

Mill's "Proof" of the Principle of Utility: A More than Half-Hearted Defense¹

By

Geoffrey Sayre-McCord
UNC/Chapel Hill

I. Introduction

How many serious mistakes can a brilliant philosopher make in a single paragraph? Many think that Mill answers this question by example -- in the third paragraph of Chapter IV of Utilitarianism. Here is the notorious paragraph:

The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experience. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it. If the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so. No reason can be given why the general happiness is desirable, except that each person, so far as he believes it to be attainable, desires his own happiness. This, however, being a fact, we have not only all the proof which the case admits of, but all which it is possible to require, that happiness is a good: that each person's happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons. Happiness has made out its title as one of the ends of conduct, and consequently one of the criteria of morality.²

¹ Earlier versions of this essay were presented at the March 2000 meeting of the International Society for Utilitarian Studies, the University of California at Riverside, the Australian National University, Tulane University, and to an informal group at the University of California at Irvine. I am grateful for the helpful comments people have offered on these occasions. I have especially benefited from conversations with and comments from Marc Baer, David Brink, Stephen Darwall, Gerald Gaus, Shelly Kagan, Dale Miller, Michael Ridge, Harriet Sayre-McCord, and John Skorupski.

² John Stuart Mill, Utilitarianism, ed. Roger Crisp (New York: Oxford University Press, 1998), IV, 3. Throughout, passages from Mill's Utilitarianism will be identified using the paragraph and line numbers provided by Crisp's edition.

The supposed mistakes in this paragraph are well-known and seem to come at every step. (I will rehearse them in section III.) Yet the idea that someone so smart should make the glaring mistakes people find in Mill's 'proof' seems beyond belief. And rightly so; for Mill did not make them. The apparently rampant fallacies, I will argue, are due to rampant misunderstandings of the argument. Indeed, once the real structure of the argument is brought out, I maintain, not one of the (in)famous fallacies is to be found. Moreover, the general strategy underlying the argument is, I believe, the only strategy available to those who think moral knowledge cannot be justified solely by appeal to nonevaluative truths. To bring this out, I will -- in a bizarre move -- be defending Mill by stressing the structure his own notorious argument shares with another one of dubious repute: Kant's defense of the Categorical Imperative (in its second formulation).³

As this essay's title suggests, though, while I will be giving a more than half-hearted defense of Mill's "proof," it will be less than full-scale. I will be defending the structure of the argument, but not the content Mill ends up putting into the structure. Along the way, I will similarly be defending Kant's argument, though also only its structure and not the content Kant ends up putting into the structure. In any case, I will argue that an appreciation of this shared structure reveals Mill's proof to be much more plausible than it otherwise seems. Where his proof remains weak, the problems are found not in the reasoning offered, but in (as I see it) mistakes of fact concerning human psychology -- concerning, for instance, what people value and why. Hence I share with Mill the view that, with the proof in hand, "all that remains is to consult practised self-consciousness and self-observation, assisted by observation of others."⁴ Where I differ substantially and crucially with Mill is in my thinking these sources reveal that the specific claims he relies on as premises -- the content of his structure -- are false. Similarly, as it happens, I think the claims Kant ends up putting into the same structure are false as well. Mill and Kant alike, it seems to me, attribute commitments to people -- in Mill's case on psychological grounds, in Kant's on a priori grounds -- that people do not actually have.

II. The "Proof"

A full proof of utilitarianism requires defending two things: its standard of conduct, or right action, and its theory of value. According to the standard of conduct, an agent has performed the right act if and only if that act is among the agent's best available options. To have taken any less than the best available

³ Immanuel Kant, Grounding for the Metaphysics of Morals, trans. James W. Ellington, (Indianapolis: Hackett, 1981), 36, (AK 429).

