HARE, SINGER AND GEWIRTH
ON UNIVERSALIZABILITY

BY W. GREGORY LYCAN

I wish to discuss and compare three recent attempts at providing a "supreme" principle of moral argument. The principles offered are: R. M. Hare's "necessary universalizability of moral judgments" (UP), M. G. Singer's "generalization principle" (GP), and Alan Gewirth's "Principle of Categorial Consistency" (PCC). Each of these is said to be "purely formal", to follow directly from the ineluctable logic of moral discourse, and hence to have a necessary normative content which cannot be sidestepped without self-contradiction by anyone seriously interested in moral reasoning. I shall set out what I take to be the most troublesome deficiencies in the formulations of Hare and Singer, and go on to argue that Gewirth's programme does not exhibit these same limitations. Then I shall raise some other questions about how Gewirth is to develop the PCC into a useful moral mechanism.

A preliminary doubt arises as to whether there need be, or even can be, any such thing as a "supreme" principle of morality. I imagine that many philosophers deny both these claims. The intuitionist, for example, although he offers the possession of a non-natural "rightness" property as the universal moral criterion, holds that one's perception of it in some class of acts does not admit of further rational justification. Relativists frequently claim that different principles are supreme with respect to different societies. Emotivists see moral judgments as merely the expression of attitudes, emotions, and wishes, having no connection with any logical kind of "supremacy" at all. Many existentialists hold that "authenticity" of attitude is a necessary and sufficient condition of the justifiability of an agent's behaviour, regardless of either the descriptive or the "rational" characteristics of that behaviour. I shall try to discuss how views of this type fare at the hands of the three theorists under investigation, each of whom would seemingly reject them.

The UP, the GP and the PCC (which three I shall hereafter take to be essentially equivalent, in that each affirms that what is right for A must be right for B, granted relevantly similar circumstances) derive their alleged indefeasibility from the premises that an individual moral judgment necessarily has a certain amount of descriptive content. Insofar as such a statement is descriptive, it is universalizable: "when we make a moral judgment about something, we make it because of the possession by it of certain non-moral properties", although these properties do not exhaust the meaning

of the moral term involved; i.e., it is always appropriate to award certain brute facts of the situation the status of reasons. This is all very well; but Hare, Singer and Gewirth differ in their attempts at providing their purely formal "supreme" principle with normative content. We must, after all, have some moral judgments to universalize.

Hare's offering can be seen immediately to be relevant and workable so far as it goes, but inadequate in scope; what we need besides "the logical framework provided by the word 'ought'"; he says, is nothing more than our inclinations (interests, wants). All that is required for moral agreement between two persons is that they should have a congruent view of the relevant facts, a sufficient amount of imagination, and roughly similar inclinations. The "logical properties of moral words" take care of the rest in a purely formal manner.

The facts are often, given sufficient patience, ascertainable. Imagination can be cultivated. But the matter of interests and inclinations is likely to be more problematic. Hare claims that the agent's imagination should enable him to empathize himself into his recipient's set of likes and dislikes; the moral argument should be generalized so that, for example, trumpet enthusiast A is prevented (by imagining himself in a Gedankenexperiment as having trumpet-hating recipient B's preferences) from prescribing that B should play the trumpet at him under such circumstances. Naturally, as Hare says, this sort of argument would not convict the truly fanatical Nazi of inconsistency, since he would himself be willing to be persecuted should it unexpectedly turn out to be the case that he was of Jewish origin. This man, Hare tells us, serves a quasi-aesthetic ideal which he reveres above the interests of others and above his own. Such ideals are themselves universalizable, and are therefore very hard to argue against. Hare claims that we shall find these fanatics to be few, once honest moral inquiry of the type described above has been fully carried out; but this is small comfort to the moral philosopher.

