

# KANT'S NEGLECTED ARGUMENT AGAINST CONSEQUENTIALISM\*

Gilbert Plumer

*The Southern Journal of Philosophy* (1991) Vol. XXIX, No. 4, pp. 501–520.

ABSTRACT: The paper interprets Kant's neglected argument at *Foundations* 401 as consisting of these two premises and conclusion: (1) It follows from consequentialism that in a natural paradise people would not be obligated to be morally good. (2) But this is absurd; one ought to be morally good no matter what. Therefore, consequentialism is false. It is shown that this argument is a powerful one, mainly by showing that independent grounds support (2) and that (1) may survive a number of strong possible objections. One that it does not appear to survive, though, is that the paradise envisioned is not logically possible.

There is perhaps no better description of a kind of paradise than David Hume's:

Let us suppose that nature has bestowed on the human race such profuse *abundance* of all *external* conveniencies, that, without any uncertainty in the event, without any care or industry on our part, every individual finds himself fully provided with whatever his most voracious appetites can want, or luxurious imagination wish or desire. His natural beauty, we shall suppose, surpasses all acquired ornaments: the perpetual clemency of the seasons renders useless all clothes or covering: the raw herbage affords him the most delicious fare; the clear fountain, the richest beverage. No laborious occupation required: no tillage: no navigation. Music, poetry, and contemplation form his sole business: conversation, mirth, and friendship his sole amusement.

Then he goes on to make his philosophical point:

For what purpose make a partition of goods, where every one has already more than enough? Why give rise to property, where there cannot possibly be any injury? Why call this object *mine*, when upon the seizing of it by another, I need but stretch out my hand to possess myself to what is equally valuable? Justice, in that case, being totally useless, would be an idle ceremonial, and could never possibly have place in the catalogue of virtues.<sup>1</sup>

Kant might very well have had Hume in mind when he wrote:

Thus (A) the moral worth of an action does not lie in the effect which is expected from it or in any principle of action which has to borrow its motive from this expected effect. For (B) all these effects (agreeableness of my own condition, indeed even the promotion of the happiness of others) could be brought about through other causes (C) and would not require the will of a rational being, (D) while the highest and unconditional good can be found only in such a will.<sup>2</sup>

(I've lettered clauses for reference.) My point is that Kant may very well have seen that Hume's claim that in a condition of natural abundance there would be no need for the virtue of justice is suggestive of a *general* claim that it follows from consequentialism that in any natural condition where "every one has already more than enough" people would not be obligated to be morally good, period. And in Kant's eyes this general claim is true ((B) and (C)), and it shows that consequentialism is false (A), since one ought to be morally good no matter what (D).

---

\* I am grateful to Marcia Baron, Christine M. Korsgaard, D. S. Shwayder, and Mark Timmons for helpful comments on an earlier draft.

I have just in effect given my considered interpretation of Kant's argument. I mean I have simplified Kant's argument in order to bring out its power. Indeed, it is his most convincing argument against consequentialism, according to the general opinion of my students. Section I will explain in some detail how my interpretation fits. I will also indicate how and why the argument has been neglected, and when not neglected, misconstrued. Sections II and III will be devoted to assessing the argument.

## I. The Argument

The treatment which Kant's argument has received in the literature (at least the English literature of this century) has been odd. Commentators don't seem to know quite what to do with it. Usually, they do nothing. Perhaps the reason for this is that (without saying so) they regard it as just a recapitulation. There is no doubt that (A) and (D) may be derived from preceding text (indeed, the argument begins with "thus"); it is (B) and (C), and in particular, "other causes" in (B), which raise the question. I take the "other causes" to lie in inanimate nature, and thereby interpret the argument as introducing something new into the discussion in the First Section of the *Foundations*. It seems that the only plausible interpretation contrary to my own here would be to take the "other causes" to be inclinations, specifically, being inclined to one's own happiness and sympathy for others, which Kant does discuss prior to the argument (esp. 14–15/398–99). But while on Kant's view these "would not require the will of a rational being," as (C) says, they do not set up a sufficient contrast to at least one of the causes indicated in (A). The "other causes" in (B) are supposed to be *other than* the causes indicated in (A).

What are these latter causes? Apparently, there are two. They are action involving the expected effect of (as is parenthetically explained in (B)) "agreeableness of my own condition" or "the promotion of the happiness of others," and action involving a "principle of action which has to borrow its motive from this expected effect." Kant may intend to emphasize the first of these action categories, as he conjoins the clauses which indicate them with "also auch nicht." Beck translates this simply as "or," but it literally translates as *therefore nor*. So the clauses are supposed to be related as proposition and consequence. Now the difficulty is that acting from the inclination to promote one's own happiness *is a case* of engaging in action involving the expected effect of "agreeableness of my own condition." And the same applies, *mutatis mutandis*, to acting from sympathy for others. On the other hand, it is true that for Kant, acting merely from inclination is not acting on principle (e.g., 16/399). But if we take the "other causes" in (B) to lie in inanimate nature, then they contrast to *both* of the causes indicated in (A), since the former causes involve no *action*.

