denselben Gründen*] which means on the basis *of the same* interest. (**SE** 78-82; **DEA** 108).[8] It lies at the heart of Habermas's disagreement

with the later Rawls that the ideal prosecution of a moral discourse ensures that this condition is met.[9]

> Whereas parties brokering a compromise can assent to the result each for different reasons, participants in argumentation aim to secure a rationally motivated consensus, if at all, then on the basis of the same reasons. (**DEA** 108)

This requirement effectively rules out the *unofficial* version, $\mathbf{(U)_1}$. Unfortunately though, it only partly clears up the ambiguity in (U).

3. Distributive and Collective Universalizability

Officially then, (U) requires that everyone be able to accept a norm for the same reason, on the basis of the same interest. But this just raises another problem. What is to count as everyone's having the 'same' interest in a norm and as accepting it 'in the same way' or assenting to it 'for the same reason'? This second problem turns on the question of how interests are individuated.

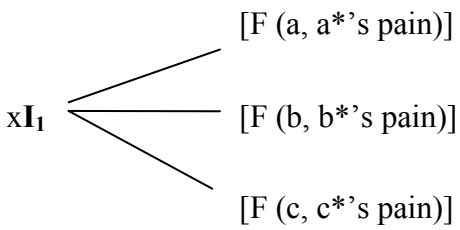
Take the norm $\mathbf{N_1}$: 'do not inflict unnecessary pain on others'. This is the kind of norm that the procedure of discourse conforming to (U) should validate. If not, something must be awry with the procedure. For the intuition that everyone has an interest in avoiding pain is very deeply rooted.[10] But can everyone be said to have the same interest in $\mathbf{N_1}$, and, if so, in what sense? Here we need to bring in the distinction between two kinds of universalizability - distributive and collective.[11]

An interest is *distributively* universalizable if and only if each person can agree that they have their own such interest. It helps to represent this with a

more formal notation. In compliance with (U) the scope of quantification is restricted to the domain of all persons potentially concerned by the general implementation of the norm, $\mathbf{N_1}$. Let 'xI' abbreviate the relation 'x has an interest in', and 'F' stand for 'avoids'. The interest, **I,** is *distributively* universalizable if and only if:

$\forall x$ (xI [F (x, x*'s pain)]) Call this $\mathbf{I_1}$.

It is a formal feature of all distributively universalizable interests that in each case the object or aim of the interest - that F (x, x*'s pain) - refers pronominally back to its subject, to whomever it belongs, x. This makes all distributively universalizable interests agent-relative.[12] Further, each universally distributed interest has a numerically distinct content in virtue of the numerical difference between the interest holders.



Practically speaking, the differences are important. Each person's agent-relative interest/reason provides that person with a different aim.[13] Morally speaking, though, each different interest counts equally. That is why the natural temptation here to say that everyone has the 'same' interest is correct. Under the description, $\mathbf{I_1}$, they do.

Of course there is cultural and historical variation in how experience of the need to avoid pain is interpreted. The Spartan warrior interpreted his need to avoid pain very differently from a late-twentieth Century academic.[14] And

Allison who chose to have a natural childbirth had a very different interest in avoiding her labor pains, than I do in avoiding my toothache. The threshold of pain to be borne or avoided depends upon whose pain and what kind of pain it is. Nonetheless it is uncontroversial that for all human beings in all situations, even those in which pain is voluntarily endured, there are degrees of pain and kinds of pain it is reasonable to want to avoid. The description, $\mathbf{I_1}$, captures formally only this very general structure, common to all interpretations of the need to avoid pain.

By contrast an interest is *collectively* universalizable if and only if 'everyone' can agree that they hold a single interest in common. But what kind of interests can be held in common by everyone? I take it that everyone's interest in there not being a global environmental catastrophe or in the planet's not careering out of its orbit are examples of *collectively* universalizable interests. But such global interests as these are pretty remote from our moral lives, and not at all the kind of thing we would be likely to appeal to in discourse, even implicitly, in order to resolve moral conflicts.