⁴ Mill, Utilitarianism, IV, 10.5.

option is, Mill thinks, to have performed the wrong act.⁵ (Although, of course, failing to take the right action may not reflect badly on the agent, depending upon why she failed to do what, as a matter of fact, was right.) Just which options count as best depends upon what is valuable and, in particular, on what is valuable as an end (since, according to Mill, everything else of value is valuable, ultimately, because it is a means to, or a part of, something valuable as an end). And according to the theory of value, happiness is the only thing valuable as an end.⁶

Mill pretty much just takes the consequentialist theory of right action for granted (as did Henry Sidgwick and G. E. Moore after him), though he is at pains, in Chapter V, to show that it is compatible with a proper understanding and appreciation of justice. In taking it for granted, Mill assumes that whatever turns out to be of value is such that we ought to maximize it, and assumes too that the value to be maximized is additive. A defense of both assumptions would be crucial to establishing utilitarianism, and Mill's failure to examine them means that the proof he does offer, of his theory of value, does not settle the issue in favor of utilitarianism, even if it is successful.⁷ Nonetheless, Mill does take the consequentialist theory of right action as given, and thinks that what needs defending is the theory of value. Indeed, he claims: "That the morality of actions depends on the consequences which they tend to produce, is the doctrine of rational persons of all schools; that the good or evil of those consequences is measured solely by pleasure or pain, is all of the doctrine of the school of utility, which is peculiar to it."⁸ Thus, as he sees it, the controversial heart of the doctrine -- and so what he tries to defend -- is "that happiness is desirable, and the only thing desirable, as an end..."⁹

The third paragraph of Chapter IV is given over to defending the first claim, that happiness is desirable as an end, while the rest of the chapter is

⁵ Mill distinguishes between the standard of conduct and "the criterion of morality," treating the latter as determined by the former (*ibid.*, IV, 9.5). In addition, at least sometimes, Mill suggests that the morality of an action turns not directly on its effects, but on whether it accords with "the rules and precepts for human conduct, by the observance of which" the best results are secured (see, e.g., *ibid.*, II, 10.10). Presumably, particular actions that are among the agent's best available options might not be in accord with the relevant rules and precepts, and particular actions that do satisfy the rules and precepts may be among the options that are less than the best.

⁶ More precisely, Mill sees happiness as a *measure* of the balance of pleasure and pain, and holds "that pleasure, and freedom from pain, are the only things desirable as ends" (because pleasure is good in itself and pain is bad in itself) and "that all desirable things (which are as numerous in the utilitarian as in any other scheme) are desirable either for the pleasure inherent in themselves, or as means to the promotion of pleasure and the prevention of pain." *Ibid.*, II, 2.10.

⁷ Among other things, it would leave completely unaddressed the suggestion that distributive considerations are (nonderivatively) relevant to what ought to be done.

⁸ John Stuart Mill, *Bentham*, in Mill, *Collected Works of John Stuart Mill*, ed. J. M. Robson (Toronto: University of Toronto Press, 1969), 10:111.

⁹ Mill, *Utilitarianism*, IV, 2.

devoted to defending the second, that happiness is the only thing desirable as an end. In both cases, the argument turns on a crucial principle of evidence, according to which "the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it." In light of this principle, what Mill needs to show is that people do actually desire happiness as an end and that it is the only thing they desire as an end. Otherwise, given the principle, he would have no evidence for thinking, first, that happiness is desirable as an end, or, second, that it is the only thing desirable as an end. This is all pretty straightforward, though the principle of evidence is, of course, highly contentious.

In any case, the straightforward becomes much less so once it is appreciated that, as Mill would have it, what is desirable as an end is "the general happiness" -- that is, happiness no matter whose and so the happiness of each and every person, not just that of the agent. This complicates things immediately since, first of all, it seems obvious that many people do not desire the general happiness as an end and that, whether they do or not, they often desire also their own happiness as an end, so the general happiness is not the only thing they desire as an end. The premises Mill apparently thinks he needs, to have evidence for his doctrine, seem simply not to be available.

Mill, of course, is aware of this difficulty and never claims that people do actually desire the general happiness, let alone that it is the only thing they desire.¹⁰ Instead, he takes it that people do at least desire their own happiness and relies on this as establishing "that happiness is a good: that each person's happiness is a good to that person" and that the general happiness is "a good to the aggregate of all persons."¹¹ Similarly, after the third paragraph, he sets out to show that what people desire as ends they always desire as a part of their own, and not the general, happiness. Presumably, following the parallel, this is supposed to establish that happiness is the only thing that is good (as an end): that each person's happiness is the only thing that is a good to that person and