As far as I can find, none of Kant's other writings are particularly helpful on this matter, and there are only two very brief discussions of his argument in the secondary literature. Each goes wrong, I believe, though the mistakes are instructive. The first is H. J. Paton's. Although he agrees there is something new going on ("Kant suggests also another argument"), in his interpretive summary he represents clause (B) as: "Other causes might produce the same results as a good will: they might, for example, produce happiness for ourselves or others."<sup>3</sup> It seems this couldn't be right, as it would have Kant denying his own view. Again, the "other causes" in (B) (which Paton does not attempt to explain) are supposed to be other than the causes indicated in (A). If these latter causes amounted to having a good will, then (A) would in effect deny that the good will has moral worth. But of course for Kant, to have a good will *is* to be morally good.<sup>4</sup> In fact, the causes indicated in (A) are ways Kant sees his consequentialist opponents defining being morally good: engaging in action involving the expected effect of promoting the

happiness of oneself and/or others, and acting on a principle such as a utilitarian principle. The difference between these two conceptions is that according to the first, you are being morally good so long as you act with the expectation or intention of promoting happiness, regardless of what it is that motivates you to act. Your motivation could be any number of things, e.g., sympathy, the threat of a sanction, or wanting to conform to a utilitarian principle because you accept it. The second conception is the same as the first, *except that* one's motivation must be like the last example just given. With (A), Kant captures two varieties of consequentialism (which he also distinguishes on 16/400), both of which are "subjective" insofar as they emphasize expected or intended results, as opposed to actual results. Following Kant, this is how consequentialism shall be understood in this paper, although I will not bother to keep the two varieties distinct since nothing further hangs on the difference. So I shall simply say that the consequentialist defines being morally good as acting so as to maximize happiness and (of course) minimize unhappiness.

In Brendan Liddell's interpretive summary there is also trouble over (B). He apparently intends to represent clause (B) where he says "good results frequently proceed from obviously immoral motives and actions, even from nonhuman causes."<sup>5</sup> But while the latter interpretation of "other causes" agrees with my own, immoral motives and actions are irrelevant insofar as contrary to clause (C), they *do* require "the will of a rational being."<sup>6</sup>

As for the other details of my own interpretation, one first needs to see that Kant is saying the following (where each point restates the clause with the same letter in the quoted passage):

(B') The good, as defined by the consequentialist, could be brought about by causes other than people being morally good, as defined by the consequentialist,

(C') such that the good would not require people's<sup>7</sup> wills.

(D') But this is absurd, for the good which is "the condition of all others" and "good of itself without regard to anything else" (12–13/396–97) can be found only in people's wills.

(A') Therefore, consequentialism is false.

Consider (B') and (C'). Let's use the term *paradise* for a situation where (as Hume puts it) "every one has already more than enough" or (eliminating the overkill) simply enough, where (as Kant puts it in (B)) there is "agreeableness of my own condition" and "the happiness of others." Kant takes (maximal) happiness to be "the sum of satisfaction of all inclinations" (15/399; cf., e.g., 21/405), and the satisfaction of inclination or desire is indeed how many consequentialists define the good (consequences) to be maximized (and its opposite, dissatisfaction, as the bad to be minimized).<sup>8</sup> Now there are three main forms of consequentialism: utilitarianism, ethical egoism, and ethical altruism.<sup>9</sup> And as we are understanding them here, the good, as viewed distinctively by the utilitarian, simply is paradise, whereas for the ethical egoist and ethical altruist, paradise includes the good, i.e., it includes "agreeableness of my own condition" and "the happiness of others," respectively. So what, *à la* (B'), are the possible scenarios in which paradise is brought about and sustained by causes other than people being morally good in the way that the consequentialist (distinctively) defines being morally good, i.e., acting so as to maximize happiness and minimize unhappiness? A restriction on these scenarios, in view of (C') and the above discussion of (A), is that paradise have a nonrational and inanimate cause—that it

be, let us say, a *natural paradise*. As far as I can see, there are only two basic scenarios. The first is exemplified by Hume's description of a condition of natural abundance. The second is suggested by the first. Suppose our genetic endowment were such that the strength and number of our desires are diminished while the prevalence of what Hume calls "*external conveniences*" remains constant (all this is relative to what the natural situation is *in fact*). Then it seems it could be claimed with equal justification, at least for some geographical areas, that "without any care or industry on our part, every individual finds himself fully provided with whatever his most voracious appetites can want, or luxurious imagination wish or desire." What this appears to mean for the consequentialist is that we would not be obligated to act so as to maximize happiness and minimize unhappiness, insofar as these would be maximized and minimized without our having to do anything to bring this about. This leads to the first premise of my considered interpretation of Kant's argument:

(1) It follows from consequentialism that in a natural paradise people would not be obligated to be morally good.<sup>10,11</sup>

The concern this raises about consequentialism might take other forms. Consequentialism, at least in its utilitarian variety, is inherently *utopian*, insofar as it tells us to work toward the maximization of everyone's happiness. So the question arises, if utopia is reached, what then? Do the demands of the theory wither away? The other side of the coin is that utilitarianism is born of woe; it even has historical roots in efforts to eliminate adversity. If it were to succeed, then wherefore utilitarianism? It may be that paradise could be brought about by everyone's conforming to any one of the three principal varieties of consequentialism—ethical egoism, ethical altruism, or utilitarianism. And suppose that with the proper TECHNOLOGY, such a paradise would sustain itself without our help; then it seems again that people would not be obligated to be morally good if consequentialism is true. The only real difference between this sort of scenario and the sort I understand to be envisioned in Kant's neglected argument is that in the latter, paradise is brought about and sustained by nature.