Fortunately not all collectively universalizable interests are so remote. There are collectively universalizable interests in basic, irreducibly social goods, i.e. goods which everyone wants and which can only be pursued and enjoyed in concert with others. My interest in freedom of expression, to use an example due to Joseph Raz, is not just an interest in my freedom to express myself but one which extends to other people's freedom of expression too. [15] The satisfaction of my interest depends on the existence of a common liberal culture which arises through the free exchange of information which occurs only where

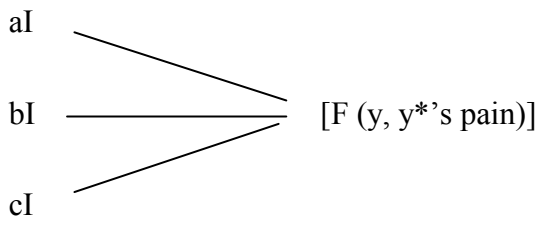
other people's interest in the freedom of expression is protected. To take a different example, my interest in my having a clean environment is also an interest in everyone's having a clean environment, insofar as the ecological balance is ultimately a global not a local phenomenon.

Collectively universalizable interests are characterized by universality on two levels; their universal *content* or *aim* as well as their ubiquitous distribution. Everyone has an interest in *everyone's* freedom of expression or clean environment. Moreover, they exist on a higher plane of abstraction than distributively universalizable interests. We can represent this formally by adding another variable and deleting the reference to the agent from the content of the interest, within the inner brackets. To revert to our first example, there is a *collectively universalizable* interest in pain avoidance, if and only if:

$\forall x \, \forall y$ (xI [F (y, y*'s pain)]) Call this $\mathbf{I_2}$.

According to $\mathbf{I_2}$ there is a collective interest in avoiding pain if and only if everybody has an interest in everybody's avoiding their own pain.

There are two things to note about $\mathbf{I_2}$. Firstly, the aim of the interest contains no pronominal back reference to the interest holder, hence the structure of the interest is agent-neutral and not agent-relative. Secondly, with $\mathbf{I_2}$, every person's interest in $\mathbf{N_1}$ has exactly the same content, namely that F (y, y*'s pain) and each has exactly the same collective aim.



The avoidance of pain is not a good candidate for a collectively universalizable interest. Yet the free expression example and the clean environment example which are *prima facie* more promising invite the following objection. It might be thought that the good provided by a liberal culture of free expression is only collective in the following sense: the satisfaction of one person's interest in free expression is conditional upon the more widespread satisfaction of that interest. The practice of free expression or environmental responsibility must be widespread enough that a common good emerges which is available to individuals qua individuals. Nevertheless it is not true that my having interest in my enjoying freedom of expression (or an unpolluted environment) entails that I have interest in *everyone else's* interest in these things. My interest in other people's interests may only be instrumental and selfish and, therefore, these other people need not be *everyone*, only numerous enough to produce the requisite social good.

To avoid this objection we must only note that the suppression of some people's or even one person's interest in free expression is deleterious to the common good of a liberal culture. Similarly, if we assume that the environment is ultimately a global not a local ecological system, we can say that everyone's environment is damaged in some way whenever anyone pollutes it. Assuming further that everyone aims at a good in the highest degree, or in its most complete form, then we are right to say that each person's interest in free

expression or in a clean environment for herself implies an interest in free expression or a clean environment for everyone. These interests are therefore collectively and not distributively universalizable in the sense outlined above.

If this is correct, interest **I** can justify norm $\mathbf{N_1}$ under two different descriptions. For both $\mathbf{I_1}$ and $\mathbf{I_2}$ fulfill the requirement that everybody be able to assent to a norm for the same reason, on the basis of the same interest. In turn, this opens up two further possible interpretations of the official version of (U): $\mathbf{(U)_2}$, a norm is valid if and only if it embodies an interest which is either distributively universalizable or collectively universalizable; and $\mathbf{(U)_3}$, a norm is valid if and only if it embodies a collectively universalizable interest.

4. Universalizability and Agent-neutrality in Discourse Ethics

So far as I can see, Habermas hedges his bets with regard to $\mathbf{(U)_2}$ and $\mathbf{(U)_3}$. In his response to the objection that (U) sets an implausibly restrictive necessary condition on the justifiability of norms, Habermas endorses $\mathbf{(U)_2}$. But when explaining how the principle captures the cognitive content of moral normativity, in contrast to the contractualism of Hobbes or Rawls, he adduces $\mathbf{(U)_3}$.