Hare relies heavily upon the moral agent's ability to "imagine" what and how strong his recipient's inclinations are. A virtue of his account (which Singer's does not share, as we shall see) is that it provides a way of ensuring "relevant similarity" of cases: the agent is supposed to imagine himself in circumstances identical to those of the recipient ("hypothetical cases will do as well as actual ones"). But the degree of "toleration" embodied in Hare's view is too great. Even if we grant the dubious contention that a sufficient amount of "imagination" could enable two people of very different sorts to empathize their way into each other's shoes, nothing guarantees the moral justifiability of the relevant inclinations themselves. Now we may allow Hare to neglect this problem in the hope that the UP can be used in some second-order way to justify inclinations; we may even want to claim that, once the UP has balanced everyone's inclinations off

2Ibid., pp. 97 3.
against each other in an impartial way, the question of "justifying" the
inclinations themselves just does not arise. I think Hare probably holds
something like the latter view. But he still cannot ascribe moral injustice
to the acts of the fanatical Nazi, who is prepared to subordinate any of his
own interests, as well as those of others, to his quasi-aesthetic ideal. Hare
logically cannot call this man "immoral"; he can only hope that any group
of such fanatics will remain an ineffectual minority.

Hare's paradigm of a "moral argument" is unquestionably useful as a
model to be appealed to by two or more people who genuinely desire to
balance their interests and ideals off against each other, and so attain a
workable arrangement aimed at promoting a maximal amount of fairness
and mutual satisfaction. We cannot, however, regard this kind of argument
as an a priori "supreme" moral authority; interests and inclinations will
not serve alone as the input to the UP, since they fail to capture some of
our most important pre-analytical notions.

Singer avoids this problem by invoking the "principle of consequences"
(PC), which states, "If the consequences of A's doing X would be undesir-
able, then A ought not to do X without a special justifying reason". Singer
calls this a "necessary presupposition or precondition of moral reasoning",
and claims that its denial is self-contradictory. It follows from the PC
that "if the consequences of everyone's doing X would be undesirable,
then not everyone has the right to do X". The GP may be construed as
requiring that "if not everyone has the right to do X, then no one has the
right to do X without a special justifying reason". These last two statements
tail a third principle, the "generalization argument" (GA), which actually
is the core of Singer's theory: "If the consequences of everyone's doing X
would be undesirable, then no one ought to do X without a special justifying
reason". I am not sure that the version of the GP cited here is in fact
equivalent to the original form; but I shall overlook this consideration in
the interests of evaluating the GA on its own merits. Several essential qualifi-
cations are added at this point, in order to prevent misuse of the GA in
"invertible" and "reiterable" cases and to establish what is to be regarded
as a "special justifying reason". I have no quarrel whatever with this
latter part of Singer's exposition. The difficulty which arises for him is that,
if the PC is to be regarded as logically true, it is also vacuous, and devoid
of normative content. If it is literally impossible, as Singer says, to deny
the PC without self-contradiction, then "undesirability of consequences"
becomes logically linked with moral wrong or with negative prescriptive
force, in such a way that any attempt to justify moral judgments or pre-
scriptions by appeal to consequences becomes circular unless an independent
criterion is given for evaluating these consequences. In short, if we regard
the PC as a logical truth, we are still entitled to ask "What kinds of con-
essquences are undesirable?" Singer is aware of this loophole, but passes
over it with surprising insouciance:

---

There are many clear cases in which there is just no question that something is undesirable. For there are many clear cases in which there is just no question that something is disastrous. An earthquake is a disaster; so is a tornado, a shipwreck, or a plane crash, and so would be a nuclear war; and this does not depend on anyone’s theory of value, or on anyone’s interests or point of view, except in exceptional circumstances. . . . I see no need for qualifications, therefore, about the fact that many of my judgments or assumptions of desirability are unsupported. They need no support.\(^4\)

One wants intuitively to side with Singer here, but it is clear that he has run right into the "fanatic" problem. Many people presumably believe that nuclear war would be a "disaster," but manage to subordinate this feeling to the service of one or more "ideals" ("Better dead than red!"). Is such a person immoral? I myself should in general say so, depending on the circumstances; but, as Hare suggests, it is very hard to make a moral impression on this kind of fanatic, having only the \(GP\) and the \(PC\) in one's arsenal.