The source of the problem might be put in terms of the good and the right, which John Rawls, for one, takes to be "the two main concepts of ethics."<sup>12</sup> On consequentialism, the highest or greatest good, i.e., happiness of oneself and/or others, is not the same thing as the right or being morally good, which raises the question of the possibility of the one existing without the other. This does not apply to Kant's theory in the *Foundations* since, according to it, the highest good, i.e., having a good will, is identical to the right.<sup>13</sup> It is true that elsewhere Kant develops a different conception of the highest good. It is the idea of being happy in proportion to and conditioned by being perfectly morally good.<sup>14</sup> But even on this conception, the highest good could not exist without the right since it includes the right. Consequentialists, on the other hand, do not even see the right as included in the highest good, notwithstanding the fact that they define the right in terms of acting so as to maximize the good.

Now for the second premise (which refers back to the first) of my considered interpretation of Kant's argument:

(2) But this is absurd; one ought to be morally good no matter what.

(And the conclusion is the same as (A'), i.e., that consequentialism is false.) An examination of (D') leads to (2). In (D') I quote two phrases as initial explications of the key terms "highest" and

"unconditional," respectively, in the (D) premise of Kant's argument. These phrases are from passages preceding the argument, passages among those in which Kant argues specifically for premise (D). Consider the second phrase first: Kant maintains that the good will is "good of itself without regard to anything else." This means that the goodness of the good will does not depend on anything else; if it did, then the good will might very well be good in some possible situations, but not in others. But in fact, the good will is not contingently good; it is good in every possible situation. It has "absolute worth" (10/394), not relative worth. Kant drives this idea home in his description of the possible situation where "by a particularly unfortunate fate or by the niggardly provision of a stepmotherly nature" a person's will is such that "even the greatest effort should not avail it to achieve anything of its end, and if there remained only the good will... it would sparkle like a jewel in its own right, as something that had its full worth in itself (10 /394). About such a situation you might at first think that in it the good will has no value. But according to Kant, "common sense" (10–11/394) indicates that the good will is still good, in other words, you ought to have a good will, even if you find yourself in a situation where you are altogether ineffectual. And for Kant, as we've seen (note 4), to have a good will is to be morally good.

That Kant holds the idea in (2) that one ought to be morally good no matter what (the situation one is in) is further evidenced by considering the first phrase quoted in (D'): Kant maintains that the good will is the "highest" good in that it is "the condition of all others," i.e., every other good. This means that for any such proposed good, it is not really good unless it is had by a good will (9–10/393–94, 14/397–98). So regardless of what ostensible good one has in a possible situation, one ought to be morally good. In particular, one ought to be morally good in a natural paradise, a situation in which everyone is happy. Kant argues (9/393) that happiness must be *deserved*, otherwise it is not in fact good. And a person deserves to be happy only if she or he has a good will, i.e., is morally good.

Kant's neglected argument against consequentialism is somewhat akin to another argument he gives to try to show that consequentialism lacks universality. Kant believes that perceptionless, desireless, rational beings are possible, e.g., God (33/415, 46/428, 54/435, 67/449, 73–74/454). Kant's convenient term for such beings is "holy wills" (later we'll see why they're all supposed to be holy). Comparing holy wills to human beings, *qua* human beings, he says they "would be equally subject to objective laws (of the good)," as he conceives such law (30/413–14), though holy wills couldn't be subject to consequentialism, assuming that it is defined in terms of desire (43/425, 63/444).

## II. Assessment of Premise (2)

I regard it as given that Kant's neglected argument against consequentialism is valid. But are the premises true? The next section will principally be devoted to evaluating premise (1) of my considered interpretation of his argument. In the present section I shall give an independent argument for the proposition in premise (2) that one ought to be morally good no matter what. I believe that this proposition is true and that its truth does not depend on any peculiarity of Kant's ethical theory. I have already indicated Kant's reasons for holding the proposition, but do not find them as convincing as the following.

It seems that *being morally good* may be generally defined as *doing what one ought to do*.<sup>15</sup> If you plug this definition of being morally good into the statement that one ought to be morally good, you get: one ought to do what one ought to do. This is necessarily true, for its

denial is self-contradictory. To hold that it's not the case that one ought to do what one ought to do would be to hold that there is something that one ought to do, and it's not the case that one ought to do it. (Here's the proof, should one be needed: "One ought to do what one ought to do" is logically equivalent to "what one ought to do, one ought to do," which is logically equivalent to "for anything  $x$ , if one ought to do  $x$ , then one ought to do  $x$ ," the denial of which is logically equivalent to "there is something  $x$  such that it's not the case that if one ought to do  $x$ , then one ought to do  $x$ ," which is logically equivalent to "there is something  $x$  such that one ought to do  $x$ , and it's not the case that one ought to do  $x$ .") As a necessary truth, the proposition that one ought to do what one ought to do is true in every possible situation (or, if you prefer, world). What would follow is the proposition in premise (2) that one ought to be morally good no matter what, and if this proposition is true, conformity to it, i.e., not being committed to a denial of it, is a necessary condition for the adequacy of a theory of morality.

