Hare on Possible People

STEPHEN M. CAMPBELL

ABSTRACT   *R. M. Hare claims that we have duties to take the preferences of possible people into consideration in moral thinking and that it can harm a merely possible person to have been denied existence. This essay has three parts. First, I attempt to show how Hare's universalizability argument for our obligations to possible people may fail to challenge the consistent proponent of the actuality restriction on moral consideration, regardless of whether this proponent is construed as an amoralist or a fanatic. Second, I raise some objections to Hare's claim that a merely possible person can be harmed. Even if Hare could successfully overcome the objection that a possible person cannot be the recipient of harm, he would still need to show that this harm is morally significant. Third, whether or not Hare is able to answer these objections, I indicate how his moral theory still supports his general position on possible people — namely, that we are ceteris paribus morally bound to bring happy people (and avoid bringing miserable people) into existence.*

Most people find it intuitively plausible that we have no obligations to bring people into existence — at least not any obligations to *them*. They would endorse what I will call the *actuality restriction on moral consideration* (hereafter abbreviated AR):[1]

(AR) Only actual people warrant moral consideration,

where 'actual people' is meant to include present, future, and perhaps past people.[2] According to AR, a person must come into our world at some time in order to be a candidate for moral consideration. I may certainly have obligations to my *future* child on this view, but I cannot have duties to a possible child unless her futurity is decided or expected. A merely possible person (i.e. a person who might exist but in fact never will, or who might have existed but never did) never comes within the bounds of moral consideration.[3]

R. M. Hare thinks otherwise, claiming that moral deliberation should take into account the preferences of possible people and, more controversially, that it can harm a merely possible person to have been denied existence.[4] It has been remarked that Hare is probably the paradigm defender of this sort of view and may be the inventor of the only genuine *argument* for it.[5] In Section I, I analyze Hare's argument concerning our alleged obligations to possible people and show how the argument can fail to challenge a consistent proponent of AR. In Section II, I raise some objections to Hare's claim that a merely possible person can be harmed. Even if these objections cannot be answered, Hare's moral theory nevertheless supports his general position on possible people. In Section III, I discuss how Hare's conception of moral thinking grounds an obligation, *ceteris paribus*, to bring happy people into the world.

I. Hare's Argument for Our Obligations to Possible People

A. The Argument

Hare has offered different formulations of his argument. One makes use of the Golden Rule, with some slight modifications. The modifications proceed as follows:

> Golden Rule: We should do to others as we wish them to do to us.
> Golden Rule$_1$:   We should do to others as we wish they had done to us.
> Golden Rule$_2$:   We should do to others what we are glad was done to us.[6]

Provided that one is willing to accept the Golden Rule and its variants, the argument is this: if you are glad that someone brought you into existence, then you should do the same for others. Most people who lead happy lives are glad that they were conceived, so they should, *ceteris paribus*, render this same benefit to others when it is in their power to do so.

A second formulation, which will be my focus in this essay, is cast in terms of the universalizability of moral judgments. Suppose that you ask a happy person to make a moral assessment of 'the actions of those who brought him into existence by begetting, conceiving and not aborting him'.[7] If he replies that his parents did what they morally ought to have done, then he has made a moral judgment. Hare can now show that, on pain of violating the logic of moral judgments, he must apply this judgment to all relevantly similar situations.[8] Unlike the Golden Rule version which makes reference to 'others',[9] this argument has the virtue of taking attention from possible people and focusing more on *actions*. If I give a positive moral appraisal of my having been conceived (no matter how my pre-existent state is characterized), I cannot consistently give a different appraisal of the relevantly similar situation of someone else's being conceived.

Hare's argument works quite well when dealing with one who would inconsistently treat his own conception as morally good and the conception of others as morally neutral. But the person who answers that his parents did as they morally ought to have done may already concede the very point in dispute — namely, that a possible person warrants moral consideration.[10] This is precisely what AR denies. If Hare's argument is to have any weight against those who consistently deny that the interests of a possible person can ground an obligation to procreate, neither he nor his hypothetical respondent can assume this point. The argument must succeed when a person gives the amoralist response.

B. The Amoralist Response

The following is based upon Hare's treatment of the amoralist response in his 1998 article 'Preferences of Possible People'. Suppose that we ask a person to give a moral assessment of her parents' act of conceiving her, and she instead says, 'What they did was neither wrong nor obligatory (i.e. it was neither the case that they did what they morally ought, nor the case that they did what they morally ought not)'.[11] And let us further suppose that this individual is nevertheless glad that she was born. This means that she prefers that her parents did what they did, though she does not make a moral judgment about it.[12] Lastly, suppose that, despite her preference for her own conception, she 'has no preference as to whether the other similar people in similar situations should be brought into existence or not'.[13]

This would be a case of amoralism. The *amoralist*, Hare tells us, is one who either refrains from making moral judgments at all or makes judgments of moral indifference, and does this either always or in connection with particular cases.[14] Hare seems to think that the amoralist who opts out of morality selectively, rather than being a total amoralist, will be motivated by self-interest.[15] We can discover this by looking at her preferences. If our respondent favours her own conception but not that of others only because the former case involves *her*, she will be an amoral egoist with respect to this sort of case. And if *her* involvement (as opposed to someone else's) can count as a relevant difference, then she must, if consistent, take a slide down the slippery slope and become an amoral egoist in many other cases. This does nothing to challenge amoral egoism. But if I get Hare right, the force of his argument hinges on the fact that most people will not want to be full amoral egoists. So by pressing for consistency 'in embarrassingly many other cases', we may shame them into abandoning their amoralism altogether.[16]

I am not satisfied with Hare's treatment of the amoralist attitude towards one's own conception.[17] The problem is that amoralism, as Hare defines it, could but need not involve amoral egoism. Say that you give the amoralist answer about your own conception, though you nevertheless prefer your having been born to your having never existed. If you do not have the same preference for the birth of others (including all of those people who might have been born), Hare thinks this will force amoral egoism in other cases. But it is by no means evident that we have an instance of amoral egoism in *this* case. I presume that the amoral egoist is one whose amoralism *is motivated by* his egoism, but amoralism can surely be grounded in other ways.

