An analysis and critique of the thesis of Universalizability in the moral philosophy of R.M Hare.

Thomas George Sharp
BA (Hons) Humanities
06829881
20th April 2010
Introduction

This dissertation is an examination and critique of R.M. Hare’s thesis of ‘universalizability’. Hare presented the first instantiation of the thesis to the Aristotelian Society in 1955\(^1\) and it formed the core of his moral theory from the publication of Freedom and Reason just under a decade later\(^2\). Although facets of Hare’s thesis are analytically complex, the intellectual problem that underpins it is relatively simple: what could be said to compel someone to act ethically and why should like cases be treated and acted upon as alike. Hare set out to present a justifiable answer to these questions that did not rely on an intuitive notion of morality. Körner states that an immoral person is one who would make profits from the violation of a law whose observance by other people is the condition of his gains\(^3\). This statement highlights the importance of universalizing concepts within moral theory; Hare argued that his universalizability demonstrated the necessity to universalize.

The first half of Hare’s universalizability is a forensic analysis of moral language; it is the language that once used, yields or should yield, a certain response, which is broadly this: A person who justifies a certain action under certain circumstances must also concede that under the exact same circumstances, another’s justification for the same action would also be justified. The three central premises that Hare expounds for universalizability are; first, moral judgements must be prescriptive; second, because they are moral judgements that they be distinguished from other judgements of this prescriptive class because of the fact that they are, third, universalizable.\(^4\)

Key to Hare’s thesis are the logical relations between prescriptive judgements, the way that some may be in contradiction with others, for example. If we accept that moral judgements are prescriptive and universalizable, then the only way to ‘get a grip of moral judgements’, as Hare put it, is to look at the logical relations between them to see if they can be rationally held against others. It is this premise which is broadly the subject matter of this paper and it will proceed as follows.

---


\(^4\) Hare, Ibid. Chapter 2
The first section is primarily descriptive and explanatory. It offers an overview of Hare’s claim that ‘moral language’ necessitates prescription and universalization by looking at the arguments he gives about the use of language and logic in moral argument. It then goes on to examine the substance of the universalizability thesis by looking at Hare’s exposition of inclination and logical consistency. Based on this, the paper provides reasons for why Hare’s thesis pursues the arguments that it does, thereby establishing the grounds for critique in the next two chapters.

The second section forms the main body of the paper. It offers an examination of Hare’s notorious ‘fanatic’. The paper provides responses to Hare’s defences of his theory in relation to the problem of fanaticism and also highlights problems in regard to Hare’s conception of his thesis; the remit of universalizability is assessing logical relations between moral judgements and it does this well, however I shall argue that the direction and function of universalizability becomes mired in confusion when Hare’s fanatic becomes ‘unassailable in argument’. This section goes on to explain that Hare takes rationality, on the part of the agent, completely for granted; I argue that because of this he expects the logic of language to instil the ability to make rational decisions into the agent by virtue of the usage and logical constraints that Hare claims are inherent in the words themselves. However, without an a priori acceptance of agent rationality that precedes his critical calculations, universalizability can only demonstrate the requirement of language to be used consistently in order for it to be intelligible. The tangibility of the thesis in relation to moral theory becomes therefore questionable.

The notion of compelling an agent to reject certain judgements in light of universalizability is distinct from Hare’s thesis as he himself conceived it. It is my contention that he is forced into this position because he is convinced that the use of the moral words necessarily imply a rational course of action, but the case of the fanatic shows that Hare is mistaken. When the moral judgements in question are wholly irrational - as in the case of the fanatic - Hare has to deploy extra-logical arguments against the fanatic in order not to sanction abhorrent moral judgements. Not only are these arguments mainly assertion, they are distinct from universalizability as originally conceived.

Moreover, Hare does not distinguish between the fanatic himself and what could be called a ‘disinterested third party’, which, I shall argue, may have kept the theory on track and

---

largely prevented it from deserting its own logic. This is because he overestimates the pertinence of logic in relation to moral arguments, so assumes that all one must do is display logical relations of moral judgements and the fanatic would be compelled to act on them. My amendments show that by concentrating on demonstrating that the fanatic is wrong, regardless what he himself thinks about his own moral judgements, universalizability could be a valuable thesis. This conceptual ambiguity leads Hare to go to great lengths in explaining how universalizability could stop the fanatic; but when universalizability begins to be ineffective, he relies more and more on a utilitarian understanding of moral reasoning to show how his thesis could ‘disarm’ the fanatic. However this is flawed, as contrary to what Hare maintains there is no necessary connection between universalizability and utilitarianism. Hare only develops this line of argument precisely because logic alone cannot disprove that the fanatic is a threat; thereby undermining the idea that a sound theory of moral reasoning can develop from moral language; because of this he then uses utilitarian arguments to show the fanatic is not a worry to his thesis.

The final section examines Kant’s categorical imperative (the moral law) and compares Hare’s universalizability to it. This is necessary as Hare believed Kant’s work to be in some ways compatible with the utilitarianism that he claimed arises out of universalizability. As these two theories fundamentally rely on universalizing, I examine what aspects of Hare’s thesis are inferior to Kant’s. I am unable to see, in terms of universalizing, the necessity for Hare’s work as opposed to Kant’s. The paper does not support a Kantian position but merely offers it as a superior model for understanding moral thinking.

The paper concludes that universalizability is a confused doctrine. In places it deviates from its own logic and moreover, because Hare judged it to be inextricably linked to utilitarianism, it absorbs even more conceptual problems in relation to the violation of groups or individuals in society. In this regard, Hare misses the point; it is not the likelihood of these problems occurring that is at issue and therefore needs defending, but that they are inherent in and the ultimate result of his utilitarian reasoning.
It is necessary to establish what Hare’s conception of Universalizability is; how it functions, what it is concerned with and what it is not. It refers, not to the moral properties of actions, but to the judgements that can legitimately be made about such actions. The content of the actions themselves is not the primary focus of Hare’s universalizability as it is primarily a logical thesis; universalizability seeks to establish that there are logical relations between moral judgements. Using language as his point of departure, Hare states that “[...] the meaning of the word ‘ought’ and other ‘moral words’ is such that a person who uses them commits himself thereby to a universal rule. This is the thesis of universalizability.”

By ‘moral words’ we mean *ought, should, must* etc. These are the subject matter of Universalizability as, according to Hare, the logical properties of these words show us when they are being misused. If we take the descriptive statement “all the players are wearing red.”; ‘all’ refers to every single object in a collection or group, and the rules of inference that govern the usage of the word ‘all’ would lead us to think that each player is wearing red. If upon checking, the observation proves to be inaccurate and at least one player is not wearing red, then the usage of the word ‘all’ is invalid as the empirical referent shows us that the word has been misused. Whether this has been done through error of observation or because of an attempt to mislead, the words usage does not correlate with its meaning in this context i.e. ‘every single object of the group’. It is through experience and usage of words that we know what inferences can be validly made, knowing these inferences allows us to know [almost] everything about the meaning of words like ‘all.’