So what can be said in favor of the general definition of *being morally good as doing what one ought to do*, thereby justifying the interchange of these terms? Space constraints do not allow a full argument to be given here. All that can be shown is that the definition is not itself incompatible with consequentialism (if it were, it would beg the question) and that it has merit. The definition is certainly not incompatible with Kant's theory; in fact, Kant seems to endorse it. He says that the laws of "moral philosophy" are "laws according to which everything should happen" (3–4/387–88; cf. 44/426). And of course these laws pertain to what people do—their actions—else we apply moral predicates to nonrational happenings, such as rainshowers and turkey gobbles.<sup>16</sup>

Perhaps the most compelling point in its favor is that the definition yields a straightforward, no-nonsense answer to the question, "why should I be moral?" One should do what one ought to or should do because to deny that one should is (as above) self-contradictory. You should do what you should do exactly because you've already determined you should do it.

Perhaps the most serious possible objection to the definition is that it is too broad. It might be said that there is on the one hand a specific moral "ought," and on the other, a nonmoral "ought" or (alternatively) an all-things-considered "ought" or (alternatively) a plethora of other "oughts": those of prudence, legality, custom, etiquette, skill, occupation, and so on. The common theme here is that morality is not alone in governing how we should act. What tends to underlie such a view is an altruistic orientation in ethics, i.e., the belief that morality is exclusively other-regarding, that "the moral point of view" is that of concern for others.<sup>17</sup> Consider (what I'm calling, following Frankena) ethical altruism, one of the three main forms of consequentialism. It distinctively defines being morally good as acting so as to maximize the happiness and minimize the unhappiness of others. Now it would appear to be wildly implausible for the ethical altruist to accept the general definition of being morally good as doing what one ought to do, thereby holding that helping others is *all* of *what* one ought to do. For the satisfaction of one's own self-regarding desires surely counts for something. So the objection is that not only is the general definition *too* general or broad, it is inherently incompatible with consequentialism *qua* ethical altruism.

First note that this does not apply to ethical egoism or utilitarianism, the two other main forms of consequentialism, since they distinctively define being morally good as acting so as to maximize the happiness and minimize the unhappiness of oneself or oneself and others (respectively). But second, I think the objection rests on a mistake in any case. The general definition need not be understood as saying that being morally good is *all* of what one ought to do, that morality alone governs how we should act. It is more plausible to take it as saying that to

be morally good is to do what one ought to do *simpliciter*. This is what can be learned from the objection. For example, something may be what one *legally* ought to do, but not exhibit moral goodness, not be what one ought to do *simpliciter*. The same applies to anything else which may be said to govern how we should act: prudence, custom, etiquette, skill, occupation, or whatever. Each may dictate actions which are not without qualification what one ought to do. Obviously, however, this would have to be modified if, e.g., ethical egoism is true. For then being prudent, being morally good, and doing what one ought to do *simpliciter* are all the same thing. A parallel remark applies to normative ethical relativism with respect to custom. So I am not by any means saying that considerations of prudence, legality, etc. are irrelevant to morality. I am just saying that the final determinant of what one ought to do is morality, regardless of what determines morality.<sup>18</sup>

If either Kant's theory or consequentialism were what are called "virtue ethics," as opposed to (misleadingly) "duty ethics," then it might be thought that they would be incompatible with the general definition. For wouldn't the virtue ethicist define *being morally good* as *being what one ought to be*? Not necessarily, for it can be argued that the essential difference between the two ethics lies in what they identify as determining what one ought to do, how people ought to conduct themselves. Is it one or more qualities of character (consider Plato's notion of a just individual), or is it one or more principles (e.g., the Greatest Happiness Principle)? Both approaches are "duty ethics"; it's just that the one is virtue-based and the other principle-based.<sup>19</sup> In any case, both Kant's theory and consequentialism are in fact principle-based (of course Kant's fundamental principle is what he terms "the categorical imperative").<sup>20</sup>

### III. Assessment of Premise (1)

In section I we saw some reason for maintaining, *à la* premise (1), that if consequentialism is true, then in a natural paradise it's not the case that one ought to act so as to maximize happiness and minimize unhappiness, which is how the consequentialist distinctively defines being morally good. What would follow, by substituting the general definition of being morally good, is that if consequentialism is true, then in a natural paradise it's not the case that one ought (*simpliciter*) to do what one ought (*simpliciter*) to do, and so a self-contradiction is possibly true, which, as premise (2) says, is absurd.

In considering whether premise (1) is true, I shall first discuss the part of the consequentialist principle which says that one ought to act so as to maximize happiness, i.e., desire satisfaction; later I'll take up the minimizing-unhappiness part. But before doing this, some preliminary elucidation of the concept of desire is necessary. It might be thought that to desire something is just to have getting or doing it as an end (aim). But there seems to be a sensuous aspect to desire that this simple elucidation misses. Kant may well capture this where he says that there is a "distinction between subjective ends, which rest on incentives, and objective ends, which depend on motives valid for every rational being," such that all and only subjective ends belong to "desire" (45/ 427). A seemingly unexceptionable example would be to contrast a particular sexual desire which some people don't have, with the end of resolving known contradictions in one's belief structure (even if one claims not to care, "the caustic of reason," as I think Michael Scriven deems it, goes to work). This understanding of desire appears to coincide nicely with Hume's description of a natural paradise, where, recall, nature is envisioned as providing "*external* conveniencies" to the extent that "without any uncertainty in the event, without any care or industry on our part, every individual finds himself fully provided with

whatever his most voracious appetites can want, or luxurious imagination wish or desire." We should be careful, however, not to take the target of Kant's neglected argument as just base hedonistic consequentialism. We must be pretty flexible about what we would allow in a full description of a natural paradise in order to give Kant's argument a fair evaluation. For instance, just as Hume says one's "natural beauty...surpasses all acquired ornaments," a similar point could be added about innate intelligence. (Actually, this may be implicit in Hume's description where he says, e.g., that "music, poetry, and contemplation" form one's "sole business.")