For example, in his response to Steven Lukes Habermas writes: 'I do not understand why he (Lukes) thinks this requirement is too strong.' He goes on to claim that there are many examples of norms embodying universalizable interests 'from traffic rules to basic institutional norms'. (**HCD** 257) Insofar as we can say that traffic regulations rest on a single universalizable interest, then it seems to me to be distributively and not collectively universalizable.[16] For, as a

vehicle user, my interest in the traffic's being able to flow depends on my interest in *my* being able to travel freely. Habermas seems to be allowing that distributively universalizable interests are sufficient to justify norms, as **(U)$_2$** has it.[17]

By contrast, in his work since 1988, Habermas has tended to endorse **(U)$_3$** and to reject **(U)$_2$**. In order to show that my argument here does not rest on any lack of interpretative charity, and at the risk of labouring the point, I shall adduce four different examples.

1. Habermas insists that there is a conceptual link between justice and universal solidarity. In *Justification and Application* he writes that 'universal solidarity' or 'solidarity with everything with a human face' is the other side of justice understood as the principle of equal respect for all.[18] The relation of solidarity usually means taking an interest in other people's interests; it is a collective rather than a distributive sharing of interests. Universal solidarity refers to the relation that obtains when each person takes interest in the interests of all others. This may imply an intrinsic and direct interest arising out of sympathy, rather than the kind of interweaving of the concern for oneself and the concern for others that we saw in the free expression and the clean environment examples. Nevertheless in each case the interest in question is directed towards the interests of all others and is, in the sense outlined above, collectively universalizable.

2. Habermas distinguishes the discourse theory of morality as a form of 'moral cognitivism' from certain forms of contractualism and from Rawls's 'political' conception of justice.[19] The latter, he argues, fail to capture the

'cognitive content' of moral norms. Part of Habermas's argument is his semantic claim that norms that are justified by agent-neutral and not merely agent-relative reasons, capture the *cognitive* meaning of moral utterances. Roughly what he means is that moral utterances have cognitive meaning because they are analogous to assertions in certain basic syntactic, semantic and pragmatic respects: they are syntactically disciplined; they make a claim to property (truth or rightness) which is absolute and stable, and they connect in the appropriate way with consensus in discourse. Thus he claims that,

> Normative reasons are - unlike mere declarations of intent or simple imperatives - not agent-relative reasons for one's own...instrumental behavior, but - as in the case of assertions - agent-neutral reasons. (**SE** 78)[20]

It is evident that Habermas must be endorsing **(U)$_3$**. On standard accounts agent-relative reasons include an essential pronominal reference back to the agent; agent-neutral reasons do not.[21] We have already seen that **I$_1$** is agent-relative, because its content, F (x, x's pain), refers back to the interest holder, and that **I$_2$**, by contrast, is agent-neutral because it does not. In the passage cited, Habermas's claim is that not just the moral norms, but also the reasons that justify those norms, must be agent-neutral.[22] The implication is that agent-neutral principles that are only justified by relative reasons are not yet moral principles. The further step to morality requires that valid norms be amenable to agent-neutral justification. For only agent-neutral reasons are 'epistemic' in the sense that they are relevantly analogous with the truth-seeking reasons justifying

assertions. (**DEA** 15: **SED** 78) And, as we have seen, only collectively universalizable interests furnish agent-neutral reasons.

3. In the context of a critical discussion of Rawls's overlapping consensus Habermas writes:

> It is counterintuitive for the moral authority of a public conception of justice to rest on reasons that are not public. Everything valid must also be publicly justifiable. Valid utterances deserve to be universally recognized on the basis of the same reasons...Such a practice of justification [i.e. moral discourse not compromise G.F.] aims at a *publicly* and *collectively* achieved consensus. (**DEA** 108: my emphasis)

Habermas uses the terms 'justice' and 'morality' interchangeably. If we replace the term 'public' with 'impartial' or 'agent-neutral', which in no way alters the sense of the passage, and if we bear in mind that collectively but not distributively universalizable interests imply agent-neutral reasons, it becomes clear again that he endorses $\mathbf{(U)_3}$.