I say ‘words like all’ as it is Hare’s contention that the words used in moral judgements are very similar to ‘all’ in that their meaning is almost exhausted by their logical properties, we can know almost everything we need to know about the moral words from inferences of usage. So, upon hearing a use of the word ‘all’ it would be inferred that every single object was being referred to as having a certain position or characteristic, “all the books are back on the shelf”, for example. They are to be seen as distinct from predicate words

---

8 Ibid pp30
which derive most of their meaning from predicate logic i.e. their place within a given sentence. So in the example “the book is red” the predicate word ‘red’ is intelligible because it tells us something about the subject which is the book. They get part of their meaning from their place in a given sentence, however unlike ‘all’, the rest of their meaning is distinct from logic. As Hare states, “[Predicate words] get the rest of their meanings (that in which blue differs from red for example) from something other than their logical properties. No amount of logic will show us the difference between blue and red.”10 The ‘moral words’ are therefore much more similar to words whose meanings are exhausted by their logical properties i.e. ‘all’ than to predicate words or adjectives.

Importantly, descriptive statements like the one above do not prescribe anything. However, when Hare gives examples of usages of ‘moral words’, they are part of imperatives. So, do the rules outlined above apply when we are looking at imperatives? With descriptive statements we can empirically check the validity of usage because the statement is referring to something. However, imperatives, because they are not purely descriptive cannot be shown to be mistaken in the same way. For example, the statement “He ought to say his prayers” differs from “all the books are red”. We cannot say that the former is incorrect in the way that we could of the latter.11

Universalizability can demonstrate what may be mistaken about the above example by revealing inconsistency in the logical relations between it and other imperatives. Secondly, moral judgements are universalizable because they contain reference to a characteristic or expectation of somebody; or an attribute of an object. In this way “Charities are good” is similar to the descriptive “The shirt is red” as anything that is identical to it in all its universal properties must also be good or red. If one were to say, “The shirt is red, but I can conceive of a situation identical in all its universal properties except that I would not describe the shirt as red,” they would have misused the word red. This shows how language, once applied in certain cases, must logically be applicable to other cases that are the same in their relevant respects. Language rules or meaning rules are created because of universal properties; once an object has a rule created for it by these properties then it

---


11 A brief point is needed here which concerns Hare’s usage of imperatives. To illustrate his argument that there are indeed some logical relations between prescriptive judgements, Hare takes the extreme case commands (imperatives) in order to show that there could be logical relations between these. It would then follow that there are logical relations between moral judgements.
seems to me that in order for language to be intelligible descriptive statements must adhere to these meaning rules.

This linguistic feature of descriptive statements is shared by imperatives in that, for Hare, there has to be consistency of usage. As aforementioned, Hare’s Universalizability does not refer to properties of the imperatives or judgements themselves but to the logical relations between statements; it shows us that a person may be committing an error of logic if he is inconsistent in his judgements. Descriptive statements are said to have a stable meaning which remains so across different classes, for example, the statement “the player’s shirt is red” and “the man’s face is red” or even an alarmist, “the light is red!” refer to a concept of red which remains consistent despite the different intentions of the statement. In this way, one cannot apply a descriptive term to one of these statements and refuse to apply it to another which is identical in all its universal properties.

Likewise, imperatives and moral judgements must also remain consistent if they are to be made about other situations that are identical in all their universal properties. Put simply, “the thesis of universalizability requires that if we make any moral judgement about this situation, we must be prepared to make it about any other precisely similar situations. Note that these do not have to be actual situations; they can be precisely similar logically possible hypothetical situations.”12 ‘Precisely similar’ is to be understood as pertinent features that would entail, logically, that the same judgement be made. Some features, place and time for example, may not be relevant, except in cases where they were directly related to the reasons given for issuing the original judgement.

Hare states that moral judgements are universalizable in the same way that other evaluative statements are universalizable. However, imperatives that contain particular reference, e.g. legal decrees, are not universalizable as they refer to a particular jurisdiction. For example, “It is illegal to marry one’s own sister’ means, implicitly, ‘It is illegal in England to marry one’s own sister.”13 (emphasis added) Moral judgements refer to a reason or grounds that necessitate universalization that is not dependent on certain jurisdictions; it is the logic of the word ‘ought’ that requires this, not the use of the word ‘moral,’ which refers to a particular class of judgement. According to Hare if one were to say, “I ought to act in a certain way, but nobody else ought to act in the same way in

---

12 Hare, Richard. M, Moral Thinking: It’s levels, method and point, OUP, Oxford, 1981 pp42
relevantly similar circumstances then, on my thesis, he is abusing the word ‘ought’.”

Hare states that the universalizability of moral judgements is due partly to the descriptive content of a judgement, because if a value-judgement is consequential on features of a situation, then the presence of the same features in another instance commits the one making the judgement to an identical judgement and secondly, that because when we make moral judgements we prescribe what to do in a given situation, universalizability implies that one who has judged in a certain way about one situation is constrained to apply the same judgment and the actions entailed to all like situations.

Furthermore, if we are to accept that a moral judgement is universalizable it follows that the principle that it refers to is universalizable. Hare states that when moral judgements are made we are then referring to a principle, if one already exists, or otherwise one is being set up by virtue of the judgement being made. So for example, X judges F in circumstances Q to be morally right, then a principle of X ought to sanction F if Q is alluded to because the properties of Q, for whatever reason, have led to the action F to be judged morally right. If the properties of Q then obtain at another time and F is not judged to be so, then the universal principle that had been established would have been violated. Note though, that this does not encourage X to do F if Q and I would argue to even compel him to judge in the same way; only that he would be inconsistent in his judgements were he not to do so.

It is here important to distinguish moral imperatives from ordinary imperatives which are, for the most part, relatively simple and general. Consider, as Hare does, the imperative, “Do not smoke.” This refers to a particular situation in which smoking is prohibited. It does not provide any reason beyond the rule itself. However, the statement, “You ought not to smoke,” implies that in other cases like this you ought not to smoke. Hare states that we should understand the latter as being governed by a universal quantifier and as not containing any individual constants. The first example is grounded in a transient notion and is particular to the current circumstances i.e. a train carriage with a no smoking sign, but the other clearly refers to a universal principle. So in universalizing a moral judgment, “He is a good man”, for example, we are stating a principle that men who act like this one

---

14 Ibid pp32
17 Hare, Richard. M, Moral Thinking: It’s levels, method and point, OUP, Oxford, 1981pp41
does are good men; and an abstract rule is thereby created by which all men who are said
to possess these characteristics must also be called ‘good’, so a logical expectation of a
similar commendation is imposed on the user of the term for all men who have those
relevant features. For Hare, description and prescription join when we make moral
judgements and it this conjunction that entails universal prescriptivism: Hare’s
universalizability.

***

In ‘Freedom and Reason’ Hare seeks to outline the bare bones of his thesis of
Universalizability with an example, which I have simplified for reasons of brevity:

B is aware that if he ought to assent to X, he will also be prescribing that X ought to be
done to all others in an identical situation; this would include him, he does not want X
done to him, so he cannot assent to both statements.  