Taken literally, the longer fragment from Hume re-quoted in the preceding paragraph seems to suppose that every desire of every person would be satisfied without anyone ever having to do anything to bring this about. Let us take this as the precise specification of what I'll call a *Kantian natural paradise*, a paradise which has a nonrational and inanimate cause (= the "other causes" in (B), as has been argued). I think that if this conception involved no difficulty, the consequentialist would indeed be in trouble, as Kant claims. For in consequentialist terms it would surely be *pointless*, hence not obligatory, to act so as to maximize happiness in a situation where happiness would be maximized without anyone ever having to do anything to bring it about (there'd be no "uncertainty in the event"). So-acting would be pointless because at best it would have the *effect* of overdetermining maximal happiness, and at worst, undermining it, insofar as meddling in the course of nature would have unhappy results due to our finite abilities. And as astonishing as it may be, such activity seems to be just the kind of thing one *would* have an obligation to engage in if Kant is correct in claiming that "the moral worth of an action does not lie in the effect which is expected from it or in any principle of action which has to borrow its motive from this expected effect," that the maxims of promoting the happiness of oneself, and that of others, derive from the categorical imperative (15/399, 41/423), and that the categorical imperative is "the principle of morality" (54/436).

I would like to reject a couple of ways the consequentialist might reply, before proceeding to what I believe is a genuine difficulty. The first is the idea that all that has been shown with respect to the consequentialist obligation to act so as to maximize happiness is that in a Kantian natural paradise the obligation would give one *nothing to do*; it has not been shown that it would *cease to exist*. After all, for example, my obligation to care for my child when she is sick doesn't cease to exist when she is well. It just gives me nothing to do.<sup>21</sup> I think that generally speaking, it is true that from the fact that an obligation gives you nothing to do it does not follow that it doesn't exist. But we are concerned with a situation or world in which the supposed obligation *never* gives *anyone* anything to do, quite unlike the parental obligation case. I propose it *does* follow that the obligation doesn't exist, otherwise, e.g., we (in the actual world) could have obligations toward merely possible persons such as Sherlock Holmes. What is queer about such obligations? It seems to me that it is exactly that they never give anyone (in the actual world) anything to do, regardless of the reason. (And it's not clear that the reason is that there can't be causal connections between any of us and Holmes; after all, Conan Doyle created or constructed Holmes.)

Secondly, it might be thought that in a Kantian natural paradise consequentialism would still obligate people to act so as to maximize happiness *by* obligating them to act so as to not interfere with the course of nature. Let nature maximize happiness on its own; to do anything more would be pointless. This approach does not seem to me to be a very promising one for the consequentialist. The reason is that it seems to contain a self-contradiction which I don't see any way of entirely suppressing. The consequentialist holds that people ought to act so as to maximize happiness. But the consequentialist is here regarded as also holding that one ought to



let nature maximize happiness *on its own*—so it's not the case that people ought to act so as to maximize happiness. The problem is analogous to the "hedonist's paradox": pleasure to be got must be forgot (or so goes the mnemonic). It's not for nothing that it's called a paradox. I realize that the approach might be regarded as just another case of "indirect" consequentialism, and according to Alan Donagan, for example, "ever since Sidgwick, with characteristic intellectual honesty, distinguished between what it is beneficial to praise, and what it is beneficial to do, it has been an open secret that, depending on circumstances, utilitarianism can justify supporting almost any practicable moral code."<sup>22</sup> All I can really do here is indicate why the approach is not convincing to me (and note that secrets and special intellectual honesty are often required when something is amiss).

The reply I wish to develop is this. If it were literally the case that one's desires were satisfied without one ever having to do anything to bring this about, not even (say) open one's mouth to let the "raw herbage" fall in from the trees (not to mention engage in mental activity), then surely, one would not be a human being (or indeed, any higher life form). The situation envisioned is at best mysterious and at worst impossible: that desires of a rational being may be satisfied in a way that is in every whit passive, as if one could be desirous, rational mush in a glorified petri dish. Certainly, it has no connection with humanity; in other words, being a human being necessarily involves playing some active role in the satisfaction of your desires. So even presuming that a Kantian natural paradise is possible, if being human beings is an essential property of us (human beings), then it would not be *us* who in such a paradise consequentialism fails to obligate to act so as to maximize happiness; rather, it would be nonactual, merely possible rational beings. On the other hand, if being human beings is not an essential property of us, then even though in a Kantian natural paradise consequentialism would fail to obligate us to act so as to maximize happiness, we would nevertheless not be human beings.

Given that this is correct, the consequentialist who is also an ethical naturalist would dismiss Kant's argument. For ethical naturalism makes no bones about being concerned only with human beings and comparable life forms (such as, possibly, extraterrestrials). I therefore think that Hume, an ethical naturalist *par excellence*, engages in hyperbole in his description of a natural paradise (this is indirectly borne out in an appendix to the *Enquiry* entitled "Some Farther Considerations with regard to Justice," esp. pp. 304, 306). The most extreme, yet seemingly still plausible, interpretation of Hume's description is that every desire of every person would be satisfied, though only given that each plays some active role in bringing this about. And the role one plays in the satisfaction of the desires of oneself or others could be, but could never be anything more than, that of a nonredundant (i.e., not overdetermining) *contributory* cause, as opposed to "the" (primary) cause, which in all cases is nature. On this construal of a natural paradise there remains a consequentialist point, hence obligation, to acting so as to maximize the happiness of oneself and/or others.