4. Finally, Habermas often presents moral discourse as a process in which all 'agent-relative' reasons, along with all ethical and evaluative considerations that presuppose the particular life-histories, projects and self-understanding of individuals, are 'uprooted' and left behind. (**MCCA** 161) Echoing Marx's famous metaphor, he claims that agent-relative reasons are part of the historical and cultural 'shell' which must be stripped away in discourse to reveal the 'rational kernel' of universalist morality. (**ED** 40). But that means that the

rational kernel of morality contains no distributively universalizable interests, and that Habermas is endorsing **(U)$_3$**, and rejecting **(U)$_2$**.

5. Objections to (U)$_3$

If I am correct, in the course of the last two decades Habermas has exploited the ambiguity in the official version of (U) by sliding between two different positions. To begin with he explicitly endorses **(U)$_2$**, but later he rejects this and commits himself to **(U)$_3$**. In my view this is a mistake. For **(U)$_3$** makes Habermas's Discourse Ethics vulnerable to the standard objection to (U). The standard objection is that (U) imposes an implausibly restrictive necessary condition of the validity of moral norms. The most cogent version of this objection is the redundancy argument. According to the redundancy argument (U) sets such strict conditions of universalizability that it leaves few survivors. Even allowing that those few norms that survive (U) capture our deepest intuitions about what counts as a valid moral norm, it still follows that moral discourse can at most play a peripheral role in our moral lives.[23] But that is tantamount to conceding that moral discourse is not up to the social and pragmatic tasks of resolving conflicts of interest and orienting interaction in the life-world that discourse ethics assigns to it. This raises two further questions. Why do we life-world inhabitants persist in making moral discourse a central part of our lives if it produces such meager results for social cooperation? And why does Habermas's political and legal theory privilege moral discourse, rather than the mechanisms to which we resort in order to regulate the many conflicts of interest, which cannot be resolved by moral discourse?

Although the secondary literature focuses almost exclusively on the redundancy argument and on Habermas's various responses to it, $\mathbf{(U)_3}$ is open to a different set of arguments.[24] Firstly, the very strong claim that adequate justifications of moral norms must exclude agent-relative reasons is certainly false. $\mathbf{I_1}$ must enter somewhere into the justification of $\mathbf{N_1}$. Each person's interest in avoiding pain to them must play apart in the justification of the norm, 'do not cause unnecessary pain to others'. For even to recognize that other people's pain deserve equal consideration to my own, I need to be able to understand and to identify with their interest in avoiding their pain, from their point of view. Achieving that level of empathy means understanding that their pain looms larger in their life than my pain does. I come to that understanding partly by observing their behavior and partly by projecting from my own relation to my pain. So, unless I keep my interest in avoiding my pain firmly in view, I cannot achieve the insight into other people's pain that moral discourse requires.

Of course, Habermas can just modify the claim that moral discourse excludes all agent-relative interests, to the weaker claim that it transforms or eliminates all *particular* agent-relative interests. But what about the weaker claim that only agent-neutral reasons are sufficient to justify moral norms, the claim that $\mathbf{I_2}$ is necessary to justify $\mathbf{N_1}$? I think it is quite plausible that agent-relative interests might sometimes be sufficient to justify norms and that $\mathbf{I_2}$ is not necessary to justify $\mathbf{N_1}$.

Imagine a moral community in which there are no universal collective goods and no collectively universalizable interests. In this community agent-relative reasons furnished by distributively universalizable interests justify

norms. In such a community my interest in avoiding pain to me, conjoined with my recognition of everyone else's interest in avoiding pain to them, is, when universalized, sufficient to justify $\mathbf{N_1}$. On this view everyone has the same reason, indeed everyone has equal reason to agree to a norm, but this reason is not impartial or neutral in Habermas's sense. Such a community need not be without first-order agent-neutral prescriptions or norms, such as 'do not steal', 'do not kill', or, to us our previous example, 'do not cause unnecessary pain to others'. Although each person has no interest in other people avoiding pain to them, they still have prudential reasons to recognize that everyone else has an interest in avoiding their own pain. Everyone's distributively universalizable interests deserve recognition in this sense. But they also deserve recognition just because they are distributively universalizable, and are thus especially good at coordinating interaction. Our imagined community could adopt Habermas's favored example of a valid moral norm, namely, the universal human right to life.[25] It could also have second order agent-neutral norms, for example that everyone should obey the valid first order norms. Yet none of these norms, even if they are themselves agent-neutral, are justified by agent-neutral reasons resting on collectively universalizable interests. Rather, they are justified by the relevant distributively universalizable interest, plus the general awareness that everyone indeed has that interest. Perhaps a more impartial outlook would be admirable. But it is not obviously necessary. All that is necessary is that I recognize that my interest in avoiding my pain counts no more and no less in favor of the adoption of a norm than every other person's interest.