He states that one of the things that is required for universalizability are certain
inclinations; in this case it is B’s inclination not to have X done to him. Someone in B’s
position would not pursue what might otherwise be pursued only because of this
disposition. B is therefore restricting his prescription because the result of him making that
prescription would be logically inconsistent with his disposition; therefore B must give up
his original ought statement or suppress his aversion not to have X done to him if he
wishes to remain coherent. As Hare puts it, “All [Universalizability] does is to force people
to choose between judgments which cannot both be asserted without self-contradiction.”

Let us look now at the central idea of logical relations in order to unravel this problem. It
is of no consequence for the universalizability test which of the above B performs - either
giving up his ‘ought’ statement or suppressing his aversion to having X done - because
both of them will result in consistency and therefore a credible challenge could not be
mounted against him in argument. What we see here is that universalizability is unable to
discriminate between what may be done to achieve consistency. In common examples this
may not prove to be a problem because if judgements have been made about one situation
then it would logically follow that the same judgement be given about situations identical
in their universal properties. However, in some cases, extremist ideology for example, a

18 The full example is included as an appendix on page 33 of this paper.
20 Ibid pp32
prescription could be readily universalized as it was personally consistent with the agent but it would produce severely detrimental consequences for the society in which that agent lived. The idea of personal consistency or integrity is an idea that is central to my critique of Hare’s thesis. It is this idea that prevents Hare from achieving his goal of getting shot of the fanatic, and because of this problem he has to cook up increasingly complicated ways of suggesting that the fanatic is not really a problem, which will become evident in the next chapter.

Examining an instantiation of the Golden Rule will crystallise the point regarding personal consistency: “And as ye would that men should do to you, do ye also to them likewise”.21 It is clear that like universalizability, this focuses on intra-personal consistency. In the case of the Golden Rule, it is an instruction that would prevent behaviour that one would not want meted out against oneself. The critical process would be as follows: X may be considering doing F against B but is aware that he would not want F done to him. Therefore he does not do F. So, the logic of the rule constrains X’s actions. But notice that because the rule is concerned with personal choice, when an anomaly occurs, the Golden Rule has to sanction things that would appear to many as wrong. For example, X very much likes violence, so much so he derives pleasure from carrying it out and having it carried out against him. The Golden Rule would therefore sanction this conduct.

This may first appear trivial, a tenuous example perhaps, but the point here is not an issue with the Golden Rule. It is to demonstrate that Hare’s universalizability suffers from a fundamental flaw; it comes to assess moral judgements with a gross generality i.e. normally people have a certain disposition and for the most part they want to live in peace and security with others and so would practice this in their conduct. A not so trivial problem arises when this is not the case.

Let us pause here and ask whether the above assessment is fair. The universally prescriptive character of moral language is what Hare sought to demonstrate; universalizability is not the same as the Golden Rule because of the universal prescriptivity of moral language. What universal prescription does, because of the use of the moral words, is to show that if one ‘ought’ to do something, then he is prescribing that in situations identical in their relevant respects, others ‘ought’ to do the same. By virtue of

this he is automatically assenting that the same thing be done to him because he is in the same position. An additional example from Kant is of use here. In the *Critique of Practical Reason* he writes, “Suppose that someone says his lust is irresistible when the desired object and opportunity are present. Ask him whether he would not control his passion if, in front of the house where he has this opportunity, a gallows were erected on which he would have to be hanged immediately after gratifying his lust. We do not have to guess very long what his answer would be.” The point here is not about punishment, but about assenting to something without thinking of detrimental implications. An agent may be less inclined to do so once they had taken stock of all the consequences that may affect people in counter positions, which as universalizability shows us, would include themselves. The use of the language, because of its nature, makes one reflect upon the affect that a certain prescription may have. However, the fact remains that if someone prescribes something for others that they would accept for themselves, there is no other criteria with which to challenge their judgement. Their own integrity in judging is what is at issue.

Contrary to this observation, Hare is convinced that universalizability would prevent actions that would prove to be massively detrimental to certain people because he assumes that a commitment to the logic of moral language itself is enough to provide sufficient reason for the agent not to assent to the action in question. We will see however that Hare should not have been so confident about this. This problem in Hare is exemplified in his conception and treatment of the Nazi fanatic. A way to understand it is to look at Hare’s *focus* of his argument in relation to universalizability. He attempts time and time again to demonstrate to someone with deeply held fanatical beliefs that they are mistaken and so should be compelled to be consistent in their assenting to judgements i.e. so they do not conflict.

However, this begs the question: what would it mean if logical coherency was of no concern to an agent? Is there even a demonstrable link between being consistent in one’s judgements and being moral? Would an agent who renounced consistency and deviated from logic be an *amoralist*? Hare states that, “a man’s moral principles [...] at the critical level, should be consistent with one another.” However, consistency does not necessarily

---

24 Ibid pp184
entail morality, as we have seen. Furthermore, if as I have suggested, an agent cares not for consistency, then where does Hare’s theory go? Hare addresses this in reference to the ‘impure fanatic’\textsuperscript{25} but does not do anything to analyse the underlying implications.

The idea of inclination is central to Hare’s thesis. It would appear that the only tangible motivation for the constraint of certain prescriptions is an inclination that is broadly self interested, i.e. assenting to a certain moral judgement would mean assenting to a rule that may harm me so one would avoid that judgement. However, Hare claims that upon universalizing, one must disregard his own inclinations and treat the inclinations of others who would find themselves in an identical situation as if they were his own.\textsuperscript{26} ‘Identical situation’ refers to the properties that are relevant to and consistent through past and present scenarios. So, one is required to give up his own inclination, factor in the inclinations of all relevant parties and then issue a prescription on that basis. This is rudimentary utilitarianism; governed by the pre-eminence Hare bestows upon the satisfaction of preferences. But if this is not a necessary move, why does Hare arrive here and pursue this line of argument. To attempt an answer to this question a thorough examination of Hare’s fanatic is necessary.

\textsuperscript{25} See pp16 of this paper.
Having offered a presentation of Hare’s thesis, I will now examine the case of the Nazi fanatic in detail and attempt to show the problems that it causes for universalizability. The damage the fanatic does to universalizability is hugely problematic in itself, but the ‘damage limitation’ Hare attempts to enact led him to some dubious and unfounded conclusions, and it is my contention that they make the universalizability thesis thereof irredeemable.

In order to get to this point however, we must look at how Hare sees and treats the fanatic. Hare separates fanatics into two categories: ‘impure’ and ‘pure’ fanatics. The impure fanatic is one that either has not understood what it means to prescribe universally, so issues immoral prescriptions, or an individual who is unwilling or unable to engage in the critical process of universalizing. The existence of the latter cannot be blamed on Hare’s universalizability; it is not the fault of a thesis that it has no effect on someone who is unwilling to understand them. It is clear then that when someone does not ‘wish’ to remain coherent Hare in effect discounts them as I suggested earlier. This represents Hare’s commitment to the idea that the logic of moral language entails morality, therefore someone who does not wish to be coherent is not involved in a moral argument- as Hare conceives it.