The criticism evidently wouldn't faze Kant. He is no ethical naturalist. And he believes that it's possible for a rational being to never have any perceptions or desires, i.e., be a holy will, so why not that it's possible for one's desires to be satisfied without one ever having to do anything to bring this about? As Kant himself holds, being a human being necessarily involves having or (if one is unconscious or dead) having had perceptions and desires (see, e.g., 33/ 415, 73–74/454). This is just part of the stratified, teleological aspect of life itself.<sup>23</sup> So holy wills and the inhabitants of a Kantian natural paradise are analogous in the respect that neither are human beings (or indeed, any higher life form), though they are supposed to be rational beings.

But in fact, it seems to me that a Kantian natural paradise is not merely inhuman, it is logically impossible. This is difficult to show, and Kant's argument is correspondingly resistant to rebuttal. Here's how I'd try. Take, for example, the desire to chew gum. Even if gum grew on trees ready-to-chew, and even if there were some mechanism by which nature automatically transferred the ripened gum to the desirous person's chewing orifice, the person would still have to do something in order for the desire to be satisfied, namely, *chew*. Yet suppose that in addition to transferring the gum, the natural mechanism somehow also made the orifice go through the motions of chewing. But in that case *the person* would not be doing the chewing, thus *the person's* desire to chew gum would not be satisfied (this is especially obvious where the orifice is externally made to go through the motions of chewing, leaving its muscles relaxed). A similar argument could be run with respect to any desire whose specification inherently involves an action term: it is logically impossible for the desire to be satisfied unless the person does something (the action) to bring this about.

However, not all desires are of this type. There are many desires which have to do with maintaining one's existence or comfort and which don't inherently involve an action term in their specification, e.g., hunger, thirst, and wanting fresh air or a soft sleeping surface. It seems that no desire is conceivable which both does not fall into one of these two categories and is not reducible to ones that do (compare the earlier point that to desire something is to have getting or doing it as a subjective end). So, could there be people (rational beings) whose desires are only of the existence-comfort type? The problem here is that it is difficult to see how people could *come to have* such desires without ever having the action type. For instance, could a hard day of *only* doing things on the order of trying to resolve contradictions in one's belief structure bring on thirst? And in biological-evolutionary terms the idea just makes no sense at all. Here there is another alternative, viz., creation *ex nihilo*, but to appeal to this would be to jump from the frying pan to the fire with respect to logical possibility.

Let us turn now to the minimizing unhappiness part of the consequentialist principle. For certain aspects of this, e.g., acting so as to remove unhappiness (desire dissatisfaction), obviously the same ideas apply as with respect to maximizing happiness. Yet what about acting so as to not inflict unhappiness? Suppose that a Kantian natural paradise were logically possible. It might be thought that there's no question that in such a paradise consequentialism would still obligate people to act *this* way. But hold on. Certainly, the question arises as to whether in a Kantian natural paradise it would be psychologically possible to act so as to *inflict* unhappiness, or if not that sadistic or masochistic, act with the knowledge that one's action would or quite possibly could (for all one knows or cares) bring about more unhappiness than happiness. Could one cheat, backbite, bribe, batter one's child, rape, engage in self-mutilation, or the like? It seems that a case could be made for maintaining that such conduct inherently results from unsatisfied desire together with there being no prospect of desire satisfaction without such conduct, and therefore, that in a Kantian natural paradise it would be psychologically impossible.<sup>24</sup> And given the truth of the ought-as-constraint principle, viz., for anything *x*, being psychologically unable to do otherwise than *x* entails not having an obligation to do *x*, it should follow on any theory of morality, not just consequentialism, that in a Kantian natural paradise people wouldn't be obligated to act so as to not inflict unhappiness. Now Kant does subscribe to the ought-as-constraint principle. Concerning holy wills he says "the 'ought' is here out of place, for the volition of itself ["according to its own subjective constitution"] is necessarily in unison with the [moral] law" (30–31/414). Kant believes that holy wills are psychologically unable to do otherwise than act in accordance with moral law, i.e., be holy, because as perceptionless,

desireless, rational beings they have nothing that tempts them to violate moral law, law which is dictated by their own rationality or reason (29–31/412–14, 44/426, 67–68/449; cf. *Religion*, pp. 30–31, 51n). Anyway, the point is that it could be that even on Kant's view, people in a Kantian natural paradise wouldn't be obligated to act so as to not inflict unhappiness (though, as we've seen, they would be obligated to act so as to maximize happiness). And this does not conflict with the idea in premise (2) that in every possible situation one ought to do what one ought to do, since it presupposes that acting in a certain way when one is psychologically unable to do otherwise is not something one ought to do, i.e., the ought-as-constraint principle. This general defeating condition on what one ought to do might be compared to logical impossibility. Acting in a certain way when it is logically impossible to do so, e.g., helping others in a world in which you're all alone, is not something you could, on any intelligible theory of morality, have an obligation to do. Finally, it seems that on the modified Humean construal of a natural paradise it *would* be psychologically possible to act so as to inflict unhappiness. For example, such conduct directed at strangers might appear to be the easiest way to play one's role as contributory cause in the satisfaction of one's own desires, or those of a friend. Given this, consequentialism would obligate people to act so as to not inflict unhappiness.