To see this point, consider Hare's questions applied to some act that most of us would deem supererogatory. Suppose that I come out of my office building to find someone handing out free ice cream sandwiches. As I walk across campus enjoying my ice cream sandwich, I reflect on how grateful I am that this kind lady was dispensing free ice cream and that I was one of the lucky ones who came outside at the right time. If Hare asks me if she did what she morally ought to have done, I will say that it was not morally required of her, and that it would not have been wrong of her to sit at home and eat all of them herself. I would therefore give an amoralist response about this sort of case, due to certain beliefs I hold about private property and supererogation. Now, am I glad that she gave me an ice cream sandwich? Certainly! Do I wish that everyone else on campus could have received an ice cream sandwich like we lucky ones did? Not especially. Nor do I have a preference that similar people in similar situations should get ice cream sandwiches. After all, I do not have as strong of a reason to favour the interaction of ice cream with *their* taste buds. But I fail to see what bearing these preferences would have on my amoralism about the dispensation of free ice cream sandwiches. Had I been unlucky, I would have held the same position about the moral status of the act. And even if I egoistically prefer that *I* (as opposed to others) should be a lucky one, this preference is beside the point — it neither grounds nor affects my moral assessment of the situation. In short, it seems that one can be an amoralist about a certain type of case and have egoistic preferences relating to that case without being an amoral egoist.

C. Fanaticism

There is another possibility that should be mentioned. Perhaps the AR-proponent, behind the philosophical rationale, *is* motivated by egoism, even if she herself is

unaware of it. The endorsement of AR has structural similarities to racism, sexism, and speciesism; members of some in-group declare some out-group unworthy of moral consideration. We could be dealing with a new sort of discrimination — call it 'modalism'. Hare views the discriminatory -isms as forms of fanaticism. A *fanatic* is one who holds some ideal, believing that it should be followed even if people's interests are harmed in the process.[18] AR might be linked to some actuality-restricted moral principle — something like 'Prevent (or at least minimize) the harm done to *actual* people (or sentients)'. Just as a racist might work under a maxim that urges the promotion of his own race's interests, this principle might qualify as a fanatical moral ideal since the interests of possible people would be completely ignored in favour of the welfare of actual people. Yet if the AR-proponent is a fanatic and can sincerely endorse an actuality-restricted moral principle even when faced with the prospect that she herself might have been denied existence, Hare (by his own admission) has no argument against her, just as he is unable to defeat the hard-core Nazi fanatic by argument.[19] This is because the hard-core fanatic, even when imagining himself in the shoes of those who have their interests violated, will stick to his principle. And the amoralist response, as we have already noted, implies that one is willing to swallow the possibility that she might have been denied existence.[20]

It might still be doubted whether one can ever *sincerely* affirm that it would have been of no moral consequence if she herself had never been born. To address this doubt, the best I can do is to mention a few cases where I think one's sincere beliefs would necessitate such an affirmation. First, a person may believe that possible people (a class of which she was once a member) lack features requisite for moral consideration — e.g. a central nervous system, consciousness, etc. This is not unlike a person believing that dead people cannot be harmed and, in turn, refusing to believe that she herself might be harmed after her own death. Second, one could make a sincere affirmation if one believes that bringing a being into existence is a supererogatory act. I can certainly grant that no wrong would have been done if I had not been given a free ice cream sandwich, and many believers would claim that God would not have done wrong by failing to create our universe. It does not matter how deeply we have benefited from an act; if we consider the act to be supererogatory, then its omission is not wrong. To deny that a sincere affirmation about the moral indifference of one's own birth could be made in these cases is to doubt that people can sincerely hold the sorts of beliefs I have sketched. For my part, I see no reason to doubt this.

II. Harm to Merely Possible People

What I take to be Hare's most controversial claim on the subject of possible people — call it the *harm claim* — is that a merely possible person can be harmed by not being brought into existence.[21] This invites what I will refer to as the *non-actuality objection* (NA):

> (NA) A merely possible person, being non-actual, cannot be harmed in a morally significant way.

While most of us probably find NA intuitively compelling, Hare has a very interesting way of generating the claim that a merely possible person can be harmed in a morally significant way. His theory of moral reasoning is future-directed. It does not simply take

preferences into account but the *frustrations* and *satisfactions* of preferences that result from a person's chosen action.[22] So rather than assessing the situation at the time of the decision, Hare's method of moral thinking involves surveying the morally relevant outcomes of one's choice options. Consider the following two choice situations, S1 and S2:

	A: Procreate	B: Don't procreate
S1	a happy person's preference to exist is satisfied	—
S2	a miserable person's preference to not exist is frustrated	—

Figure 1.

Let us assume that the preference-satisfactions and -frustrations of all other people in the world will be the same whether A or B is chosen in any given situation S; therefore, all that is morally at stake in the decision is shown in the outcome boxes. The decision to procreate can result in the birth of a person with a satisfied existence-preference (in S1) or a person with a frustrated existence-preference (in S2). Hare's method lends itself to the claim that a person can be benefited or harmed by being brought into existence, a view defended by Derek Parfit in Appendix G of his 1984 *Reasons and Persons*.[23] But Hare wants to go further than Parfit and claim that a possible person can be harmed on outcome S1B.[24]

At this point, it may be objected that Hare's system of moral thinking cannot accommodate this sort of harm: any harm will need to be reducible to preference-frustrations, but the frustration of the preference in S1A will carry no moral weight. In his 1990 article 'External and Now-For-Then Preferences in Hare's Theory', Mane Hajdin argues that, by its very nature, Hare's method of moral thinking excludes certain types of preferences. Hajdin writes:

> The requirement that in moral thinking we imagine what it is like for other people to have their preferences frustrated (and not merely what it is like for them to *have* them) therefore quite naturally, without any ad hoc moves, precludes now-for-then, as well as external preferences from having any real effect on the results of our moral reasoning.[25]

In his 1998 article 'Preferences of Possible People', Hare concedes Hajdin's point — namely, that his theory of moral reasoning can only take into account experiential now-for-now and then-for-then preferences.[26] Hare thus rules out consideration of *non-experiential* or *external* preferences, which he describes as 'preferences for things other than experiences of the preferrer'.[27] While I suspect that Hare's definition of non-experiential preferences may be ill-chosen,[28] the point that Hajdin establishes is clear: Hare's theory can only give moral weight to *experienced* preference-satisfactions and -frustrations. Because Hare concedes this point, he would seem to face a devastating objection to his claim that merely possible people can be harmed. It is conceptually impossible for one to *experience* the frustration of her preference to exist or her preference to have been conceived. Therefore, it seems that there cannot be morally significant harm in S1B without challenging the structure of Hare's own moral theory.