Let us look then at a possible example of an ‘impure’ fanatic. One wants to challenge a person in argument who advocates the dictum, “We ought to limit the rights of Asian people.” As this is a prescriptive moral judgement it is universalizable; in the sense that it would entail identical judgements to be made about all cases identical in all their universal properties. Thus the thesis demands of him that if he states that he/we ought to do a certain thing to a certain person or persons, he must be committed to the view that the very same thing be done to him, were he in exactly the same situation, including having the same characteristics and motivational states. So, could he still logically issue his ought statement, given the knowledge that he would now be subject to his own prescription? Universalizability reveals that the prescription and his disposition to not be subjugated would be in direct contradiction; universalization shows that the ‘ought’

28 Ibid pp108
statement is untenable as there would be an intra-personal conflict between it and his disposition.\textsuperscript{29}

The racist may well want to give up his prescription in light of this argument, or he may not, but universalizability has done its job by revealing the contradiction.

A case of a ‘pure fanatic’, a fervent Nazi for example, has a conversely problematic outcome for Hare. Let us imagine that we are confronted with a Nazi fanatic. The individual holds the ideal that to create a better society all Jews must be eliminated. So, in order to convince him that he is mistaken, that particular ideal is universalized. It can be correctly stated that the individual is not in error if he adheres to the belief despite taking account of the harm that his ideal would cause from the perspective of a Jew. Moreover, he can retain the ideal even if he takes on the interests of the Jew \textit{as if they were his own}. Hypothetically then, the Nazi becomes the Jew that he wants to persecute. However, given that the individual is a committed fanatic, he will hold on to the belief even if it flies in the face of his own self interest as the hypothetical Jew.

Hare approaches the case of the fanatic under the assumption that following the logic of moral language implies rationality, but it clearly does not; in fact universalizability cannot demonstrate that the fanatical argument is logically flawed, because following Hare, it would not be: “Moral judgements are universalizable in only one sense, namely that they entail identical judgements about all cases identical in their universal properties.”\textsuperscript{30}

And it would appear that the Nazi has done this: If a person believes in something firmly enough, to the degree that they would forego their own life, and because they then remain true to their ideals, they are not therefore ‘wrong’ following universalizability. Furthermore, given that Hare’s Universalizability is not morally substantive and only highlights logical inconsistencies, we would not be able to show through argument that there is anything mistaken in his beliefs because there would be no logical grounds to do so.

Despite this, Hare draws parallels between the Nazi’s ideal and an aesthetic principle, to show that the principle can be helpful in these extreme cases. Put simply, Hare likens the desire to rid a society of Jews for its supposed betterment to the story of Heliogabalus

\textsuperscript{30} Ibid pp107
who, according to Hare, took a liking to slaughtering people just because he enjoyed the colourful display of spilt blood on green grass. Hare claims that, “[The] ideals have, on the face of it, nothing to do with self-interest or with a morality which can be generated by universalizing self-interest; they seem much more akin to [aesthetic] evaluations [...] the enormity of Nazism is that it extends an aesthetic style of evaluation into a field where the bulk of mankind think that such evaluations should be subordinated to the interests of other people.”

But the Nazi’s desires were not aesthetically guided, unlike Heliogabalus’s actions, the only reason that they could be described as distinct from moral judgements and more like aesthetic judgements is that prescriptions when universalized fly in the face of self-interest and Hare cannot account for this. By drawing this parallel, Hare is attempting to show that the fanatic can still have arguments brought against him because his abhorrent prescriptions can be logically universalized. The Nazi’s prescriptions may be aesthetic in style but are very different in kind, as Hare himself sets out in some detail. Consider, “[The Nazi] accepts the principle that the characteristics which Jews have are incompatible with being an ideal or pre-eminently good [man]; and that the ideal, or even [a good] society cannot be realized unless people having these characteristics are eliminated.”

Ideals like these, I would suggest, are distinct from the actions of Heliogabalus. Hare writes, “Heliogabalus, [must] if he is consistent, desire that anyone (even if it be himself) should be slaughtered, if that is necessary in order to gratify the taste of someone who likes red and green juxtaposed.” This shows us that an ‘aesthetic fanatic’ is, unlike the Nazi, sheer fantasy, because there is no evidence whatsoever that anybody like this has ever existed; unlike the fanatical Nazi. I would suggest that the comparison is something of a rearguard action from Hare; an attempt to show that the real world fanatic is something so extreme that it would not pose problems for his theory, for he writes “That we have to look for such a farfetched parallel (i.e. Heliogabalus) as this indicates that we have at least diminished the problem of dealing in argument with the Nazi. For it shows that real intractable Nazi’s are rarer than might be thought.” They are not parallel; Hare’s contention that these individuals are rare almost to the point non-existence is mere

---

31 Ibid pp161
33 Loc. Cit.
34 Hare writes of Heliogabalus: “I have been unable to find the source of this story so cannot therefore say whether it is true.
assertion and has nothing to do with the conceptual problems of universalizability that the occurrence of fanaticism highlights.

In order to convince us of the efficacy of universalizability in the face of fanaticism, Hare brings home to the Nazi the implications of issuing a truly universal prescription:

“We say to him, ‘You may not know it but we have discovered that you are not the son of your supposed parents but of two pure Jews; and the same is true of your wife’; we produce firm evidence to support the allegation. Is he likely to say-as he logically can say- ‘All right then, send me and my family to Buchenwald!’? Then let us imagine saying to him that was a deception; the evidence was forged, but now having faced this possibility, do you still feel the same way about Jews?” 36

Hare claims that the purpose of this is to get him to make a universal prescription in which the Nazi really believes or discard his prescription. But let us pause here. Hare’s thesis already admits that when faced with a ‘hard-core’ fanatic there may result no logical error and therefore no grounds to challenge him in argument. We must remember that according to Hare, “the thesis of universalizability was established by arguments of the philosophical-logical sort“[most important of which] is in showing that a person who makes different moral judgements about cases which he admits to be identical in all their non-moral universal properties encounters the same kind of incomprehension as is encountered by logical inconsistency (for example self contradiction)” 37

A critical problem for Hare is that the ‘hard-core’ fanatic, the one who does demand that his family be sent to Buchenwald, is in possession of impeccable logic. This is what Robert Fullinwinder has called the ‘reversal test’: “To meet the universalizability requirement the Nazi must adopt a consistent stance, either by giving up his prescription to kill Jews or by overriding his desire not to be killed if he is a hypothetical Jew.” 38 The Nazi’s position is consistent and universalizability alone can do no further work in tackling him.