In sum, my assessment of Kant's neglected argument against consequentialism is this. While his argument is typical of his ethics for its bold forcefulness, it does not appear to succeed. Although its second premise seems true, its first—that it follows from consequentialism that in a natural paradise people wouldn't be obligated to be morally good—does not. This would follow from consequentialism only if the natural paradise were a Kantian natural paradise. But, contrary to what the premise presupposes, there does not appear to be any such logically possible situation. It is no fault of consequentialism, or of any theory for that matter, that it fails to hold in logically impossible situations.

Moreover, if my speculation that it would be psychologically impossible for people in a Kantian natural paradise to act so as to inflict unhappiness is false, then, supposing that such a paradise were logically possible, consequentialism *would* obligate them to be morally good in the limited sense of avoiding such conduct. This would at least make premise (1) misleading.

There is a subtle objection which might arise at this point, and which ultimately concerns defining *being morally good* as *doing what one ought to do*. As was mentioned in indicating that Kant subscribes to this definition, he says that the laws of "moral philosophy" are "laws according to which everything should happen." And we saw in section I that he holds that holy wills are "subject to objective laws (of the good)," yet we've just seen him also saying that "the 'ought' is here out of place." Isn't this inconsistent? I think an adequate response would be that "properly" speaking, holy wills are morally good in that they do what we (non-holy-wills) ought to do (Kant may be saying exactly this on 67–68/449). Similarly, for example, it might be held that it follows from ethical altruism that in a world in which you're all alone, you would not be obligated to act so as to maximize the happiness and minimize the unhappiness of others (because so-acting would be logically impossible), i.e., be morally good, i.e., do what we (who are not in such a world) ought to do. In any case, the main point is that the definition of *being morally good* as *doing what one ought to do* does mean that if you're morally good, then you're doing what *you* ought to do, provided only that no general defeating condition on what one ought to do (such as psychological or logical impossibility, as discussed) applies to *you*. This preserves the interchangeability of the defined and defining terms, which otherwise would have counterexamples.

## NOTES

<sup>1</sup> *An Enquiry concerning the Principles of Morals*, L. A. Selby-Bigge, ed., 2nd edition (Oxford Univ. Press, 1902), pp. 183–84. First published in 1751. Cf. Hume's *A Treatise of Human Nature*, L. A. Selby-Bigge, ed. (Oxford Univ. Press, 1888), pp. 494–95. First published in 1740.

<sup>2</sup> Immanuel Kant, *Foundations of the Metaphysics of Morals*, L. W. Beck, trans. (Indianapolis: Bobbs-Merrill, 1959), 17/401. The second numeral is the pagination of the Prussian Academy edition. Future references to this work will be given in this manner in parentheses. The original German text, first published in 1785, reads as follows:

Es liegt also der moralische Werth der Handlung nicht in der Wirkung, die daraus erwartet wird, also auch nicht in irgend einem Princip der Handlung, welches seinen Bewegungsgrund von dieser erwarteten Wirkung zu entlehnen bedarf. Denn alle diese Wirkungen (Annehmlichkeit seines Zustandes, ja gar Beförderung fremder Glückseligkeit) konnten auch durch andere Ursachen zu Stande gebracht werden, und es brauchte also dazu nicht des Willens eines vernünftigen Wesens; worin gleichwohl das höchste und unbedingte Gute allein angetroffen werden kann.

See *Immanuel Kant's Sammtliche Werke*, K. Rosenkranz and F. W. Schubert, eds. (Leipzig: Leopold Voss, 1838), p. 21.

<sup>3</sup> *The Categorical Imperative*, 5th ed. (London: Hutchinson, 1965), p. 58.

<sup>4</sup> "The pre-eminent good can consist only in the conception of the law in itself (which can be present only in a rational being) so far as this conception and not the hoped-for effect is the determining ground of the will. This pre-eminent good, which we call moral, is already present in the person who acts according to this conception, and we do not have to look for it first in the result" (17/401). Cf., e.g., *Kant's Religion within the Limits of Reason Alone*, T. M. Greene and H. H. Hudson, trans. (New York: Harper & Row, 1960), p. 31.

<sup>5</sup> *Kant on the Foundation of Morality* (Bloomington: Indiana Univ. Press, 1970), p. 67.

<sup>6</sup> As Thomas E. Hill, Jr., points out in "Kant's Argument for the Rationality of Moral Conduct," *Pacific Philosophical Quarterly* 66 (Jan./April 1985), p. 3, Kant's *Foundations* is sometimes read as propounding "the outrageous view that immoral acts are unfree and not even willed." Hill argues against this interpretation.

<sup>7</sup> "Rational beings are designated 'persons'" (46/428; cf. 57/438). A rational being is simply any being "endowed with reason" (67/449; cf., e.g., 28/411–12).

<sup>8</sup> Such consequentialists have come to be called "preference" consequentialists. For discussion, see, e.g., T. L. S. Sprigge, *The Rational Foundations of Ethics* (London: Routledge & Kegan Paul, 1988), pp. 23–26.

<sup>9</sup> Cf. William K. Frankena, *Ethics*, 2nd edition (Englewood Cliffs: Prentice-Hall, 1973), pp. 15–16.