But Hare has a response. Granted, the harm done to merely possible people does not involve the experience of preference-frustrations, but it does involve the loss of experienced preference-satisfactions.[29] He thinks the harm just consists in the loss of the benefit of preference-satisfactions in S1A. He writes:

> True, [a possible person] does not exist to be harmed; and he is not deprived of existence, in the sense of having it taken away from him, though he is denied it. But if it would have been a good for him to exist (because this made possible the goods that, once he existed, he was able to enjoy), surely it was a harm to him not to exist, and so not to be able to enjoy these goods. He did not suffer; but there were enjoyments he could have had and did not.[30]

Hare thinks that 'to avoid a harm *is* to confer a benefit'[31] and, based on the passage above, to deny a benefit is to render a harm. So he endorses the following interpretation of S1:

	A: Procreate	B: Don't procreate
S1′	an actual person is benefited (in terms of preference-satisfactions)	a merely possible person is harmed (in the loss of S1′A preference-satisfactions)

Figure 2.

By ruling out non-experiential preferences, Hare's moral thinking only ignores preferences where one's experience will be completely unaffected regardless of whether they are satisfied or frustrated. For example, Cheops' desire for a big funeral cannot make any difference to moral deliberation since there can be no experienced preference-satisfaction or -frustration for the preferrer.[32] But suppose that you are secretly robbed of an inheritance left to you by some unknown relative. In ordinary usage, it seems right to say that someone has harmed you. It is true that this act of robbing you will not produce experienced preference-frustrations, but had the thief let you receive your inheritance, you certainly would have had the experience of preference-satisfactions. According to Hare's moral theory and our own moral intuitions, harm is not always reducible to experienced preference-frustrations.[33]

The NA-objector could press her case here, arguing that Hare still cannot defend the claim that a merely possible person can be harmed in a morally significant way. I will briefly consider two lines of attack. The first line of attack involves the claim that a merely possible person cannot be harmed at all. Harm must have a recipient, but the NA-objector can plausibly deny that a merely possible person is a *type* of person. Just as Epicurus claimed that 'when death arrives, we are not',[34] it might be claimed that, before conception, we are not. Being always non-actual, a merely possible person is, in a sense, no-thing. There might be some support for this conclusion in Hare's (1981) *Moral Thinking*. Leery of endorsing an ontology of *possibilia*, he claims that possible situations and possible worlds 'exist only as objects of thought'.[35] If Hare's espoused view on possible situations and worlds also applies to possible people, then a possible person is a mere thought-object. In that case, NA gains plausibility as an objection since it is not at all clear how a mere thought-object can be the recipient of real harm.

A second line of attack allows that a merely possible person can be harmed (in the way Hare specifies) but denies that it is morally significant harm. Harm, it can be

claimed, is only morally significant when it is done to an entity that has moral standing. Consider the possible outcomes of my decision about whether or not to rob you of your inheritance. On one outcome, you will happily enjoy your status as a beneficiary; on the other outcome, you will live out your life oblivious to the fact that harm has been done to you. The NA-objector can point out that, on each of these options, we are faced with an actual person who deserves moral consideration. This is not so in existence-decisions.[36] While there is an actual person with preferences on both S1′A and S1′A, the B-options (if they have anything) have a non-existent who lacks consciousness and sentience and thus is unable to have concerns or preferences.[37] So even granting that a merely possible person is harmed in the sense of being denied existence, the NA-objector can claim that it will not be morally significant harm since a merely possible person lacks the properties requisite for moral consideration.

I believe that these two lines of attack raise problems for Hare's claim that a merely possible person can be harmed in a morally significant way. If his claim is to be defended, it first needs to be established that a merely possible person can be the recipient of harm and then shown that such harm is morally significant.

III. Hare's Moral Theory and Our Obligation to Make Happy People

Even if Hare's universalizability argument fails to challenge many proponents of AR and even if the harm claim cannot be adequately defended, this is no threat to the defensibility of his general position on our obligation to bring happy people into the world.[38] In this section, I will show how his moral theory as presented in his *Moral Thinking* (1981), coupled with the developments in his essay on possible people (1998), yields this same result.

A. The Critical Level

Hare develops a utilitarian theory with two levels of moral thinking: the critical and the intuitive. The *critical level* is the level at which we would always operate if, like archangels, we had 'superhuman powers of thought, superhuman knowledge and no human weaknesses'.[39] An archangel would be a perfect act-utilitarian — or, what comes to the same thing, a perfect 'specific rule'-utilitarian whose rules are allowed to be of unlimited specificity.[40] To the often considerable extent that we fall short of the archangelic ideal, we need to work at the *intuitive level* of moral thinking, following *prima facie* principles that are selected at the critical level. Presumably, we will adhere to intuitive level principles (e.g. 'Don't tell lies') unless we find some reason to diverge and feel confident that we have adequate information for doing critical thinking. Nonetheless, Hare thinks that the critical level is primary while the intuitive level is just a practically necessary subordinate.[41] Further, while intuitive level thinking is crucial to our practical moral lives due to the weaknesses and epistemic limitations that characterize the human condition, these factors do not come into play in the consideration of Figures 1 and 2 where we have archangelic knowledge of the possibilities. If it is stipulated that an additional happy person could be born without affecting, in a morally relevant way, the preference-satisfaction of anyone else in the world, we seek to know whether we have a moral reason to bring this person into existence. Our concern, therefore, is with the critical level since it tells us the morally best thing to do, all things considered.

To see that Hare's moral theory yields the same results as he sought to get from his harm claim, I wish to highlight two features of critical level thinking: first, it must take into account possible outcomes, and second, it does not recognize a class of supererogatory acts. These features follow from the nature of critical level thinking, which involves, effectively, putting oneself in the shoes of all those who are affected by some action and making an impartial assessment. As Hare tells us in a summarization,

> the method of critical thinking which is imposed on us by the logical proper-
> ties of the moral concepts requires us to pay attention to the satisfaction of the
> preferences of people . . . and to pay attention equally to the equal preferences
> of all those affected (because moral principles have to be universal and there-
> fore cannot pick out individuals).[42]

The upshot is this: what ought to be done in a given choice situation is fixed by the preferences (with their respective strengths) of the people affected by the action. We can look at one of Hare's examples to draw out the first feature mentioned above.