This imagined community is sufficiently like our own to undermine the intuitive basis of the claim that only collectively universalizable interests are sufficient to justify norms. It is true that morality is more than a matter of enlightened self-interest, indeed more than enlightened 'agent-relative interest' which need not be self-concerned. We can even grant, although it is debatable, that all moral norms are agent-neutral. Certainly very many are. We can accept all this and still deny Habermas's claim that moral norms can only be justified by agent-neutral reasons, and that agent-relative reasons are not sufficient to justify moral norms. Intuition, then, does not support the view that norms must be amenable to a consensus on the basis of agent-neutral reasons furnished by the collectively universalizable interests of participants in discourse.

Not only does Habermas not have intuition on his side here, he has good reason not to hold that moral norms must be agent-neutral all the way down to the reason that justify them. For this is uncomfortably close to a doctrine of Kant's that Habermas rejects, namely that moral actions must be justified by *pure* practical reasons, that moral maxims must be adopted solely on the basis of a priori or 'purely rational' interests. (**MCCA** 197: **OCCM** 345) Habermas studiously avoids all talk of 'pure' or *a priori* interests which, he claims, is saddled with the baggage of Kant's two worlds metaphysics. Yet he himself separates impartial *moral* reasons rigidly from agent-relative reasons. He maintains that it is epistemologically possible to distinguish moral norms strictly from all other evaluative considerations on the grounds that the former are agent-neutral all the way down.

Secondly, by banishing agent-relative reasons from the sphere of moral discourse, Habermas loses the advantages offered by interest-based accounts of practical reason, like Hegel's. One advantage of an interest-based account is that by exploring the close connection between moral reasons and the lived experience of real individuals, it can offer a very plausible account of moral motivation. Discourse Ethics as conceived by Habermas, namely as a program of the justification of the moral principle or the moral standpoint (**MCCA** 43, 78-86, 96: **OCCM** 347) does not include an account of moral motivation. However, it is supposed to be consistent with an empirical moral psychology. The more Habermas insists on the radical impartiality of moral justification, the harder it will be to hook up Discourse Ethics with a plausible moral psychology.

Thirdly and finally, by thus committing himself to $\mathbf{(U)_3}$ Habermas fails to do justice to what is an essential feature of interests. Let me illustrate this point with an example of tryingly clever conversation from *Ally McBeal*.

Ally, what makes your problems bigger than everyone else's?

They're mine!

Ally's problems, like most people's, are due to the frustration of her interests. Notice she does not say that her problems (interests) are bigger (or seem bigger to her) than everyone else's, because they are about her, but because they *belong* to her. These are different and independent claims. For not all of her problems are about her; not all of her interests selfish. What makes her interests bigger than everyone else's is just this, that they are hers. It is a question of perspective.

So far I have not distinguished this perspectival feature of interests from their 'agent-relativity'. But it is not just agents who have interests. Strictly speaking the focus of Discourse Ethics is on morally justifying reasons, rather than on reasons to act - practical reasons. It may be more appropriate, then, to talk about the 'relationality' of interests, and leave the terms of the relation open. But 'relationality' does not quite capture the first-person nature of interests. For the want of a better English term, Heidegger's '*Jemeinigeit*' most aptly captures the feature of interests I am talking about - the fact that interests are in each case mine.[26] Whatever an interest is an interest in, it belongs to someone. Even a collective or group interest belongs to the group only by virtue of belonging to the individuals who comprise the group. To my way of thinking any interest-based account of moral reasons worth its salt has to do justice to the irreducible *Jemeinigkeit* of interests. Discourse Ethics, in its present form, does not. Moral discourse conforming to (U) is a procedure which begins from the perspective of each participant, but in the process 'the reasons adduced lose the actor-relative meaning of practical motives and assume an epistemic meaning under the aspect of symmetrical consideration. (**OCCM** 355: **DEA** 60)

But then it seems that the 'epistemic' basis of the moral requirement that I give equal weight to the interests of all, friends and strangers alike, cuts moral discourse adrift from a fundamental, orientating and perspective-giving feature of interests, namely, their being in-each-case-mine.[27] This sits ill with the aspiration of discourse theory to provide an intersubjective or interpersonal rather than an objectivist model of the validity of moral norms.