<sup>10</sup>The reader will have noticed that in (1) I've freely substituted the phrase "be morally good" for the phrase "act so as to maximize happiness and minimize unhappiness" which appears in the consideration that leads directly to (1). For consequentialists whose definition of the former phrase in terms of the latter propounds *synonymy*, whether as a report or just a recommendation, such a substitution is unquestionably legitimate. On the other hand, consequentialists whose definition only asserts *co-extensiveness* (i.e., given their actual meanings, the two phrases and their cognates apply to or are true of the same things) might question whether the context is opaque, and thereby whether the substitution is legitimate. With respect to co-extensive phrases, an opaque context is one in which the interchange of them does not in all cases preserve truth-value. But the only uncontroversial way of showing that a context is opaque is to furnish an uncontroversial case where such an interchange does not preserve truth-value, and as far as I can see, there is no such case for the kind of context we are concerned with. (I take it that all consequentialists are committed to asserting at least the co-extensiveness of some such phrases. Note that with respect to definitions, I speak in the conceptual mode, rather than the linguistic—except when the latter is convenient, as in this note.)

<sup>11</sup>In order to understand Kant as having an argument against consequentialism, he does not actually need to be understood as appealing to the idea of paradise. Again consider (B) and (C) in light of (A). Strictly speaking, all Kant needs is one case where the expected consequentialist good of an action could have been brought about through a natural (nonrational and inanimate) cause. However, I read (B)'s "all these effects" in context as implying *for any action, its expected effect* could be brought about by natural causes. Of course we don't *have* to suppose that these causes *uniformly* fall under the category of a natural paradise, such as Hume's description of a natural abundance. But surely, that is the *simpler* interpretation. And the idea of paradise is floated by Kant himself earlier in the First Section (11/395). There the paradise is one which is brought about by "instinct alone," rather than in conjunction with reason, guiding "the creature's" conduct. So although the cause of the paradise is nonrational, it is *animate*.

<sup>12</sup>*A Theory of Justice* (Cambridge: Harvard Univ. Press, 1971), p. 24.

<sup>13</sup>Kant identifies the highest good with having a good will at 12/396, and this with being morally good at 17/401 (cf. my note 4). Again, by "the right" Rawls and I mean being morally good, and to avoid misunderstanding it is worth noting that Kant holds that "it is not sufficient to that which should be morally good that it conform to the [moral] law; it must be done for the sake of the law" (6/390).

<sup>14</sup>*Critique of Practical Reason*, L. W. Beck, trans. (Indianapolis: Bobbs- Merrill, 1956), 114/110ff.

<sup>15</sup>With this definition I am propounding synonymy, and I am propounding it as a report *and* a recommendation. In other words, I think "being morally good" in fact has the same meaning as "doing what one ought to do," and I see no good reason for change. To avoid misunderstanding, it should be noted that I quite agree with Rawls (*op. cit.*, p. 130) that "the merit of any definition depends upon the soundness of the theory that results; by itself, a definition cannot settle any fundamental question."

<sup>16</sup>It might be objected that the definition is incompatible with Kant's view that being morally good requires that one act "from duty," as opposed to merely "in accordance with duty" (13/397ff.; cf. my note 13). It appears that the definition only requires the latter. This appearance is correct! But just because the definition *requires* nothing more than that one act in accordance with duty—do what one ought to do—in order to be morally good, that does not mean that you cannot accept the definition and then, in your theory, go on to regard duty as reflexive in that it includes acting from duty.

<sup>17</sup>A couple of examples of this orientation are Kurt Baier, *The Moral Point of View* (Ithaca: Cornell Univ. Press, 1958), p. 215, and Lansing Pollock, "Evaluating Moral Theories," *American Philosophical Quarterly* 25 (July 1988), p. 229.

<sup>18</sup>For further explanation and argument in favor of the sort of view I'm advocating here, see Robert B. Louden, "Can We Be Too Moral?," *Ethics* 98 (Jan. 1988). Cf., e.g., Nancy Davis, "Acting Utilitarians," *Pacific Philosophical Quarterly* 66 (Jan./April 1985), p. 138: "utilitarians are agents who have the Greatest Happiness Principle as the foundation of their evaluative standpoint: they regard it as the final word in determining what they ought to do."

<sup>19</sup>Cf. Bernard Mayo, *Ethics and the Moral Life* (London: Macmillan, 1958), p. 200ff.

I have heard it said that *being* morally good is obviously a state of character, and so cannot naturally be defined as *doing* what one ought to do. I can only suppose that this person's parents never admonished him to be good.

<sup>20</sup>Onora O'Neill interprets Kant in a nonstandard way as offering us "primarily an ethic of virtue rather than of rules." "Kant after Virtue," *Inquiry* 26 (Dec. 1983), p. 404. John Waide, it seems to me, effectively defends the standard interpretation in "Virtues and Principles," *Philosophy and Phenomenological Research* 48 (March 1988), pp. 461-63.

<sup>21</sup>I owe this point to Christine M. Korsgaard. Cf. her "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66 (Jan./April 1985), p. 32.

<sup>22</sup>*The Theory of Morality* (Univ. of Chicago Press, 1977), p. 199.

<sup>23</sup>Cf. David Hull, *Philosophy of Biological Science* (Englewood Cliffs: Prentice-Hall, 1974), e.g., p. 103.

<sup>24</sup>Cf. the remaining part of Hume's description of a natural paradise: "It seems evident that, in such a happy state, every other social virtue would flourish, and receive tenfold increase; but the cautious, jealous virtue of justice would never once have been dreamed of" (*Enquiry*, pp. 183-84).