Hare imagines that he wants to move someone's bicycle in order to park his car, where Hare's preference to move the bicycle is stronger than the bicycle owner's mild preference not to have his bike moved. Both parties, if they have properly engaged in critical thinking, will conclude that 'the bicycle ought to be moved'.[43] Given Hare's 1998 concession that he can only take into account now-for-now and then-for-then preferences, and given that critical moral thinking involves envisioning the satisfaction and frustration of these preferences, it is clear that Hare's moral theory must look to possible outcomes. We cannot arrive at a moral evaluation simply by looking at the situation in which the bicycle gets moved since that only represents half of the story. Consider situations S3 and S4, where Biff is the owner of the bicycle:

	A: Move the bicycle	B: Don't move the bicycle
S3	Hare is pleased. Biff is slightly annoyed.	Hare is extremely annoyed. Biff is pleased.
S4	Hare is pleased. Biff is slightly annoyed.	Hare is slightly annoyed. Biff is blissfully happy.

Figure 3.

If we exercise impartiality between the sets of experiential synchronic preferences in the A-B options, it seems clear that in S3 the bicycle ought to be moved but that in S4 it should not be moved. My purpose in this figure is not to give a precise rendering of how Hare's treatment of multilateral situations would be applied to possible outcomes. Rather, I only intend to illustrate that knowing what it is like for the bicycle owner to experience having his bike moved and what it is like for Hare to get the parking space (which I treat as identical in S3A and S4A) cannot, in itself, settle the question of whether the bicycle ought to be moved. We must take all of the possible outcomes into consideration.

Moving now to the second feature: If we take a supererogatory act to be an action whose performance is morally good but whose omission is not morally wrong, then supererogation does not exist at the critical level.[44] Critical level thinking is simply not equipped to

make the distinction between the obligatory and the supererogatory. The critical thinker identifies with the preferences of all affected parties, on each possible outcome, developing her own preferences as to what should happen to her were she to find herself in their shoes. The prescription she accepts about what ought to be done will coincide with her all-things-considered preference among the outcomes. An outcome with a given set of preference-satisfactions (e.g. people enjoying ice cream sandwiches) will be preferred over an outcome that is equal in all respects except for the absence of that set.[45] So by giving weight to all of the experiential synchronic preferences affected by the action in question, the critical level tells us what *ought* to be done, all things considered.

Utilitarian theories are sometimes criticized for being too stringent since they are commonly thought to make no room for the supererogatory. Hare's solution is to work supererogation (or at least something resembling it) into the intuitive level.[46] But at the critical level, one *ought* to choose that possible outcome which maximizes preference-satisfactions and minimizes preference-frustrations; one *ought not* choose any other option.[47] With this fact in mind, we are now in position to see how Hare's critical thinking leads to a *ceteris paribus* obligation to create happy people.

B. Some Mandates of the Critical Level

We can start with a typical moral decision, like Hare's bicycle/car example, that does not involve people coming into or going out of existence. In S3, Hare and Biff exist on both the A- and B-option, and they also exist prior to the execution of the decision. But this latter fact plays no role in deciding what to do and is therefore beside the point in critical thinking. This is because critical thinking only looks at the satisfaction and frustration of experiential synchronic preferences. If this is correct, then Hare's critical thinking involves what we might call a 'tunnel-visioned consequentialism' where only certain features of the possible outcomes of a given choice (and not the pre-choice situation) are relevant to deciding what ought to be done. In S3, the archangel would recommend that the bicycle be moved just in virtue of the comparison between S3A and S3B.

Now let us move to cases where the existence of the affected individuals is not constant. Consider situations S5 through S10:
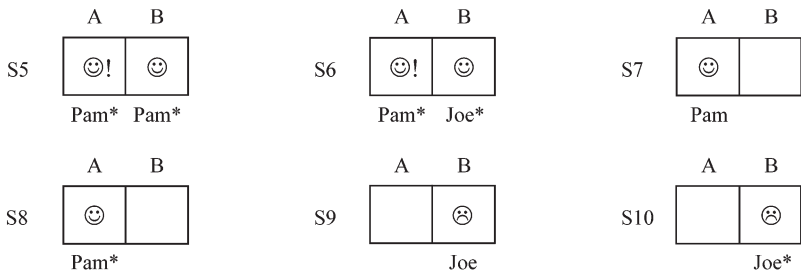
Key:
☺   Individual with a life worth living
☹   Individual with a life not worth living
☺!   Individual with a life more worth living than the life on the other alternative
*   Does not exist prior to the execution of the decision

Figure 4.

A few qualifications concerning Figure 4 are in order: (1) The decision in each situation S is between bringing about either outcome A or outcome B; (2) As simplifying assumptions, we will say that, in any S, A and B represent the only possibilities, and all morally relevant differences between A and B are captured in the diagrams; (3) A life worth living or not worth living is to be understood in terms of experiential synchronic preference-satisfactions and -frustrations; (4) The information listed below the outcome boxes will play no role in the critical level decision about which option to select.

Concerning this fourth point, it will make no difference whether or not the person exists prior to the agent's acting on A or B since Hare's theory only looks to the satisfaction and frustration of experiential *synchronic* preferences (see note 26). Further, the identities of the affected individuals will not be relevant because 'individual references are excluded [from moral principles] by the requirement of universality'.[48] (These facts provide some further reasons why the critical level is not equipped to handle some standard cases of supererogation.)[49]

In Figure 4, critical thinking mandates that we bring about outcome A in each situation S5-S10. I will consider the cases in order:

S5: Pam is a future person since she will exist no matter what we choose. We can make her life better or worse, but in either case it will be a life worth living. Critical thinking will bid us to do A, maximizing the quality of Pam's life.