¹⁰ Nonetheless, some have criticized Mill for implicitly assuming that people, either individually or collectively, do desire the general happiness. For instance, F. H. Bradley suggests that "[e]ither Mill meant to argue, 'Because everybody desires his own pleasure, therefore everybody desires his own pleasure'; or 'Because everybody desires his own pleasure, therefore everybody desires the pleasure of everybody else.' Disciples may take their choice." Of course, the first claim is trivial and of no help and the second is obviously fallacious. Fortunately, it is pretty clear that Mill, whatever he meant, did not mean to offer either of these arguments. See F. H. Bradley, *Ethical Studies*, 2nd ed. (Oxford: Clarendon Press, 1927), 113-14, fn 3. Along the same lines, C. D. Broad sees Mill as committed to thinking that something being desired is one and the same with it being desirable, so that when Mill holds that the general happiness is a good that is desirable to the aggregate of all people, he is, according to Broad, committed to holding (falsely) that the aggregate desires something -- that is, the general happiness. Fortunately, again, it is pretty clear that Mill neither equated being desired and being desirable nor meant to hold that the aggregate desires anything at all. See C. D. Broad, *Five Types of Ethical Theory* (Paterson, NJ: Littlefield, Adams, 1959).

¹¹ Mill, *Utilitarianism*, IV, 3.10.

that the general happiness is the only thing that is good (as an end) to the aggregate of all persons.

In the post-third paragraph discussion, the main effort is given over to acknowledging and accommodating two facts. First, people often "desire things which, in common language, are decidedly distinguished from happiness."¹² Second, desire aside, a person can willingly pursue ends "without any thought of the pleasure he has in contemplating them, or expects to derive from their fulfillment" and can even knowingly act in ways where the benefits expected "are outweighed by the pains which the pursuit of the purposes may bring upon him."¹³

The first fact, Mill argues, raises no difficulty for his view, once properly understood. There are, he acknowledges, a number of things originally distinct from happiness that people can come to desire for their own sake. Yet when this change happens, when someone actually does come to desire money, say, or virtue, for its own sake, she has simultaneously become someone for whom acquiring money or acting virtuously is a pleasure. "Whatever is desired otherwise than as a means to some end beyond itself, and ultimately to happiness, is desired as itself a part of happiness, and is not desired for itself until it has become so."¹⁴ Admittedly, there are interesting and significant difficulties here surrounding Mill's psychological thesis. Specifically, he seems to run together the idea that people get pleasure from securing what they value for its own sake with the ideas of their valuing it for the pleasure that securing it brings.¹⁵ However, since my concern is to defend the structure of Mill's argument, not the truth of its premises, I will leave the psychological claim unchallenged.¹⁶

The second fact, Mill argues, likewise raises no difficulty for his view, since what matters to his argument is what people *desire*, not merely what they might be motivated to do. According to Mill, "Will, the active phenomenon, is a different thing from desire, the state of passive sensibility, and though originally an offshoot from it, may in time take root and detach itself from the parent

¹² *ibid.*, IV, 4.4.

¹³ *ibid.*, IV, 11.5.

¹⁴ *ibid.*, IV, 8.2. In this, Mill is largely following Aristotle, who identifies valuing something for its own sake with taking pleasure in it directly, rather than thanks to what it produces.

¹⁵ T. H. Green seems to be picking up on this point when he writes, "It is the realization of those objects in which we are mainly interested, *not the succession of enjoyments which we shall experience in realizing them*, that forms the definite content of our idea of true happiness, so far as it has such content at all" (emphasis added). I think, though, that Green goes astray in equating happiness with the realization of what is valued. See T. H. Green, *Prolegomena to Ethics* (Oxford: Clarendon Press, 1890), Bk. iii, Chap. iv, sec. 228, p. 244.

¹⁶ As I have said, the source of my less than full-hearted support of Mill's proof is largely found in my unwillingness to embrace the psychological claims upon which the proof turns.

stock."¹⁷ It is desire -- the passive sensibility -- that is supposed by Mill to afford evidence of value, not the will nor particular motives that serve as the springs of action. Mill is here marking an important difference between what people desire and what they are merely motivated to pursue. The difference will turn out to be crucial to understanding Mill's argument, and especially to understanding the way in which desire is supposed to be analogous to the senses.

These two facts taken account of, Mill thinks that the "practised self-consciousness and self-observation, assisted by observation of others" that I mentioned above will reveal that people do desire, as ends, all and only what constitute parts of their own happiness. This fact, he holds, provides the evidence, both necessary and sufficient, to establish that happiness, no matter whose, is desirable, and the only thing desirable, as an end. Before turning to the (more than half-hearted) defense of Mill's argument, let me briefly rehearse the familiar objections to it.