Moreover, the evaluation of consequences is far more central to moral argument than Singer seems to admit, and disagreement over such matters lies at the heart of many, if not most moral disputes, especially those of wide and crucial concern. ("But if everyone refused to join the Army, no one would be left to fight the war."—"Great! Fine!!") It could plausibly be held—I think Singer would maintain this—that conflicts over the evaluation of consequences generally rest on open or latent factual disagreements about what consequences such-and-such a deed would actually have. But I suspect that such a dispute would more often than not degenerate, once the factual debris had been cleared away, if in fact it could be, into a conflict between an array of disastrous consequences on the one hand and the faithful serving of an ideal on the other. What is important is that the fanatic regards the serving of his ideal as a desirable consequence, or at least as a dummy value whose function it is to offset undesirable consequences. In this way he is enabled to claim that the consequences of his act have been desirable "on the whole", even though he admits that it is too bad about those six million Jews. So the \(GP\), \(PC\) combination will not suffice to rule out his fanatical behaviour.

Singer's system has yet a third difficulty: Hare has avoided the problem of finding criteria of "relevant similarity" by basing his moral-argument paradigm on the agent's ability to imagine himself in an identical case; but Singer's moral agent, in formulating the maxim of an act, always has to mention the act under a certain description. To do so successfully he needs criteria of relevant similarity. Some guidelines are provided by the essential generalizability of exceptions to the \(GA\) (as described on p. 83 of his book), requiring as it does that an argument for exemption not be reiterable. But this restriction would suffice only if the disputants concerned shared the same evaluative view of the projected consequences, since it is, for Singer, only another "undesirability" consideration that defines a class of exemptibles. So this way of handling relevance criteria seems to fall prey to the same criticisms as does the \(PC\).

\(^{4}\)Singer, op. cit., pp. 94-5.
Finally, it does not at first glance seem impossible, as Singer claims it is, that the GP and the PC should appear to conflict in a practical case. It is certainly reasonable to suppose that in some situation the consequences of treating two people in similar circumstances equally might be very undesirable indeed. But there is really no formal incompatibility here: any application of the PC which violates the GP may be validated simply by adjusting the closeness of the ‘similar circumstances’ required by the GP. (“The circumstances aren’t similar; he’s a Negro, a person whose presence in my restaurant would cause mass heart-failure.”) On the other hand, we could also resolve the dilemma by adjusting our idea of “undesirability” in the case in question. The weakness of Singer’s programme is not that the GP and the PC can ever conflict irresolvably (they cannot do so, since they are purely formal principles), but rather that he gives us no metacriteria by which to balance the “relevant similarity” criterion and the “undesirability” criterion off against each other in a particular problematic case. Even in the face of the GP and the PC, in short, moral dilemmas will still arise, and prove almost as difficult as they do now. The GP and the PC will be helpful in clarifying moral issues, but they are too indeterminate to serve as “supreme” moral principles.

So our problem is still that of finding content, an unappealing normative and/or descriptive input, for the principle of universalizability (which principle—GP or GR—I shall hereafter accept without reservation; I take it that a satisfactory choice of content should itself provide some criterion of relevant similarity, and so free the GP from the burden of building in its own). Hare’s and Singer’s respective appeals to inclinations and consequences as content let important indeterminacies into their systems; but these particular weaknesses are avoided by Gewirth in his formulation of the PCC. His content comes rather, he claims, from a second-order application of the GP to the “categorical rules of human action”: it is a logically necessary condition of some instance of behaviour’s being action that the behaviour be (1) voluntary, and (2) purposive. So “whenever an agent acts, he necessarily applies to himself these two categorical rules,” viz., that his behaviour be voluntary and purposive. It follows from the GP that, if agent A enters rightfully into a social “interaction” with recipient B (and hence applies the two categorical rules to himself in acting), then B must have the right to apply the same rules to himself, i.e., to enter the interaction voluntarily and purposively. (A and B are relevantly similar in that they both are “potential agents” in the interaction.) The PCC thus states, “Apply to your recipient the same categorical rules of action that you apply to yourself,” and it has as corollaries the requirements that (1) no agent may justifiably force another person to participate unwillingly in an interaction with him, and (2) no agent may “frustrate his recipient’s purposes.” (Gewirth later provides, as he clearly must, for qualifications of these two principles.)