S6: Pam and Joe are both possible people since, for each of them, there is a possible outcome on which s/he will not exist. As in S5, critical thinking will direct us to choose A and bring into existence Pam, whose life will be more worth living.[50]

S7: Pam is an actual person who now has a life worth living. Let us say that the decision concerns whether or not to terminate her existence; we could kill her painlessly while she is sleeping. Critical thinking will select A, instructing us to refrain from killing her. Needless to say, it would be a major problem for any ethical theory if it treated the secretive, painless killing of happy people as a matter of moral indifference. (The decision in question could also have concerned whether we should take active steps to prevent Pam from dying.)

S8: Pam is a possible person who will have a life worth living if she is brought into existence. Given our simplifying assumption, it will not affect the quality of life of existing people to bring this new person into existence. Critical thinking, which must give an identical treatment to S7 and S8, will mandate that we choose A and so bring about this happy person.

S9: Joe is an actual person who now has a life not worth living, perhaps due to some debilitating and painful medical condition. Critical thinking instructs us to choose A, killing, or allowing to die naturally, anyone whose life is not worth living. This conclusion is not quite as scary as it first appears. Presumably, we are talking about a 'miserable person' (as in Figure 1) who has no hope of having a life worth living. If Joe's life was only *barely* not worth living, then we would surely have a third outcome, C, at our disposal on which we find a way to raise Joe's quality of life. If we could possibly help him to have a life worth living, the critical level would then endorse C.

S10: Joe is a possible person who will have a life not worth living if he is brought into existence. As in S9, critical thinking will mandate that we choose A, preventing the birth of this miserable person.

So, to summarize some of the *ceteris paribus* mandates of the critical level:

S5:   Maximize the quality of life that future people will have.
S6:   Given a choice between two sets of possible people, actualize the one that will have a higher quality of life.
S7:   Keep in existence people whose lives are worth living.
S8:   Create people whose lives would be worth living.
S9:   Send out of existence people whose lives are not worth living.
S10: Do not create people whose lives would not be worth living.

The critical level of moral thinking clearly mandates that, *ceteris paribus*, we procreate in S1 and we not procreate in S2. Hare's claim about harm to merely possible people is unnecessary to establish this conclusion since he can do so either by endorsing Parfit's more modest conclusion that a person can be benefited in S1A or harmed in S2A by being brought into existence, or simply by selecting the best outcome in each situation.

IV. Concluding Remarks

I have set out to show three things. First, Hare's universalizability argument for our obligations to possible people has no force against those who would endorse AR consistently and with good, non-egoistic reasons. In his 1998 essay, Hare takes seriously the fact that one might give the amoralist response to his argument, but I claim that his treatment of this response rests on a distorted picture of amoralism and its relation to egoism. And in the event that the AR-proponent is a consistent fanatic, Hare admits that he is not able to defeat her by argument.

Second, Hare's claim about harm to merely possible people is open to at least two lines of attack from the non-actuality objection. Whether or not the Harean *could* overcome these objections, we should first ask whether the battle is worth fighting at all. Indeed, why did Hare defend this curious thesis in the first place? There are some places where Hare espouses the view that utilitarians cannot show wrong without showing harm,[51] though he seems less committed to it elsewhere (see note 47). His defence of the harm claim can be seen as an attempt to bring the results of an impersonal total utilitarianism, which follow from his critical level thinking, in line with the intuition behind person-affecting views (see note 1). Hare attempts to straddle these two positions in consequentialist population ethics by expanding the *person*-affecting view to include all *possible persons* (see note 2 [3]). Of course, in making this move, Hare loses the intuitive appeal of the person-affecting intuition, which is captured by AR.

Third, in light of the concessions that Hare makes in the 1998 essay, I have tried to show why his critical level already yields the results that he would get if the harm claim had succeeded. The tunnel-visioned consequentialism of critical thinking is able to satisfy many of our moral intuitions: we should not kill happy people; we have a moral reason not to bring miserable people into existence; we should not adopt a policy of depletion (see note 50). Yet this tunnel vision is not without its costs. Hare's theory seems unable to take into account certain features of situations that, for many of us, are morally significant — for instance, the state of the world prior to the execution of the decision, and the identities of the affected individuals across outcomes. Some

would find the view that we have an obligation to procreate to be an effective *reductio* of Hare's position. Yet, in fairness to Hare, all views in consequentialist population ethics seem to have their unwelcome or counterintuitive implications. Hare's two-level utilitarianism may have a special edge since his intuitive level allows him to soften some of the repugnant conclusions required by critical level thinking.[52]

In any case, the general goal of the arguments given in this essay has been to shift the focus away from the universalizability argument and harm claim and toward Hare's moral theory. The harm claim is open to attack from NA, but Hare can simply rely on his critical thinking to achieve the same results. Likewise, in Section I, I suggested that a person might have good reasons for her consistent endorsement of AR. But the reasons that were proposed (that conferring existence is a supererogatory act, and that possible people lack features requisite for moral consideration) are both challenged by Hare's moral theory since, at the critical level, there is no supererogation and the preferences of possible people are given as much consideration as those of future people and of present people who will still be around on the relevant outcomes. Thus, when faced with Hare's proposal that we have an obligation to make happy people, I think we are ultimately pushed to ask: must moral thinking be as Hare proposes? I have staked no claim on that issue here.

Stephen M. Campbell, Department of Philosophy, The University of Michigan, 2215 Angell Hall, 435 South State Street, Ann Arbor, MI 48109-1003, United States. cmpbell@umich.edu

NOTES

1 This restriction is closely related to the person-affecting intuition, which Melinda Roberts describes as the 'idea that a choice cannot be bad unless it affects *some person or another* badly, and cannot be morally wrong unless it wrongs *some person*'. M. Roberts, 'Is the person-affecting intuition paradoxical?', *Theory and Decision* 55 (2003): 1–44 at p. 1. However, AR is meant to emphasize what is only implicit in the person-affecting intuition — that 'person' means an *actual* person.

2 Four notes of clarification: (1) For simplicity's sake, I am speaking only of people, but AR can of course be extended to other actual sentients. (2) I will not address the question of whether and when a foetus or infant becomes a 'person' or 'human' in the richer cognitive sense. This essay focuses on the case of unconceived possible people, though it certainly has implications concerning conceived potential people. (For more on the latter issue, see note 7.) (3) Strictly speaking, we can think of the class of possible people as being comprised of two subclasses: *actual people* (possible people who — in the past, present, or future — become actual) and *merely possible people* (possible people who could have, could now, or could in the future become actual but do not). This follows the breakdown given in R. Hare (1988a), 'Possible people', reprinted in *Essays on Bioethics* (New York: Oxford University Press, 1993), pp. 67–83 at pp. 68–69.