III. The Familiar Problems

Problems come up at the very beginning of the third paragraph when Mill sets to defending the principle of evidence upon which the rest of the argument turns. According to Mill, "The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experience. In like manner," he famously maintains, "the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it."

Right away there is reason to think this principle of evidence is exaggerated, even as applied to visibility and audibility. After all, we often have evidence that some particular thing is visible (or audible) even though we have never seen (or heard) it. Mill's actual view, and in any case all that he needs to rely upon, is that the ultimate source of all our evidence that something is visible (or audible) is found in what people see (or hear), whether or not the particular thing is itself seen (or heard).

Regardless, it has seemed to many that the argument for extending the principle of evidence from visibility to desirability utterly collapses once one notes that "desirable" may have either of two meanings: it might mean "capable of being desired" as "visible" means "capable of being seen," or it might mean "worth desiring" in a way that makes plausible an identification of it with "good" or "valuable." Read in the first way, the analogy with vision, hearing, and the other senses is apt and well supports the principle of evidence, but is irrelevant to the question of what is actually worth desiring. Yet the claim to be defended concerns what is worth desiring. Read in the second way, the

¹⁷ Mill, *Utilitarianism*, IV, 10.

principle of evidence, if true, would be relevant, but the analogy with our senses seems totally beside the point. Moreover, when read in the second way, the principle looks pretty clearly to be false. The mere fact that someone or other desires something seemingly provides us with no reason at all to think that what is desired is good. Thus, either the principle is well-supported but irrelevant, or relevant but totally unsupported -- and implausible as well. As Moore points out, "Mill has, then, smuggled in, under cover of the word 'desirable,' the very notion about which he ought to be quite clear. 'Desirable' does indeed mean 'what it is good to desire'; but when this is understood, it is no longer plausible to say that our only test of *that*, is what is actually desired."¹⁸

Putting aside the apparently irrelevant analogy with vision and the other senses, many (influenced by Moore) have thought the principle of evidence itself requires supposing that an evaluative concept can be defined in naturalistic terms, and so embodies the "naturalistic fallacy." Moore himself claims that Mill's reliance on the principle reveals "as naïve and artless a use of the naturalistic fallacy as anybody could desire." The fallacy comes, Moore urges, when Mill "pretends to prove that 'good' means 'desired'."¹⁹ Why Moore believes Mill makes any such pretensions, however, is a mystery. What Mill says is that "the *sole evidence* that anything is desirable is that it is desired" [emphasis added]. While that might be overstating the case -- there might be other sorts of evidence -- it is not at all to hold that being desirable and being desired are one and the same property, nor is it to hold that "desirable" and "desired" have the same meaning. Still, one might think, the idea that what people actually desire is the sole evidence we might have as to what is in fact valuable may seem to borrow all its plausibility from the assumption that being desired and being desirable are one and the same. If its plausibility does depend on that assumption, then the principle would seem to require identifying an evaluative property with a natural one. Whether this would be genuinely fallacious, of course, depends upon whether the identification on offer, if there is one, is correct. As I will argue, however, Mill is making no such assumption and has, instead, a different conception of the relation between desire and value.²⁰

In any case, suppose that one accepts the principle of evidence, at least for the sake of the argument. Even then, the third paragraph invites complaints and even ridicule. This is because Mill moves blithely from the observation that

¹⁸ G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), 67.

¹⁹ *Ibid.*, 66. Broad similarly claims that Mill "starts by assuming that 'desirable' means 'desired by someone'" in Broad, *Five Types of Ethical Theory*, 183.

²⁰ Specifically, on Mill's account, our concept of value and our evidence that the concept is satisfied are traceable to our capacity to desire, just as our concept of color and our evidence that the concept is satisfied are traceable to our capacity to see. Yet the concepts of value and of color are not concepts of the experiences that the capacities make possible.

people desire their own happiness as an end, which (in light of the principle of evidence) provides grounds for thinking that their own happiness is desirable as an end, to the required -- but still completely unsupported -- conclusion that *the general happiness* is desirable as an end. Given the principle of evidence, this conclusion would of course follow from the observation that people desire the general happiness as an end. Yet Mill does not rely on this (optimistic) observation.²¹ Mill instead apparently tries to move from the claim that each person's happiness is desirable as an end, to that person, to the conclusion that the general happiness (which is just the happiness of all) is "a good to the aggregate of all persons." As Sidgwick points out, though, "an aggregate of actual desires, each directed towards different parts of the general happiness, does not constitute an actual desire for the general happiness, existing in any individual."²² Thus, Mill could not in this way try to establish that the general happiness is the object of some individual's desire. Moreover, if Mill thought he only needed to show that someone or other desires the general happiness, his own case would have been enough. After all, he does desire the general happiness. However, Mill is clearly not trying to establish any such thing.