I am inclined to accept the PCC as true. It does not share the evident inadequacies of Hare's or Singer's programme, since it requires no further evaluative considerations as content. And it does convict the Nazi of inconsistency: he is acting on the maxim, 'Only Aryans are relevantly similar to me', which he has unilaterally decided for himself. In doing so he is acting on the second-order maxim, 'I, as agent, have the right to decide my own relevance criteria'; but he denies this same right to his victims. Since they are potential agents as well, and so are relevantly similar to him in a second-order way, he contradicts himself in refusing to universalize his second-order maxim.

The PCC, moreover, catches an intuitive feeling we have about moral interactions: I refrain from punching Joe in the nose, not because of the undesirable consequences of everyone's punching someone in the nose, but because I have less right to punch Joe than Joe has to be allowed to escape unpunched. (Hare's system, too, reflects this personal quality fairly well.) So the PCC is prima facie satisfactory. But there are some puzzling logical moves in its development which bear investigation.

First, I fail to see how referring to the "categorial rules of action" as "descriptive a priori" rules counters the objection that a logically inviolable rule is no rule at all. Gewirth states that "no agent can fail to conform" to the categorial rules. If we regard this claim as true, it is analytic. To say that someone had "acted involuntarily", on Gewirth's view, would be to misuse the verb 'to act', at least in the strict sense. So "involuntary action" is logically impossible (it is not "something that we cannot do", because it is not a "something" at all). For purely linguistic reasons, then, we have no concept of what would constitute a violation of the voluntariness "rule"; but if we have no concept of violating here, we have no concept of following or obeying either, and hence no real "rule" at all. What the PCC seems to be saying is simply that no one has the right to initiate an interaction unless his recipient has the right to participate in it voluntarily, i.e., the right not to be coerced. In other words, one should not force another person to do something; one should treat him as an "end" rather than as a means. This contention is venerable and true (in that its acceptance seems to be, as Singer says, a pre-condition of the possibility of using moral language workably and sensibly at all), and Gewirth makes a good point in tracing its logical origins by exhibiting its formal connection with the second-order judgment that no one has the right to make arbitrary unilateral decisions about relevance criteria vis-à-vis other human beings. This latter claim is indeed a consequence of the OP, but none of it follows from the definition of 'action'; nothing follows from such rules of language save other statements about language.

Gewirth will surely want to deny that the "categorial rules of action" are stipulative, or definitional, or otherwise purely linguistic, as I claim they are. What makes the PCC "necessarily true", he says, is some sort of intrinsic conceptual affinity between the notions of action, voluntariness,
and purposiveness. Being voluntary and being purposive are "the constitutive [italics mine] a priori conditions" of some event's being called an "action". I take this to mean that the link between voluntariness-purposiveness and action is not merely a criterial connection, such as that which holds between pain-behaviour and pain; if the connection is to be spoken of as "constitutive", I fail to see how we can avoid saying that voluntariness-purposiveness is a defining characteristic of action. In short, there is nothing which Gewirth will allow to count as a violation of the "categorical rules". Perhaps he wants to say that only voluntary and purposive actions admit of moral evaluation, but this is only a way of making his linguistic recommendation plausible, albeit a perfectly sound way. And perhaps he still wants to hold that we are dealing with a necessary conceptual connection which is not merely a matter of linguistic usage; but I see no way of making any sense of this. I think the burden is on him to show how such a connection could be possible.