Hare’s quite ingenious attempt to show that the logical properties of moral language demand that moral prescriptions be universalised is a project that arose because he wanted to provide grounds for moral reasoning that did not rely on what he saw as arbitrary intuitionism. It is an effort undertaken to provide a firm basis for the refutation of certain moral judgements by showing that they had been made in error, and as a consequence,

36 Loc cit
37 Hare, Richard. M, Moral Thinking: It’s levels, method and point, Oxford pp115
38 Fullwinder, Richard. K, Fanaticism and Hare’s moral theory pp166
could not be sustained. The problem here lies not in the concept that we should be compelled through moral language to treat like cases alike—but in the substance of some cases where the thesis of universalizability is applied; in these cases moral language is not enough.

Through an examination of the fanatic we have seen that immoral judgements can be universalized logically. It would seem that Hare’s initial adherence to logic allows him little room to manoeuvre when he is confronted with the fanatic; after the hardcore fanatic passes the ‘reversal test’, Hare has to look beyond logic with assertions about the occurrence of ‘hardcore’ fanatics and later by reference to strengths of competing desires and preferences. The credibility of this latter defence will be addressed later.

Presently, I will turn to a different aspect of universalizability and examine whether Hare’s thesis adds anything to our ability to reason morally. It is necessary to outline what it could do when we are considering differing moral judgements. For this section I will look at the position of a rational third party or bystander, uninvolved in the actual dispute. Let us take a contemporary example; whether terrorist suspects should be detained for 42 days without charge. The statement relies first on something about the suspects; be it that they are too dangerous or that they represent a ‘new kind’ of criminal, or whatever. Universalizability would demand that if the same features were present in another instance then the same judgement be made, secondly that one who has judged in a certain way about a situation is constrained to apply the same judgment and the actions entailed to all like situations. It must be noted that presumably government legislatures are not hard core fanatics and that if the ‘reversal test’ was put to them they would not want to be subjected to this arbitrary detention; it is presumed they are rational and would respond if their position was shown to be illogical. In any case, the reversal test demonstrates analytically to a third party their error in argument.

However, aside from logical inconsistency, what would universalizability demonstrate to a third party about the morality of the prescription? They could have no quibble with the fact that Universalizability reveals inconsistency, but could still be convinced by the empirical evidence relied upon by the person issuing this particular prescription and so deem it to be morally justifiable. I would argue that a possible third party should be concerned with the initial justification for implementing the legislation and the empirical evidence that allegedly backs it up, taking an analytical approach to it. As if we were to
concentrate solely on the issue of consistency as revealed by universalizability, then like the Nazi fanatic, if the prescriber were to believe in their prescription strongly enough, the position could be universalized. The prescriber would be unassailable in argument- as they would pass the ‘reversal test’- and most importantly of all the rational third party would have to endorse a wholly irrational position even if he was not convinced by it merely because it was consistent.

The example highlights the problem that arises from trying to convince or compel a person who holds an incoherent position; because if they believe it firmly enough, the universalizability thesis cannot challenge them. Hare could have gotten round this problem by showing who exactly he was demonstrating the efficacy of his thesis to.

***

Hare relies fundamentally on a general inclination of self interest; of agents not being wantonly destructive. He writes:

“Let us [consider] the case of someone who really wants to kill everybody, himself included. Should it disturb [us] that no argument can stop him wanting to do this, if his desire is sufficiently intense and unshakeable? Should we not rather say, ‘[At any rate, thank goodness] not many people feel like that? [This is drawing] comfort from the happy fact that most people want to live.”

This is a statement with which most of us would agree. It is clear that to be rational is not to be like the man in Hare’s example; this is precisely the problem that Hare has to take account of because his reliance on the logic of moral language to make safe the prescriptions of an irrational agent fail. The agent in question can be completely logical as per Hare’s criteria whilst being wholly irrational which results in pure immorality, as I have shown in the case of the fanatic. Logic without rationality will inevitably lead to the sanctioning of immoral prescriptions precisely because it is rationality that is a prerequisite for being a moral agent.

---

39 Hare, Richard. M, Freedom and Reason, OUP, Oxford, 1963 p172-173 (Interestingly this quote is characteristic of Hare’s impatience with what he viewed as ‘fantastic examples’ that hold no weight in arguing against the virtues of a moral theory. The quote itself is the most extreme kind of one of these examples and is in fact a caricature of legitimate criticisms because it is so extreme. I would suggest this serves to denigrate and show as fanciful critical points that are like this. In any case, he does not address the fact that he is having to entertain such criticisms because of very real flaws in his theory.)
This is a fundamental problem of Hare’s Universalizability when the fanatic comes into play. Logic alone is useless and actually turns against him. All Hare can do is rely on a general intuition about humanity to defend his theory; and it was unpredictable intuition that he had hoped to extricate himself from by implementing a theory based on the logical properties of moral language.

As mentioned, the above quote from Hare is uncontentious, however this is beside the point as it is not the validity of the empirical claim about the nature of human beings that is at issue. Hare set out to rid himself of any intuitive notions of morality and to set out a system of moral reasoning that relied on language and logic alone, the case of the fanatic demonstrates that it would seem one cannot employ arguments based on the logic of moral language without first declaring that certain prescriptions are, by their nature, immoral and are therefore to be discounted. This is a problem for Hare specifically because by not addressing a rational third party to which to show the irrationality of a judgement that declares that everyone who is Jewish needs to be murdered, he enters a bargaining process with an agent whose prescriptions should never under any circumstances be entertained; though of course his reliance on logic compels him to do so. Furthermore, his utilitarian reasoning compels him to satisfy them. In order for this satisfaction not to obtain, Hare falls back on a notion of intuition but this merely looks like an ad-hoc and unconvincing argument. Worse still, his reliance on arguments about the empirical nature of humanity reveal that there are critical flaws in Hare’s reasoning, demonstrated by the very fact that he is forced into such a position. It serves to cover up fundamental problems that he is unable to address. But this is not all. It would appear that hard core fanaticism is the most extreme example of a further problem in Hare’s universalizability which is magnified by his increasingly utilitarian explanation of how he can rid himself of the fanatic.

In the essay *Wrongness and Harm* Hare comments that what lets the fanatic in is the frustration of future desires at the expense of present desires. He writes, “[Fanaticism] consists in setting so much store by some prescription or other that I am prepared to see it realised even if my other prescriptions etc., in hypothetical or even actual circumstances are thereby frustrated.”

---


to the contrary, even if they suggest that great suffering would be afoot if the original
prescription is adhered to. It is this belligerence that is the first problem for Hare; the
fanatic just has to possess an almighty desire to see his preference satisfied, for he writes in
Moral Thinking “[The fanatic] has to claim that his own preferences are so strong and
unalterable that they will continue to prevail over those of the others whom his actions
will cause or allow to suffer. If this claim be granted, then critical thinking will endorse the
universal prescription that in such cases the fanatic’s preferences should be
implemented” (emphasis added)\textsuperscript{42} However the fanatic need only claim that he is in
possession of the stronger desire and that his preference outweighs all others. Hare can
only claim, as he has to, that this is such a fantastic example that we should not be
concerned that universalizability has allowed the prescription to be satisfied. His response
in sum is, “If we address [the theory] to queer cases, we shall get queer answers.”\textsuperscript{43} Again,
Hare relies on empirical assumptions about the world which are assertion, although they
are presented as the sum of argument. In any case arguments of this type are distinct from
the logic between moral judgements and what we described as the ‘reversal test’.