If I am right, all these difficulties can be avoided if Habermas adopts **(U)$_2$** and rejects **(U)$_3$**. So the question arises, 'why does Habermas adopt the stronger and less plausible position?' The only explanations I can think of are the following.

1. Habermas may mistakenly assume that the distinction between universalizable and non-universalizability (particular) interests lies parallel with that between agent-neutral and agent-relative reasons. In fact, the former distinction is orthogonal to the latter: some agent-relative interests/reasons can be universalized.[28] Once Habermas allows the two distinctions to cross, he can hold that (participants in moral discourse must assume that) everyone can assent to a norm for the same universalizable reason, without insisting that this reason be agent-neutral into the bargain.

2. Alternatively Habermas allows that universalizability does not always reliably correlate with agent-neutrality. However, he assumes that discourse meta-ethics requires that moral reasons be impartial, i.e. both universalizable and agent-neutral. In this case the culprit is the supposed analogy between 'truth and moral rightness', or, as he also puts it, between epistemic and moral reasons. (**SE** 76-82: **OCCM** 344) If it is the analogy that requires **(U)$_3$**, then Habermas presumably has grounds for assuming, as the fixed end of the analogy, that all epistemic reasons are impartial. His assumption may be that beliefs and assertions are justified by the facts, not by beliefs about facts. For example, the reason that justifies my belief that it is dark outside is given by the fact that it is dark outside, and can thus be fully articulated without essential reference to me, the holder of the belief. This is a fair enough line to take. Even so, it does not

warrant the conclusion that all epistemic reasons are impartial, and hence that there are no thinker-relative epistemic reasons. My reason for believing that I have a headache, or that your joke is funny, or that this letter is addressed to me is not agent-neutral. For the facts that justify them contain an ineliminable reference to me; the fact that my head aches, the fact that I am amused, and the facts that the letter is addressed to GF and I am GF.[29] These counterexamples suggest that not all epistemic reasons are impartial, or, at very least, that this cannot be assumed without argument. Hence the analogy between epistemic and moral reasons cannot on its own drive the conclusion that all moral reasons must be impartial. No doubt there are similarities between epistemic and moral reasons, one of which may be that very many epistemic reasons and many moral reasons are impartial. But this is not sufficient grounds for rejecting **(U)$_2$** and accepting **(U)$_3$**.

6. Conclusion

The reasons why Habermas should reject **(U)$_3$** and endorse **(U)$_2$** instead are overwhelming. Phenomenologically speaking it would give him a richer conception of what an interest is, and permit him a more plausible and robust account of moral motivation. He would gain a broader and less revisionist account of moral reasons. This would itself be a good thing. For universalizable agent-relative reasons are some of the deepest-rooted and most powerful reasons we have, and no universalist, deontological moral theory worth its salt can afford to ignore them. It would also men that (U) would be less restrictive and not so vulnerable to the standard objection.

Of course the revision I am proposing will require adjustments elsewhere in the theory. Habermas will have to redraw the 'razor-sharp' distinction that (U) is supposed to make between the moral and the ethical. This is just as well. For Habermas's thesis that ethical values and moral norms divide neatly up along the lines of the agent-relativity/agent-neutrality distinction, seems to rest on wishful architectonic considerations rather than on sound moral phenomenology. Indeed, in the face of much criticism, Habermas has begun to concede that relation between the good and the right, between ethics and morality may be much closer and much messier than he initially supposed. (**ED** 44: **OCCM** 343)[30] It which case there is all the more reason to investigate the role of universalizable agent-relative reasons in moral discourse. This would open up a vista onto the complex dialectic that exists between agent-relative the agent-neutral reasons, which we more readily recognize as moral.