The PCC does show that the Nazi cannot claim to be morally justified in his actions without contradicting himself, since he is denying to his victim, a human being and a potential agent, the right to make meta-decisions about relevance criteria which he claims for himself merely on the grounds that he himself is a potential agent. But Gewirth's logical apparatus seems to me a little elaborate; we do not need to refer to the definition of 'action' at all in applying the GP in this latter way. Moreover, he calls the PCC a "tautology"; we arrive at the PCC by denying the miscreant's maxim which entails a self-contradiction, thereby affirming that the statement 'In acting toward R I should allow him to participate voluntarily in our joint interaction' and its equivalents are necessarily true. But a necessarily true proposition is not tautologous merely because to negate it would be to misuse ordinary language. Gewirth claims that the PCC has the form of the tautology 'p → p', the antecedent referring to the agent's right to act and the consequent to the recipient's co-ordinate rights. But these two propositions, though logically related, are obviously not identical; and so the PCC does not literally have the form 'p → p' (its form, if we have to represent it in symbolic notation, is probably more like '(x) (y) (Rx = Ry)', this implication being established through the semantic interpretation of the predicate 'R' rather than through syntactic structure). If the PCC were a tautology, then it would have no "substantive content"; a tautology is true by virtue of the meanings of its logical connectives alone, while the PCC points out something essential about the use of the word 'right'; what Gewirth has shown is that, if I am open to moral discussion at all, i.e., if I am interested in using moral language and doing so correctly, then I am committed to calling certain acts morally wrong on the grounds that it would literally be nonsensical—unintelligible—to call them morally justifiable. This is a significant, and, I think, true conclusion, but we do

*Gewirth, op. cit., p. 293.
not need to call the PCC a tautology in order to obtain it. There is no such thing as a "descriptively significant" tautology, let alone a "morally significant" one. On the other hand, Gewirth is correct in noting that a person cannot deny, if he wishes to justify some action of his, that his action was voluntary and purposeful; and this does free us to some extent from the contingency of Hare's premise ('I have such-and-such inclinations') or Singer's ('Such-and-such consequences would be undesirable').

The question arises here as to whether there can be an argument to show that one must use moral language. How, i.e., is the moral philosopher to address himself to a seemingly "amoral" intelligence? I am inclined to say that there are only immoral intelligences, not amoral ones; but the problem is far too complicated to fall within the scope of this paper.

Gewirth’s paper, being a short and preliminary one, leaves us in doubt as to how he would handle certain common moral situations. First, he acknowledges that "many forms of human association inevitably involve conflict and hence coerciveness and frustration of wants". The persons concerned in such cases must therefore accept a system of certain first-order "specific rules" which will serve to balance wants and interests against each other, i.e., each person must consent to the frustration of one or more of his purposes. This would be in accord with the PCC, whose aim it is to "foster freedom and beneficence" in society at large; each member of society resolves ("coerces" himself) to accept certain rules in pursuit of this goal. The case of real coercion "for the good of society" is dealt with similarly:

Whatever sacrifices of individual interests they may sanction, in penal imprisonment, military conscription, exercise of eminent domain, taxation, and the like, these sacrifices, if they are to be justifiable, must serve the end of having a society free and beneficent or at least non-maleficient persons and hence must advance the welfare of persons having these traits.

The problem here is in determining what kind of coercive "sacrifices" do serve this end. When we try to discuss such considerations in a real situation, we frequently become embroiled in a nebulous debate very similar to that which can arise in connection with Singer’s PC. Gewirth will have to provide us with more specific criteria for applying the PCC in cases of this kind. The lack of such criteria becomes more acute when we examine interactions which, though morally wrong for social reasons usually expressed by some form of the GA, actually maximize satisfaction for everyone immediately concerned. How, for example, is the PCC to rule out my bribing the traffic cop who arrests me for speeding, provided that he takes the bribe voluntarily? Perhaps bribery of traffic cops tends to decay freedom and beneficence in the long run, but it will take a fairly impressive argument to demonstrate this. Gewirth could probably construct such an argument by showing that laws against speeding and bribery conform to the PCC in the two ways described on p. 298 of his article. Thus we would be provided

*Ibid., p. 298.
with a logical basis for institutional justice. On the other hand, in order to bring PCC-justified laws into prescriptive play, we need in every case the additional premise: 'It is morally wrong to break a law which conforms procedurally and instrumentally to the PCC'. Although I am perfectly willing to accede to this claim in the interest of justice, I cannot see that it proceeds directly out of the logical topography of moral language, as the PCC itself does. Certainly it is no "tautology".