However, for my purposes here, I will generally use the designation 'possible person' to imply that it is an open question as to whether or not a person will be actual or just merely possible. The phrases 'actual person', 'future person', and 'merely possible person' will indicate that the question of whether those persons have existed or will exist is closed. (4) It is worth noting that our duty 'to give equal weight to the equal preferences of all possible people, whether they are actual or merely possible' (*ibid.*, p. 70) will not necessarily lead to an obligation to actualize a possible person. Indeed, under certain conditions, one can have an obligation *not* to procreate. What I will sometimes refer to loosely as an 'obligation to procreate' is attended by a crucial *ceteris paribus* clause. Assuming that a possible person would lead a happy life, that it is in my power to bring this person into existence, and that it will not create a greater cost of unhappiness to others (including other possible people), Hare thinks that I ought to bring this person into existence. Cases where this obligation is present will be our primary concern here.

3   In counterfactual moral reasoning about what our obligations would have been if things had gone differently, we might of course be talking about obligations to people who would in that case be actual but who are not 'actually' actual.

4   Hare expresses his views on potential/possible people in R. Hare (1974) 'The abnormal child: moral dilemmas of doctors and parents', reprinted in *Essays on Bioethics* (New York: Oxford University Press, 1993), pp. 185–91; R. Hare (1975) 'Abortion and the Golden Rule', reprinted in *Essays on Bioethics* (New York: Oxford University Press, 1993), pp. 147–67; Hare (1988a); R. Hare (1988b) 'When does potentiality count?', reprinted in *Essays on Bioethics* (New York: Oxford University Press, 1993), pp. 84–97; R. Hare (1989a) 'A Kantian approach to abortion', reprinted in *Essays on Bioethics* (New York: Oxford University Press, 1993), pp. 168–84; and R. Hare (1998) 'Preferences of possible people', in Frehige and Wessels (eds.) *Preferences* (Berlin: de Gruyter, 1998), pp. 399–405.

5   C. Frehige and U. Wessels, 'Introduction to possible preferences', in Frehige and Wessels op. cit., p. 379.

6   Hare (1975) op. cit., p. 153.

7   Hare (1998) op. cit., p. 401. This grouping is by no means insignificant. Since Hare is using the richer cognitive notion of 'person' when speaking of a possible person [see Hare (1975) op. cit., p. 152], he draws no moral distinction between, on the one hand, the person who could develop from a now existing foetus and, on the other hand, the person who could develop from a given sperm/egg combination (even if the sperm and egg in question do not yet exist). [For a vivid depiction of this point, see the foetus/ Andrew example in Hare (1974) op. cit.] As a result, Hare groups 'begetting, conceiving, and not aborting' together, and his Golden Rule/universalizability argument is intended to apply uniformly to abortion-decisions and procreation-decisions.

However, this grouping gives us skewed results if the respondent believes, *contra* Hare, that there is some significant moral difference between aborting and not conceiving. For the sake of clarity, I will distinguish and isolate the begetting/conceiving component from the not-aborting component of Hare's question. If I am right in thinking that most people accept AR, they will likely give the 'amoralist' response to the begetting and conceiving component (see Section 1.B). If, on the other hand, we ask people how their parents did in not aborting them, I expect that more people will answer that their parents did the morally obligatory thing. Are they inconsistently abandoning AR here? More likely, they believe that *they themselves* were actual during the foetal stage. They would therefore disagree with Hare that the foetus is not a person. AR allows flexibility as to what constitutes an actual 'person'.

For the purposes of this essay, I will concern myself only with unconceived possible people and the begetting/conceiving component of Hare's question, setting aside the sticky issue of whether 'not aborting' should be included in the grouping.

8   This thesis is defended in R. Hare, *Moral Thinking* (Oxford: Clarendon Press, 1981).

9   Some would question whether, or to what extent, a possible person qualifies as an 'other'. This point is raised in K. Chan, 'The Golden Rule and the potentiality principle: future persons and contingent interests', *Journal of Applied Philosophy* 21,1 (2004): 33–42, at p. 35. See also the Epicurean-style objection raised in Section II.

10 To decide this issue, we need to know *why* the respondent thinks that his parents did as they ought to have done. Presumably, they would have done wrong if they had not conceived him. What could ground the immorality of that act-omission? One possibility, which I am assuming here, is that it would have wronged or harmed *him* to deny him existence. As we shall see in Section II, Hare may have difficulty defending the view that a merely possible person can be harmed in a morally significant way. Given Hare's emphasis on harm to merely possible people (see note 21) and duties to merely possible people [see, e.g. Hare (1988a) op. cit.], we have good cause to think that he has this option in mind.

But there are at least two other possibilities. Here is one: The respondent's parents, by failing to create him, would somehow have wronged themselves or other actual people. I will set this possibility aside since it clearly falls short of what Hare seeks to establish, and AR generates no objection to this proposed ground of the immorality. Another possibility: In failing to create him, his parents would have failed to maximize the good in the world, and that is wrong. I will deal with this possibility in Section III, but it should be noted that some AR-proponents would reject this possibility along with the first.

11 Hare (1998) op. cit., p. 401.

12 For Hare, 'I prefer' is general enough to include 'I am glad that' and 'I desire that'. R. Hare (1989b) 'Brandt on fairness to happiness', *Social Theory and Practice* 15:1 (1989): 59–66, at p. 59.

13 Hare (1998) op. cit., p. 403.

14 Hare (1981) op. cit., p. 183.

15 See, for instance, Hare (1981) op. cit., pp. 185–86; Hare (1989a) op. cit., p. 178; Hare (1998) op. cit., p. 403.

16 Hare (1989a) op. cit., p. 178.

17 Hare says that he himself is 'not yet satisfied with this way of dealing with the amoralist position'. Hare (1998) op. cit., p. 403.

18 R. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963), pp. 105, 175. 'To have an ideal . . . is to think of some kind of thing as pre-eminently good within some larger class', *ibid.*, p. 159.