Hare has employed a utilitarian argument, whereby the strength of preference necessitates
an endorsement of that preference regardless of the content. Because universalizability is
unable to deal with it and in a sense endorses it, he is now obliged to assert that the
likelihood of anything like it ever happening is near impossible because, as Hare himself
says, “It is obvious that [I will have] to say that such a heroic desire, if it occurred, \textit{ought}
to be satisfied.” (emphasis added)\textsuperscript{44} However, the utilitarian route does not appear to
rescue universalizability and it is difficult to see where this takes his argument. Consider:

“Let us suppose that there are ten thousand Nazis, all of whom have a moderate desire not to
look at [Jewish] noses. And let us suppose that there are a thousand people with [Jewish] noses.
[...] It is obviously possible to imagine a case in which the numbers involved, and the intensity
of the respective desires, are such that satisfactions really would be maximised, if those with
[Jewish] noses were painlessly exterminated. [...] [But] I believe, that is to say, that as the
intensity of the desire required is diminished to a plausible value, the number of fanatics
required to compensate for the diminution of the intensity of their individual desires increases
to an implausible value. So that the case, taken as a whole, remains implausible”\textsuperscript{45}

\textsuperscript{42} Hare, Richard. M, \textit{Moral Thinking: It’s levels, method and point}, OUP, Oxford pp181
\textsuperscript{43} Ibid p182
\textsuperscript{45} Loc cit
Put simply then, if we are to put the ‘super man’ type fanatic to one side and replace him with many run of the mill fanatics, the number of them that would be required to exist for their preferences to be satisfied would be so big as to be incredibly unlikely to occur. However, the invocation of a scenario where many fanatics exist actually make things worse for Hare. If we are talking about the persecution of minority groups, which in this case we are, it would seem that the likelihood of circumstances obtaining that would demand this persecution is far from implausible. Hare’s confidence about this is dangerous; consider the high number of supporters for Enoch Powell after he delivered his ‘Rivers of Blood’ speech, does this not represent or come close to an actualisation of what Hare viewed as near impossible?46

The case of the fanatic shows us that because of the strengths of preferences of certain groups or individuals, that universalizability would actually sanction judgements that would violate other members of society. I deliberately use the word ‘affect’ because there are a variety of examples within which we can see this problem - the fanatic is just an extreme example of a pervasive problem in Hare’s reasoning. A notable example is given by Fullinwinder, “Although Hare claims that it is implausible that the Nazi’s desire to kill Jews could ever outweigh the Jew’s desire to live, he cannot seriously maintain that, say, a large majority’s ideal of decency could never outweigh the desires of a limited minority to see pornographic films.”47

Much like his arguments that concentrate solely on moral language, Hare’s utilitarian defence is flawed because it does not preclude anything; if something is wanted or desired, and the preferences and numbers are strong enough, then it ought to be satisfied. It is what Nozick has called an ‘end result’ theory, whereby the legitimacy or acceptability of claims to certain goods or satisfactions are not looked at.48 The main flaw with Hare’s defence is that once he entered into a utilitarian defence of universalizability, he was obliged to sanction preferences that other moral theories, Rawl’s theory of justice49 for example, would have disregarded from the outset as illegitimate. A moral theory should not be entertaining the fanatic and engaging with ideas of the kind he advocates because firstly,

---

46 Goldbourne, Harry, Race relations in Britain since 1945, Macmillan Press Ltd., Hampshire pp55
47 Fullwinder, Richard. K, Fanaticism and Hare’s moral theory pp173
48 Hondrich & Burnyeat (ed), Philosophy As It Is, Pelican Books, Middlesex pp92
49 Briefly Rawls claimed that it is a defining feature of our sense of justice that ‘interests requiring the violation of justice have no value and so the presence of illegitimate preferences would therefore not distort our claims upon one another.’ See: Kymlicka, Will, Contemporary Political Philosophy: An Introduction, OUP, Oxford, 1990 pp 41 and pp52
they are by their nature irrational, secondly, despite the fact that the ideas are irrational, he will never be convinced of this because he is a fanatic, third, universalizability is not something that could compel anyone to judge morally, let alone an individual who is so vehement in his beliefs that he would want himself destroyed if he were found to be a Jew. The case of the fanatic shows us that Hare’s thesis admits anything; worse still Hare was convinced that by using the logic of moral language in argument, all unseemly prescriptions would be weeded out. Following Hare, fanatical arguments, indeed any arguments that infringe on the rights of others are points of view worthy of consideration.

Much of this could have been avoided if Hare had made clear that the theory was a tool to demonstrate that the Nazi fanatic is mistaken. To show why the fanatics ideas are morally wrong to a third party is something, I would argue, that is extremely valuable because it shows that they are fanatics because they are mistaken in their reasoning.

Hare states, quite convincingly, that if we judge the similar circumstances, including hypothetical circumstances in which I would be in precisely his position and if (and I stress the if) I am seeking universal prescriptions or prohibitions or permissions to guide people’s conduct in circumstances like these, I shall not be likely to accept a system of such prescriptions which does not contain a prohibition on such acts- acts which, if I were in precisely his circumstances, including having the same desires, would harm me. What is problematic is the idea of ‘guiding conduct.’ To guide conduct could have meant to demonstrate by example, to show how moral reasoning is to be done or not done. Demonstrating the flaws of a moral argument to a disinterested third party would have helped Hare’s thesis; it would not have deviated from its own logic, because it would not be trying to use logic as an imperative to convince someone they must take a certain course of action. Universalizability would elucidate the virtues and highlight the contradictions of moral claims; as evidence is presented to a jury. The approach Hare took seems akin to convincing a defendant of his guilt whilst at the same time trying to rehabilitate him, when what matters is demonstrating to others whether he is guilty or not. No distinction is made between demonstrating inconsistency and using logical incoherency as an imperative in itself.

Hare’s work on the fanatic appears to be that it is them we must convince if our theory is to be of any use. Categorically demonstrating to others the error of the fanatic’s ways

---

seems to me a very good way of ensuring that people do not adopt fanatical views in the first place.
In this last section, I will examine how Hare’s thesis developed in the way that it did. It will be necessary to look to Kant in order to understand why Hare embarked on his thesis of universalizability.

Hare’s universalizability is intellectually rooted in the categorical imperative. Kant explains that an action is moral if ‘I can also will that my maxim should become a universal law.’ Or, more explicitly, “act only on that maxim through which you can at the same time will that it should become a universal law.”  

51 The question to ask is what exactly is it that Hare’s universalizability does better than Kant’s categorical imperative? Or has Hare simply produced an inferior concept?