What is not so clear is why he thinks the value, to each, of his or her own happiness in any way goes to establish the value of the general happiness. Indeed, the reasoning seems to require one or both of the following: either an unjustified inference from the value (to someone) of each part of the general happiness to the value of the whole (which risks a fallacy of composition), or an unjustified inference from the value of the general happiness to all taken together (since each piece of the general happiness is valuable to someone) to the value of it to each, taken singly.²³

The worries multiply if we move on beyond the third paragraph. The remainder of the chapter inherits the above problems because it simply redeploys the third paragraph argument to show that happiness is the *only* thing desirable as an end on the grounds that it is the *only* thing desired as an end. It then adds to the difficulties quickly by embracing, apparently, either a simple-

²¹ Elijah Millgram argues that Mill is relying on the optimistic prospect that people in the future will desire the general happiness, and using this to support the claim that the majority of people (past, present, and future) give a decided preference to the general happiness even if people of his day do not. The success of the proof, on this view, turns on whether what is being "proven" ends up ultimately being accepted, and does not itself constitute any sort of argument for accepting it. See Elijah Millgram, "Mill's Proof of the Principle of Utility," *Ethics* 110, no. 2 (2000): 282-310.

²² Henry Sidgwick, *The Methods of Ethics*, 388. Broad presses the same point, at Broad, *Five Types of Ethical Theory*, 184.

²³ Thus, to say that the general happiness is "a good to the aggregate of all persons," on the grounds that each person's happiness is a good to that person, trades on ignoring the difference between "all" being used collectively (i.e., for all of us, taken together, all of it, taken together, is good) and its being used distributively (i.e., for each of us, considered individually, all of it, taken together, is good). See Alan Ryan, *John Stuart Mill*, (New York: Pantheon Books, 1970), 200-201.

minded version of psychological hedonism or a question-begging assumption concerning the connection between desiring something as an end and desiring it as a part of happiness (or both).²⁴

I propose to leave these additional difficulties to one side, though, because they are raised by Mill's attempt to establish the truth of one of his premises -- that people desire only happiness as an end -- and not by the argument's structure. It will be enough of a job to try to make sense of the overarching argument that gives point to Mill's trying to determine what people in fact desire.

IV. The "Proof" Again Part I, Desirability and Visibility

How is the argument supposed to go, if not by way of these multiple fallacies? Let us start with the principle of evidence and the analogy Mill draws between visibility and desirability. What is the analogy supposed to be if not one that commits Mill to interpreting "desirable" as "capable of being desired"?

When it comes to visibility, no less than desirability, Mill explicitly denies that a "proof" in the "ordinary acceptance of the term" can be offered.²⁵ As he notes, "To be incapable of proof by reasoning is common to all first principles; to the first premises of our knowledge, as well as to those of our conduct."²⁶ Nonetheless, support -- that is, evidence, though not proof -- for the first premises of our knowledge is provided by "our senses, and our internal consciousness." Mill's suggestion is that, when it comes to the first principles of conduct, desire play the same epistemic role that the senses play, when it comes to the first principles of knowledge.

To understand this role, it is important to distinguish the fact that someone is sensing something from what is sensed, which is a distinction

²⁴ These problems circulate around Mill's claim that "to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility" (Mill, Utilitarianism, IV, 10). I think it is pretty clear that Mill doesn't intend this claim to be true by definition, but it is hard to see why he thinks the empirical evidence would support it.

²⁵ Some, hoping to defend Mill, have jumped on this disclaimer to excuse Mill for arguments that appear invalid. If no proof is to be had, they note, then Mill must not be offering the considerations he does as deductively valid grounds for his conclusion. Still, the argument he offers had better be reasonable, and so far the argument under discussion seems to violate even that requirement by turning on ambiguities, implausible definitions, confusions between an individual and the group to which she may belong, and so on. Complicating matters, Mill (at least in A System of Logic) advances a distinctive view of what counts as a genuine proof, according to which deductively valid "proofs" are not proofs at all, since anything to be found in the conclusion of such a "proof" must already be present in its premises, and so is assumed rather than proven. Rather, he supposes, genuine proofs need to establish conclusions that go beyond their premises -- a proof establishes something new. John Stuart Mill, A System of Logic, in Mill, Collected Works, VII, Bk. II, Chap. I, sec. 2, 7: 58-162). Whether Mill thinks the "ordinary acceptance of the term" "proof" includes deductively valid arguments is unclear. See Millgram, "Mill's Proof."