19 'If there are people so wedded to some fanatical ideal that they are able to imagine, in their full vividness, the sufferings of the persecuted, and who can still prescribe universally that this persecution should go on in the service of their ideals, even if it were they themselves who had to suffer thus, then they will remain unshaken by any argument that I have been able to discover', *ibid.*, p. 184.

20 Putting AR in the fanatic camp alongside the racist, sexist, and hard-core Nazi might seem an uncomfortable solution. But most of us will find the justificatory reasons offered for racist and sexist discrimination to be flimsy, if not grossly misinformed, and this lack of good reasons will therefore suggest to us that 'group egoism' may be what is really motivating those positions. But there can be good reasons for discrimination. Most everyone agrees that some things have no moral standing (e.g. the colour red, the present king of France, etc.), and a strong case can be made for the claim that actuality is requisite for moral consideration. Thus, even if the proponent of AR is a fanatic, we can certainly question whether 'modalism' is on par with the other discriminatory -isms.

21 See, e.g. Hare (1974) op. cit., p. 191; Hare (1975) op. cit., p. 166; and Hare (1998) op. cit., p. 402.

22 See M. Hajdin, External and now-for-then preferences in Hare's theory, *Dialogue* 29 (1990): 306–7. See also Hare (1981) op. cit., Ch. 5, and Hare (1998) op. cit., p. 400.

23 D. Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984).

24 In symmetrical fashion, Hare would argue that the possible miserable person is benefited on S2B by not being brought into existence. I will focus on the harm claim since Hare himself does, though Hare's 'benefit claim' should be no less controversial.

25 Hajdin op. cit., p. 308.

26 Hare (1998) op. cit., p. 404. Now-for-now and then-for-then preferences are *synchronic* preferences (i.e. preferences pertaining to a given time t that are also held at t). In contrast, a now-for-then preference, which is *asynchronic*, is a future-directed preference that one holds though it may not be held at the specified future time [Hare (1981) op. cit., p. 101 f., Hare (1998) op. cit., p. 399]. For example, I may now prefer that I should go to the dentist tomorrow though I will very likely lose the preference once I am suffering under the dentist's instruments.

27 *Ibid.*, p. 399. Following Dworkin, Hare uses 'external' preferences in Hare (1981) op. cit., p. 104 f.; he shifts to 'non-experiential' preferences in Hare (1989b) op. cit. and (1998) op. cit.

28 Many of our desires do not take our own experiences as their object but are instead desires for *states of the world*. For defence of this point, see H. Sidgwick, *The Methods of Ethics*, 7th edn. (Indianapolis, IN: Hackett, 1981), I.4.6, and J. Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Clarendon Press, 1986), Ch. 1 — esp. pp. 9–10, 13. There is a difference between my preference that the Braves win the World Series and my preference that *I see* the Braves win the World Series. The object of the latter preference is my own experience of a world-state (and so it is an experiential preference); the former is just for the world-state (a non-experiential preference).

Yet the frustration or satisfaction of a world-state preference can still fall within our experience in such a way that it could play a role in critical thinking. Consider Tanya, who has a strong preference that her nephew Joe graduate from high school. Setting suspicions of psychological egoism aside, this is not a preference for her own experience, for she wants his graduation to occur even if she never reaps any

experiential joy from its occurrence. Now, let us say that the principal dislikes Joe and wants to keep him from graduating (specifically, by coming on stage at the appropriate time and yanking the diploma from his hands); but the principal, being a good Harean, changes his mind after doing a calculation of preference-outcomes. In part, he imagines what it will be *like* for Tanya (who will be out in the audience) to experience the frustration of her preference. I submit that such a preference-frustration would carry moral weight in critical thinking. If this is right, then Hare is only justified in invoking an 'experience requirement' that says preference-frustrations and -satisfactions only count when they come into our experience [see Griffin (1986), pp. 13, 16–17]. An experience requirement would exclude certain types of preferences (e.g. desires for certain world-states to obtain after one's death), but it would also exclude some preferences for the reason that their satisfaction or frustration just happen, contingently, not to fall inside one's experience. Synchronic preferences that take one's experience as their object would, by their very nature, always be included.

   That being said, it may be that all of this falls perfectly in line with what Hare meant by 'experiential' preferences; I find it rather difficult to tell from the various passages where he discusses these matters (see note 27). Regardless, in the remainder of this essay, I will assume only what I think Hajdin has established — that only *experienced* preference-satisfactions and -frustrations carry weight in moral thinking — though I will, for convenience's sake, use Hare's terminology (e.g. experiential, synchronic preferences) to mark the concessions that have been made. This will not have much effect on the arguments of this paper, but it does invite some very difficult questions about the nature of preferences and of Hare's particular brand of preference-utilitarianism.

29 I am indebted to Adam Shriver for bringing this point to my attention.

30 Hare (1975) op. cit., p. 166.

31 Hare (1989b) op. cit., p. 65.

32 Cheops' preference for a big funeral will also be excluded on the grounds that it is a now-for-then preference. The Cheops example comes from Allan Gibbard's contribution in D. Seanor and N. Fotion (eds.) *Hare and Critics* (Oxford: Clarendon Press, 1988), pp. 57–72.

33 For Hare, suffering must be an actual experience, but one can be harmed without knowing anything about it. Hare (1981) op. cit., p. 92.

34 C. Bailey, *Epicurus: The Extant Remains*, Greek Texts, with Trans. and Notes by C. Bailey (Oxford, Clarendon Press, 1926), pp. 124–25.

35 Hare (1981) op. cit., p. 115. He continues: 'Why cannot I say that I distinguish *in thought* between worlds *A* and *B* by calling *A* actual, and that this means that it is the world in which *I* am, without being committed to saying that *B* is in some shadowy sense actual too? I should only be committed to this absurdity if to *think about B* already presupposed its actuality; but does it presuppose this?'.

36 I have assumed that only beings with moral standing deserve moral consideration. Hare, whose theory of moral thinking involves imagining oneself in the position of those affected by an action, is willing to give moral *consideration* to any affected object whatsoever (e.g. a stove):

   The stove differs from me in not having concerns. It is sometimes said that stoves, mountains, and trees are outside the scope of morality (we cannot have duties to them) because we cannot put ourselves in their positions. I think that this is badly expressed. We can put ourselves in their positions, but, since when we do this we have no sentience and therefore no concerns, it simply does not matter to us what happens to us if we turn into such things, any more than it matters to us if we are going to be put on the rack after we are dead. So moral arguments relying on the move 'Are you prepared to prescribe that it should be done to you if you were a tree?' get no grip (cf. Seanor and Fotion op. cit., p. 283).