Furthermore, the revision I am proposing is minor and involves low costs to the theory of Discourse Ethics. This marks it out from almost all current criticisms from the direction of communitarianism or the 'ethics of care', and which are, in my view, much too ready to give up on the central deontological and universalist aspirations of Discourse Ethics.[31] If such criticism is well-aimed, it is not easy to see how it can be accommodated without drastic revisions to the theory. The alternative I have offered is more appealing. I have shown how Discourse Ethics can be made plausible with the help of a more nuanced account of the conditions of the universalizability of interests, and still live up to its original aim of providing a genuinely intersubjective, deontological and universalist moral theory.

[1]Abbreviations of Habermas's works: **DEA** = *Die Einbeziehung des Anderen* (Frankfurt a/M: Suhrkamp, 1996): **ED =** *Erläuterungen zur Diskursethik* , (Frankfurt a/M: Suhrkamp, 1991) **HCD** = *Habermas: Critical Debates*, J. Thompson & D. Held eds.(London: MacMillan, 1982): **JA =** *Justification and Application*, (Cambridge: Polity 1993); **MCCA** = *Morality and Communicative Consciousness*, (Cambridge: Polity Press, 1990): **OCCM** = 'On the Cognitive Content of Morality', *Proceedings of the Aristotelian Society,* 1997: **SE** = 'Sprechakttheoretischer Erläuterungen zum Begriff der kommunikativen Rationalität', in *Zeitschrift für Philosophische Forschung* 50 (1996), pp.65-91:

[2] See Albrecht Wellmer, 'Ethics and Dialogue', in *The Persistence of Modernity*, (Cambridge: Polity Press, 1991); Agnes Heller, 'The Discourse Ethics of Habermas: Critique and Appraisal', *Thesis Eleven* 10/11, (1984-5), pp.5-17; and Simone Chambers, *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse* (Ithaca; New York: Cornell University Press, 1996), p.145.

[3] Seyla Benhabib, *Critique, Norm and Utopia* (New York: Columbia University Press, 1986), ch. 8 and 'Liberal Dialogue Versus a critical theory of discursive Legitimation' in Nancy Rosenblum ed., *Liberalism and the Moral Life* (Cambridge: Harvard University Press, 1981) pp. 143-156.

[4] See Thomas McCarthy 'Legitimacy and Diversity Dialectical Reflections on Analytical Distinctions', *Cardozo Law Review* 17/4-5 (1996), pp.1083-1127.

[5] The notion of reason rests loosely on the notion of a need, and the concepts of need and desire are take left deliberately vague.

[6] The English translation fails to render the German 'genau dann...wenn' as 'if and only if', and makes (U) look like it contains only a necessary condition of the validity of norms. (**DEA** 60: **OCCM** 354)

[7] E.g. the earlier formulations at (**MCCA** 197: **ED** 12 & **FG** 139, but not **BFN** 109!).

[8] For this reason I think that Rehg's example of $Norm_{xy}$ regulating a conflict 'you want x and I want y' is not one Habermas could be happy with. For this norm might not be acceptable to both parties for the same reasons. W. Rehg, *Insight and Solidarity*, (University of California Press, 1994) p. 72.

[9] J. Habermas, 'Reconciliation through the Public Use of Reason', *Journal of Philosophy*, 92/3, 101-131 & (**DEA** 113-5)

[10] Sometimes, as when we go to the dentist, it is necessary to undergo severe pain. But I take it that this is usually for the sake of pain-free future. I am not claiming that this interest overrides all others, just that all of us have it.

[11] Philip Pettit makes a procedural distinction between distributive and collective consensus in 'Habermas on Truth and Justice', in *Marx and Marxisms* G.H.R.Parkinson ed. Royal Institute of Philosophy Lecture Series 14, (Cambridge: Cambridge University Press, 1982) p.215.

[12] See McNaughton and Rawling: 'Value and Agent-Relative Reasons' *Utilitas* vii, (1995) p. 33. 'A reason is agent-relative if its full articulation would involve ineliminable pronominal back-reference to the agent: agent-neutral otherwise'.

[13] Derek Parfit, *Reasons and Persons*, (Oxford: Oxford University Press, 1984) p. 27.