²⁶ Mill, Utilitarianism, IV, 1.

mirrored in the contrast between the fact that someone is desiring something and what is desired. In the case of our senses, the evidence we have for our judgments concerning sensible qualities traces back to what is sensed, to the content of our sense-experience. Likewise, Mill is suggesting, in the case of value, the evidence we have for our judgments concerning value traces back to what is desired, to the content of our desires. Ultimately, the grounds we have for holding the principles we do must, he thinks, be traced back to our experience, to our senses and desires. Yet the evidence we have is not that we are sensing or desiring something but what it is that is sensed or desired.²⁷

When we are having sensations of red, when what we are looking at appears red to us, we have evidence (albeit overrideable and defeasible evidence) that the thing is red. Moreover, if things never looked red to us, we could never get evidence that things were red, and would indeed never have developed the concept of redness. Similarly, when we are desiring things, when what we are considering appears good to us, we have evidence (albeit overrideable and defeasible evidence) that the thing is good. Moreover, if we never desired things, we could never get evidence that things were good, and would indeed never have developed the concept of value.²⁸

Recall that desire, for Mill, like taste, touch, sight, and smell, is a "passive sensibility." All of these, he holds, provide us with both the content that makes thought possible and the evidence we have for the conclusions that thought leads us to embrace. "Desiring a thing" and "thinking of it as desirable

²⁷ The interpretation I am advocating thus differs from those who suggest that Mill is appealing to the fact that people desire happiness as an end as establishing that it is *possible* to desire happiness as an end. According to these accounts, the appeal to what people desire plays out against the assumption that what we ought to desire is constrained by what we can desire, and is meant to show that happiness can be desired and is the only thing that can be desired as an end. On this interpretation, happiness emerges as the only candidate for being what ought to be desired. Mill's own observation that "[i]f the end which the utilitarian doctrine proposes to itself were not, in theory and practice, acknowledged to be an end, nothing could ever convince any person that it was so" is often taken as supporting this interpretation. This interpretation, though, leaves Mill with no positive argument for thinking happiness is good -- it would simply have survived as the only candidate. Moreover, the argument has not established even this much by the end of the third paragraph, where Mill infers that happiness has been shown to be at least among the things that are desirable as an end. Thus, this interpretation has trouble both with finding Mill a positive argument for thinking happiness is valuable and with making sense of the structure of the text. For discussions of this interpretation, see James Seth, "The Alleged Fallacies in Mill's 'Utilitarianism,'" *Philosophical Review* (1908): 469-88; R. F. Atkinson, "J. S. Mill's 'Proof' of the Principle of Utility," *Philosophy* (1957): 161-62; Norman Kretzman, "Desire as Proof of Desirability," *Philosophical Quarterly* (1958): 246-58; and George Clark, "Mill's Notorious Analogy," *Journal of Philosophy* (1959): 652-56. In contrast, on my reading, Mill's point here is not guided by 'ought' implies 'can' but instead, as I will suggest, by an appreciation of the fact that evaluative conclusions require evaluative premises (i.e. no 'ought' from only 'is's).

²⁸ Whether the experience is of something as red, or of it as good, the evidence we have is constituted by the content of our experience, not the fact that we had the experience.

(unless for the sake of its consequences)" are treated by Mill as one and the same, just as seeing a thing as red and thinking of it as red are one and the same.²⁹ Accordingly, a person who desires x is a person who *ipso facto* sees x as desirable.³⁰ Desiring something, for Mill, is a matter of seeing it under the guise of the good.³¹ This means that it is important, in the context of Mill's argument, that one not think of desires as mere preferences or as just any sort of motive. They constitute, according to Mill, a distinctive subclass of our motivational states, and are distinguished (at least in part) by their evaluative content.³² Thus, Mill is neither assuming nor arguing that something is good because we desire it; rather, he is depending on our desiring it as establishing that we see it as good.