   Even if we give moral consideration to, say, a stove or a tree, these never carry any weight in moral decisions. Learning this about certain classes of objects, we cease to give them moral consideration in the future. So I do not take my use of 'moral consideration' to conflict with Hare's more liberal usage. The same results follow either way.

37 Hare (1988a) op. cit., p. 73. See also note 36.

38 This point has been effectively made by Roger Crisp in his review of Hare's *Essays on Bioethics* [R. Crisp, 'Essays on bioethics (review essay)', *Bioethics* 9,2 (1995): pp. 161–63]. Further, at various points in his writings Hare himself acknowledges total utilitarianism as a fallback position [see, e.g. Hare (1988a) op. cit., p. 67 and Hare (1998) op. cit., pp. 403–4], though he still attempts to defend his claim about harm to possible people. What I hope to contribute in this section is an explanation, in light of the developments in his 1998 essay, of just why Hare's theory leads to an obligation to make happy people.

39 Hare (1981) op. cit., p. 44.

40 *Ibid.*, p. 43.

41 There seems to be ample evidence for this claim; the following points come from Ch. 2 and 3 of Hare (1981) op. cit. The intuitive level's principles are selected by critical thinking and, even then, are only *prima facie*; in moral conflicts or uncharacteristic cases, it may be necessary to engage in critical level thinking (2.5). Intuitive level principles must be overrideable, while critical level principles are overriding (3.8). It is irrational to feel remorse (as opposed to regret) over a critical level decision (2.2). Also, Hare repeatedly censures those who are stuck at the intuitive level (e.g. thinking that one should not lie under any circumstances whatsoever), but I do not think he would censure people for being stuck in the critical level *if* they are in fact equipped to do critical thinking. Of course, no human *is* always equipped to do critical level thinking, so Hare would surely criticize anyone who fails to see this (see 2.4).

   In short, the critical level is the ideal moral standpoint. The intuitive level is only valuable insofar as it gets us as close as possible to the results of the archangel's case-by-case choices (see p. 46). Of course, all of this is perfectly compatible with the claim that we humans should use intuitive level thinking the vast majority of the time; Hare certainly gives that impression at various points.

42 *Ibid.*, p. 91.

43 *Ibid.*, p. 109 f.

44 Perhaps there is an exception: if I must do either A or B and they lead to outcomes that are equally good, then each will be supererogatory, as I have defined the term. For it will be good to do A but A's omission will not be morally wrong since my only other choice is equally good — and the same for B. Perhaps critical thinking would pass no judgment in such a situation. Or perhaps it will just be indifferent between these two options but nevertheless mandate that one of them be done (which could be seen if an action C, leading to a lesser outcome, is introduced). In any case, we can ignore this exception since it does not affect my point about creating happy people.

45 This is confirmed by Hare's bun example, *ibid.*, pp. 128–29.

46 *Ibid.*, 11.6–8.

47 There are several places where Hare implicitly or explicitly endorses a maximizing total version of utilitarianism. See, for instance, Hare (1988a) op. cit., p. 67, p. 70 f.; R. Hare (1993) 'Why I am only a demi-vegetarian', reprinted in D. Jamieson (ed.) *Singer and His Critics* (Malden, MA: Blackwell, 1999), pp. 233–46 at p. 239; Hare (1998) op. cit., pp. 403–4.

48 Hare (1981) op. cit., p. 63. See also the quote cited in note 42.

49 The supererogationist may believe that, all else being equal, we have a duty not to lower a person's quality of life but that we have no corresponding duty to raise it, even if it is praiseworthy for us to do so. Critical thinking, which only looks at the satisfaction and frustration of experiential synchronic preferences on different outcomes, will not be able to track whether a lowering or raising of well-being has taken place. Take a situation S5′, which is just like S5 except that Pam exists prior to the execution of the decision. On outcome S5′B, Pam's life is worth living (☺); on outcome S5′A, Pam's life is more worth living (☺!). To know whether we are obligated to choose A, the supererogationist will need to know Pam's quality of life in the pre-choice situation. If it was ☺!, then we must choose A. If it was only ☺, then we are only obligated to choose B though it will be admirable to choose A. Since the pre-choice situation plays no role in critical thinking, this type of supererogation cannot be accommodated by critical level thinking.

50 This result allows Hare to avoid some embarrassing problems that Parfit has raised for person-affecting views, connected to what he calls the 'Non-Identity problem'. Parfit op. cit., Part IV, Ch. 16. In one of his conundrums, we as a society can choose a policy that depletes certain kinds of resources; the bad effects of Depletion will not be felt for 200 years. Since any policy we choose will surely affect when people are born, we can expect there to be different people who are born depending on what we do. In that case, there is no problem with Depletion according to the person-affecting view (which says, roughly, that *someone* must be harmed for there to be wrong). The people on the Depletion outcome (corresponding to Joe★ in S6) may not have the quality of life had by the different people on the Non-Depletion outcome (Pam★), but they will still have lives worth living; were it not for Depletion, they would not have existed at all; *ibid.*, pp. 361–62. The inability to criticize Depletion is usually taken to be a major problem for the standard person-affecting view; as S6 illustrates, critical thinking will tell us to avoid Depletion.

51 '[Brandt and I] are both utilitarians of related though slightly different sorts; and therefore we have to make what we morally ought to do depend on the avoidance of harms to the interests of people, and the conferring on them of benefits'. Hare (1989b) op. cit., p. 65. Also: 'For utilitarians like Singer and myself,

doing wrong to animals must involve harming them. If there is no harm, there is no wrong. Further, it has to be harm overall . . .' Hare (1993) op. cit., p. 238.

52 One of Hare's signature moves is to accept some intuitively abhorrent result as being the best option in a hypothetical situation but then take refuge in the fact that, due to epistemic limitations and/or other factors, it should never be chosen in our actual world. See, for instance, the hospital example in Hare (1981) op. cit., 8.2 and his treatment of the Repugnant Conclusion in Hare (1988a) op. cit.