Hare starts with language as his base. The logic of moral language allows one to see errors of moral judgement. It is for this reason that Hare’s Universalizability differs from Kant’s categorical imperative as it does not provide any basis for saying that a certain judgement is actually immoral. Consider one of Kant’s own examples, “If I believe myself short of money, I will borrow money and promise to pay it back, although I know that this will never happen.” 

52 Hare’s universalizability would show that the prescriber would be inconsistent if and only if he had an inclination or disposition not to have money borrowed from him and not repaid. This may seem obvious, but Hare requires an inclination to exist for the universalizability thesis to reveal any inconsistency, and for a subsequent challenge to be issued. If the prescriber accepts that the possible damage done to him by people prescribing this action unto him is outweighed by the possible benefits gained then he can universalize the prescription consistently and will not be committing an error of logic. The key difference that is revealed by this comparison is that Hare’s thesis would seem to be something that is concerned with language and its correct usage. For example, if I were to demand of someone that they ought to do something, and then later in a situation precisely similar involving myself, I were to issue a different prescription, then the word ought would have been misused as its usage pertains to a general principle that is being alluded to, for example, “You ought not to smoke around children.” So, what universalizability would show us is that the word ‘ought’ has been misused and an inconsistency has arisen as a result. The person can therefore be shown to be in error in a

52 Ibid p137
demonstrable way as opposed to saying, treat all situations in an equal manner, as that has no basis other than the rule itself. From this idea of usage rules, universalizability shows that someone cannot assent to judgements that reveal preceding judgements to have been made in error, as the many examples I have given throughout the paper attest to. Key to Hare’s thesis then is that judgements have to be judged against other judgments.

To refer back to the money lending example, Kant claims of this problem: “If this maxim were to be elevated into a universal law such promises and the purposes they would serve would become impossible, so the maxim is not moral.” Kant’s criteria are much more rigid than Hare’s. It needs no other judgement or personal inclination to be present for a problem to arise; whereas with Hare, the inconsistency lies between self interest and the prescription made i.e. to prescribe is to universalize and to universalize is to prescribe unto me, which the prescriber’s inclination forbids. Kant’s ‘impossibility’ it should be properly called, arises in this example because the maxim is elevated to a universal law against the empirical world and is not therefore an intrapersonal contradiction, as with Hare, but one that simply could not be universalized.

The fact that Hare’s universalizability highlights logical inconsistency primarily against the prescriber’s self interest is curious. It would appear that because it is only concerned with relations of logic between prescriptions and self interest it is something that takes measure of the prescriber’s integrity. By this I mean if the person believes in something enough to be subject to his own prescription he has logical integrity. To refer back to the fanatic, the only difference between him and his opponent is that he is willing to stick to his judgements even when they conflict with his own interest in hypothetical cases and therefore, he has logical integrity. For Kant, it does not matter if someone would be willing to accept the consequences of never paying back a debt despite promising to do so, as the individual’s integrity is not at issue; what is at issue is whether or not the maxim could possibly be universalized and for the concepts to remain intelligible, i.e. how could there be such a thing as a promise if there was no expectation of it ever being kept?

***

From this reference to Kant’s categorical imperative, it is my contention that Hare’s universalizability suffers from a major drawback which Kant’s work does not. Put simply,

53 Ibid p137
as universalizability focuses on logical relations between judgements, it would appear to be something that takes stock of a prescriber's logical integrity and allows us to see when moral language has been misused, providing tangible grounds to undertake moral argument. I have shown by reference to fanaticism that this is flawed. However, even if this was not the case, I would suggest that the logic of moral language is too problematic to ground a moral theory in. It is clear that Hare wanted to develop a theory that was purely analytic; a theory that did not rely on a concept from outside the theory that would demonstrate its validity (on this point he took great issue with intuitionist philosophers) and a theory upon which all the terms could be universally agreed.54

But what actually resulted was, in my view, something completely distinct from these aims. Due to Hare using the logic of moral language as the main criteria for universalizability he has to unfortunately sanction irrational and immoral prescriptions. Because of this, he then has to use a utilitarian argument to suggest that this would not be a problem for his theory, which by its nature has to make assertions about eventualities in the world i.e. a fanatic’s prescriptions would never be strong enough to outweigh a Jews desire to stay alive, but again this is unconvincing.

To return to the example of promise breaking, it could be said that Kant’s example suffers from exactly the same problem because to a certain extent it measures a maxim in terms of its viability as a possible universal law in the world. However, Kant’s example seems much more cut and dried. The empirical extrapolation concerning the universalization of promise breaking is used to show that the maxim that underpins it is in violation of the moral law (categorical imperative) and that is why, on a Kantian view, it would be immoral. The maxim is something upon which a subject acts or can act; the moral law itself contains only the necessity that the maxim should conform to this law.55 Hare has no underlying concept to rely upon. He is solely reliant on assertions that the likelihood of events obtaining in the world which would allow fanatical prescriptions to be satisfied would be rare. His position is therefore weaker than Kant’s.

***

54 Hare states that the meaning rules of language are not specific to any one language and it is in this way that they would seem to provide a universal starting point for moral theory. (See chapter 1 and 2 Hare, Sorting out Ethics)
I have argued that because Hare’s universalizability is based on the logic of moral language, this resulted in it suffering subsequent problems that Kant’s categorical imperative does not. I am of the view that because Hare’s universalizability is actually quite similar to Kant’s work on the moral law- in the way that it requires prescriptions to be universalized- he believed his and Kant’s work were compatible. Due to Hare advocating utilitarianism, I would argue that he sought to vindicate it by attempting to show whether Kant could have been a utilitarian. To elaborate, because of the damage that was done to Hare’s thesis by the fanatic he steadily resorted to utilitarian arguments to limit this damage; his universalizability developed into something that implied utilitarianism and on page fifteen of this paper I gave brief reference to this. I will now return to it and see whether in fact universalizability naturally moves into a certain kind of utilitarianism or whether it is something that arises simply in response to the fanatic.

Hare focuses on preferences and inclinations, for example when universalizing one is to treat the interests of others as if they were his own. An example from Hare will clarify: “[Another] party wants me not to move his bicycle, but I want to move it in order to park my car. I am fully aware of the strength of his desire, and therefore have a desire of equal strength that, were I in his situation, the bicycle should stay where it is. But I also have my desire not to move it in order to park my car. This latter desire wins by superior strength.”56 The example is long winded and goes on for some time; however this extract is enough to give an impression of how the move into utilitarian reasoning develops. All preferences are taken into account and the strongest one wins. Universalizing allows someone to fully represent himself to the situation of the other and in doing so equal weight is given to all parties concerned; and this according to Hare, is to obey Bentham’s injunction “‘Everybody to count for one, nobody for more than one,’ and if we are to treat equal preferences as of equal weight, utilitarianism is the result.”57 However, universalizing prescriptions would not seem to be necessarily related to overall preference satisfaction (the defining feature of the type of utilitarianism advocated by Hare) most obviously because one could properly universalize a prescription say for example ‘all egregious tax evaders ought to go to jail for a long time’; which would entail myself being subject to my own prescription if I were to evade tax in the extreme. The result of this may be a depletion of overall preference satisfaction but this does not mean the prescription has

56 Hare, Richard. M, Moral Thinking: It’s levels, method and point, OUP, Oxford p111
57 Sorting out ethics p151
been universalized improperly because I can assent to it as well as accept that is to be done to me. What should be of concern is that this kind of reasoning, at its heart, can be reduced to something quite dangerous: If enough people assent to something and strongly enough, it ought to be satisfied.