We are tempted to say that Gewirth's system will have to allow for justifiable lawbreaking (civil disobedience in resistance to tyranny, etc.). But the premise I have introduced in the foregoing paragraph logically precludes this possibility. Perhaps this is as it should be: if a law is fully justified appropriating freedom and beneficence, then it is immoral to break it. What we customarily call "civil disobedience" generally arises only when the law in question is capricious and/or tyrannical.

One also wonders how the PCC would be applied to a situation in which some friend whom I know very well is almost neurotically undecided between two alternative courses of action, and I coerce him in some way to pick one over the other, knowing that his choice will make little difference in the end (furthermore, he later thanks me for seeing the right course, and for saving him a good deal of exhaustion and worry). In short, when may I force someone to do something for his own good, granted that he is psychologically normal enough still to count as a "potential agent", but also that I have a much better perspective on his particular circumstances than he? It might be held that he does count as an agent only so long as it is possible for me to argue rather than force him into acting. Gewirth might also reply that such an action is justified when it is beneficent, when I am trying to use my superior perspective to further my recipient's real purposes. But again, this is a very hard sort of thing to determine in a practical situation unless we are given more specific criteria for applying the PCC in such cases.

What of supererogation? The PCC cannot enjoin one to do supererogatory deeds, since if it did, they would no longer count as supererogatory. The PCC is a principle of duty, not a way of measuring degrees of virtue. And, more important, what about cases in which something we would ordinarily call supererogatory seems to become a duty, such as that of the Good Samaritan? If someone is helplessly in trouble, we want to hold, it is our duty to aid him if we can. It might be claimed that the PCC requires us to act in this way on the grounds that to do so is to promote beneficence again. But these appeals to the fostering of a beneficent society are beginning to sound suspiciously like appeals to our own inclinations or to rule-utilitarian consequences. The most central question that arises here is: can the PCC alone be construed as prescribing to me certain actions vis-à-vis a situation resulting directly from someone else's violation of the PCC, in which I am in no way involved? I think we must answer in the negative, if we are to preserve the strictly formal character of the PCC; it cannot enjoin me on purely logical grounds to repair damage caused by another.
We must fall back on inclinations or on the GA here. But this is not an undesirable move; Hare's and Singer's arguments are perfectly good where they are applicable.

I do not mean in the foregoing paragraphs to doubt that Gewirth can tighten up the POC mechanism so as to arrive at moral judgments about these various kinds of action. But he will have to do so if he is to build it into a workable moral system, and so justify his claim that it is the "supreme" principle of morality. I suspect that such a system will necessarily involve reference to people's inclinations and to the consequences of actions—I agree with Hare and Singer as to the indispensability of these considerations—but perhaps the PCC will be found to entail judgments about what kinds of inclinations are justifiable and what kinds of consequences would be undesirable. If so, it will itself provide us with independent logical criteria governing such evaluations, and so the criticisms I have levelled against Hare's and Singer's uses of them will not apply.

I think that all three universalizability principles, as Gewirth says, will suffice to refute the various non-cognitivist positions I mentioned in beginning, since each of these denies that morality has a rational basis. The PCC is an a priori principle of justice, whose observance is a logically necessary condition of the possibility of using moral language intelligibly at all. Having a logical basis, then, a PCC-based morality a fortiori has a rational basis. It remains to be seen whether the evaluation of inclinations and consequences can be built into such a system, but I see no reason why this should not be feasible.

University of Chicago.