[14] See Life of Lykurgus, ch XVII. *Plutarch's Lives*, (London: G. Bell and Sons, 1925) p. 85.

[15] The example comes from Joseph Raz, *Ethics in the Public Domain*, (Oxford: Clarendon Press, 1994): p. 54.

[16] Just as plausibly traffic regulations could rest on a constellation of different interests - i.e. those of private vehicle users, pedestrians, road-hauliers and residents.

[17] Pettit claims that whatever Habermas means by 'generalizable interests' they surely include 'universal self-regarding desires which each can fulfill compatibly with respecting similar desires in others.' 'Habermas on Truth and Justice', in *Marx and Marxisms* ed. Parkinson (Cambridge University Press, 1982) p.225.

[18] (**JS** 15 translation amended**,** 68 & 134, **MCCA** xi-xii.).

[19] See especially (**DEA** chs. 1, 2 & 3). When Habermas says that moral utterances are cognitive, he means not that they are truth-apt, but that they are amenable to justification.

[20] Habermas writes that *moral* justification is only possible on the basis of rational reasons [*Vernunftgründe*]: 'But in contradistinction to the empirical varieties of contractualism, these rational reasons are no longer understood as agent-relative motives, with the result that the epistemic core of the validity of ought-claims [*Sollgeltung*] remains in tact.' (**DEA** 15)

[21] McNaughton and Rawling: 'Value and Agent-Relative Reasons' *Utilitas* vii, (1995) p. 33. See also Thomas Nagel, *View from Nowhere*,(Oxford: Oxford University Press, 1986) p.153.

[22] Note that the claim is not that agent-neutral principles are only justifiable by agent-neutral reasons, which is certainly untrue. Even a community of rational egoists could agree to a system of morality which contained some agent-neutral principles. Habermas's claim is that *moral* principles cannot be justified by agent-relative reasons.

[23] Thomas McCarthy 'Practical Discourse: On the Relation of Morality to Politics' in *Ideals and Illusions,* (Cambridge Mass.: Cambridge University Press, 1991) p.198; and Maeve Cooke, 'Habermas and Consensus' *European Journal of Philosophy* 1:3 pp. 257-8; & *Language and Reason: A Study of Habermas's Pragmatics*, (Cambridge Mass.: MIT Press, 1994) p. 153-4.

[24] McCarthy's important articles (see notes 3. & 24 above) have set the agenda here, not just for commentators on Discourse Ethics but, to an extent, for Habermas too.

[25] (**JA** 90-91: **MCCA** 205: **HCD** 257) Habermas does not specifically adduce the universal right to life, but he must have it in mind.

[26] The fact that *Dasein* is in each case my own is central to Heidegger's discussion of *Eigentlichkeit,* authenticity, or literally 'ownness' in *Being and Time*. *Authenticity* which has become a *faux* contintental word for self-realization gets a lot of coverage in contemporary debates in social, and political philosophy. *Jemeinigkeit* gets virtually none. *Being and Time* §9, tr. J. Macquarrie and E. Robinson, (Oxford: Blackwell, 1987): 67-70.

[27] In a way *Jemeinigkeit* is more appropriate than the term 'agent-relativity' which is a category of practical philosophy. I take agent-relativity to be a practical instantiation of the former, more general 'property'.

[28] Furthermore, arguably, some agent-neutral interests/reasons cannot. 'Agent-neutral Reasons: Are They for Everyone' B.C.Postow, *Utilitas* 9/2 July 1997 pp. 249-258.

[29] Of course some people may claim that there is no fact of the matter about my experience of pain or the amusingness of jokes. Such an objection does not apply to the third example however.

[30]He now claims that the good is, in Hegelian fashion, 'preserved and canceled' [*aufgehoben*] in the right. However he thinks the good is only preserved formally. It persists in 'the form of an intersubjectively shared ethos in general and therefore the structure of membership in a community.' (**DEA** 45)

[31] For a synopsis of current criticisms of Discourse Ethics (in 1994) see W. Rehg *Insight and Solidarity*, (Berkeley: University of California Press, 1994): p.11. On the ethics of care see Stephen Darwall's treatment in *Philosophical Ethics*, (Boulder, Colorado: Westview Press, 1998), pp. 217-228.