Hare was preoccupied with ‘strength of preference.’ This has two problems; again this echoes back to the logical integrity of a prescriber, if someone wants something enough and deems it more important than other preferences then on Hare’s view it should be satisfied, regardless of the legitimacy of the agents claim. Secondly, why does Hare choose preference and the strength of it by which to arbitrate scenarios? In this way Hare’s universalizability is, in my view, dangerously egocentric.

This is the part of Hare’s thesis that shows that it is distinct from Kant’s moral law. For Hare, the satisfaction of preferences or ‘wants’ is the sum of morality and as such his thesis seeks to find a way of discriminating between competing preferences; this results in him having to adopt a utilitarian calculation, for it would make no sense to arbitrate between preferences without the end of maximising preference satisfaction overall. This characterisation of morality in terms of the satisfaction of personal wants creates the illusion that universalizability creates utilitarianism, but is simply Hare’s conception of what morality is that produces his utilitarian reasoning.

Hare concentrates on what Kant termed material principles, inclination strength of desire etc. As a result he is at the mercy of these factors. Conversely, Kant’s moral law states that these material principles are subjective and are adopted as effects of his action and in every case only relative. Hare has substituted material, consequential principles as a replacement for formal principles. The Kantian position posits that we are moral beings only because we subordinate our inclinations and desires to rationality and when an agent acts morally he should do precisely this, otherwise he is not acting morally at all. A final quote from Kant will hopefully clarify my point:

“Practical principles are formal if they abstract from all subjective ends; they are material, on the other hand, if they are based on such ends and consequently on certain impulsions. Ends that a rational being adopts arbitrarily as effects of his action (material ends) are in every case only relative; for it is solely their relation to special characteristics in the subjects power of appetition which gives them their value. Hence this value can provide no universal principles.
[...] Suppose, however, there were something whose existence has in itself an absolute value, something which as an end in itself could be a ground of determinate laws; then in it, and in it alone, would there be the ground of a possible categorical imperative. Now I say that man, and in general every rational being exists as an end in himself. [...] All objects of inclination have only a conditioned value.”

In sum, Hare’s criteria for moral argument are material whereas Kant’s are strictly formal. Moreover, it would appear that in Could Kant Have Been a Utilitarian? Hare claims that Kant’s reasoning is compatible with utilitarianism but it is clear that the utilitarianism Hare advocates can sanction using rational beings as means as opposed to ends in themselves; whereas Kant’s moral law could never allow this.

---

Conclusion

“It is much easier to point out the faults and errors in the work of a great mind than to give a clear and complete exposition of its value.”

Schopenhauer’s statement resonates strongly, particularly as it was indeed difficult to locate aspects of universalizability that outweighed its drawbacks, despite it being a profound and extensive thesis. In light of this, one hopes the reader does not feel that this paper has been unduly critical of Hare’s work. He was without doubt an extremely gifted philosopher who contributed a great deal to our understanding of secular ethics, his subject matter and writing style have both challenged and entertained me. It is therefore unfortunate that a lot of his energy, particularly in his later years, was expended on attempting to reconcile the work of Kant with a certain kind of utilitarianism. I would suggest that this developed because Kant is seen to be at the opposite end of the intellectual spectrum to utilitarianism, if Hare could demonstrate the similarities between universalizability and Kant’s moral law, then he may have silenced a lot of anti-utilitarians. It also arose due to the fact that Hare had to make more and more complex utilitarian arguments to defend his theory; showing that Kant was conceptually similar in his work on morality would have been somewhat of a vindication for Hare. Alas, this did not happen, because as I have argued, Kant’s reasoning is fundamentally incompatible with Hare’s utilitarian reasoning.

Hare attempted to give a different account of how one is to universalize moral principles to Kant, one that was solely reliant on logic, presumably because he did not see rational beings as ends and therefore he sought a more reliable criterion upon which to base his universalizability. The thesis itself suggested that the logic of moral language acted as a kind of imperative, but logic alone simply allowed one to see when language deviated from set rules. Importantly, when the thesis was confronted with an agent who advocated a wholly irrational position i.e. the fanatic, it broke down. It would seem to follow then that rationality is of more importance than the logic of language; for one can use language logically and consistently in advocating an irrational position. Furthermore, as Hare’s examples showed us, it would seem that irrationality begets immorality precisely because rationality is a requirement of being a moral agent.

'Universalizing', as opposed to Hare’s universalizability, needs to be a key feature of any cogent moral theory. For if there is morality at all, it would necessarily be applicable to all; as such a credible ethical doctrine has to, as far as possible, demonstrate the universal implications of making moral judgements. Otherwise people would be treated, as they far too often are today, as somehow outside of our moral sphere. Without universalizing how else could we avoid the descent into some confusion of community based relativism? Hare’s instantiation of universalizing however is found wanting because he claims that the usage of moral words, constrained as they are by logic, necessitate rationality, which they do not. Furthermore, he adopts purely ‘material’ principles as ‘formal’ principles in defence of his theory. One cannot therefore see the benefits of adopting Hare’s universalizability as opposed to Kant’s moral law when attempting to both undertake theoretical ethics and for our understanding of how we are to treat others.
Appendices

(From page 10)

“A owes money to B and B owes money to C. It is the law that creditors may exact their debts by putting their debtors into prison. B asks himself, “Can I say that I ought to take this measure against A in order to make him pay? He is no doubt inclined to do it or wants to do it. Therefore, if there were no question of universalizing prescriptions, he would assent to the singular prescription let me put A into prison. But when he seeks to turn this prescription into a moral judgement, and say, “I ought to put A into prison because he will not pay what he owes,” he reflects that this would involve accepting the principle “Anyone who is in my position ought to put his debtor in prison if he does not pay.” But he then reflects that C is in the same position of unpaid creditor with regard to himself (B) and that the cases are otherwise identical; and that if anyone in this position ought to put his debtors into prison, then so C ought to put him (B) into prison. And to accept the moral prescription ‘C ought to put me into prison’ would commit him (since we have seen he must be using ‘ought’ prescriptively) to accepting the singular prescription ‘Let C put me into prison’ and this he is not ready to accept and can [therefore not accept that he ought to put A into prison.”

Bibliography


Fullinwinder, Robert K., ‘Fanaticism and Hare’s Moral Theory’, *Ethics*, Vol. 87, No. 2 (Jan., 1977)


Honderich, Ted & Burnyeat, Miles (ed.), *Philosophy As It Is* (Middlesex: Pelican Books 1979)


The British and Foreign Bible Society, *King James Bible*, London, 1945