At the same time, while desiring something is a matter of seeing it as good, one could, on Mill's view, believe that something is good without desiring it, just as one can believe something is red without seeing it as red. While desire is supposed to be the fundamental source of our concept of, and evidence for, desirability, once the concept is in place there are contexts in which we will have reason to think it applies even when the corresponding sensible experience is lacking. Indeed, in Chapter IV, Mill is concerned not with generating a desire but with justifying the belief that happiness is desirable, and the only thing desirable, as an end, and so concerned with defending the standard for determining what should be desired.³³

²⁹ Mill, Utilitarianism, IV, 10.10. Here Mill first equates desiring something with thinking it pleasant, and then he equates thinking of it as pleasant with thinking of it as desirable. (He actually writes "finding it pleasant" rather than "thinking it pleasant," but goes on immediately to equate aversion to a thing with "thinking of it as painful," not "finding it painful," so in this passage he seems clearly to be using "finding it" and "thinking of it" interchangeably.)

³⁰ Within this broadly empiricist approach to concepts and evidence, there is an important question of how to understand the original experiences that give rise to concepts and provide evidence for their application. These experiences are often characterized not only by using the concepts they are supposed to explain, but in a way that seems to suppose the concepts are already available to the person having the experience. This obviously needs to be avoided on pain of circularity. In the case of desire and value, then, Mill needs to suppose that while desiring x is something like seeing x as desirable or good, the experience of desiring x does not require already possessing the concepts of desirability or value.

³¹ Views similar to Mill's in this respect (though not others) have recently been defended in Dennis Stampe, "The Authority of Desire," *Philosophical Review* 96 (1987): 335-81, Warren Quinn, "Putting Rationality in Its Place," in Quinn, Morality and Action (Cambridge: Cambridge University Press, 1993), and T. M. Scanlon, What We Owe to Each Other (Cambridge, MA: Harvard University Press, 1998).

³² While he believes that "[i]n what we call Desire there is... always included a positive stimulation to action," he also thinks there is always included (to use his father's phrase) "the idea of something good to have." See James Mill, Analysis of the Phenomena of the Human Mind 2nd ed., Chap. XIX, and John Stuart Mill's notes to that edition, *ibid.*, 194-95. The notes can also be found in Mill, Collected Works, XXXI:215.

³³ In some cases, of course, Mill thinks that the correct standard will recommend desiring something other than the general happiness. At the same time, there is, on Mill's view, an important difference

Mill's aim is to take what people already, and he thinks inevitably, see as desirable and argue that those views commit them to the value of the general happiness (whether or not their desires follow the deliverances of their reason). Those who, like Mill, desire the general happiness already hold the view that the general happiness is desirable. They accept the claim that Mill is trying to defend. As Mill knows, however, there are many who do not have this desire -- many who desire only their own happiness, and some who even desire that others suffer. These are the people he sets out to persuade, along with others who are more generous and benevolent, but who nonetheless do not see happiness as desirable, and the only thing desirable, as an end. Mill's argument is directed at convincing them all -- whether their desires follow or not -- that they have grounds for, and are in fact already committed to, regarding the happiness of others as valuable as an end.

Mill recognizes that whatever argument he might hope to offer will need to appeal to evaluative claims people already accept (since he takes to heart Hume's caution concerning inferring an 'ought' from an 'is').³⁴ The claim Mill thinks he can appeal to -- that one's own happiness is a good (i.e. desirable) -- is something licensed as available by people desiring their own happiness. Yet he is not supposing here that the fact that they desire their own happiness, or anything else, is proof that it is desirable, just as he would not suppose that the fact that someone sees something as red is proof that it is. Rather, he is supposing that if people desire their own happiness, or see something as red, one can rely on them having available, *as a premise for further argument*, the claim that their own happiness is desirable or that the thing is red (at least absent contrary evidence).³⁵ As he puts it in the third paragraph, "If the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end nothing could ever convince any person that it was so."

Thus, in appealing to the analogy between judgments of sensible qualities and judgments of value, Mill is not trading on an ambiguity, nor does his argument here involve identifying being desirable with being desired or assuming that "desirable" means "desired." He is instead relying consistently on an empiricist account of concepts and their application -- on a view according to which we have the concepts, evidence, and knowledge we do only thanks to our having experiences of a certain sort. In the absence of the relevant experiences, he holds (with other empiricists), we would not only lack the

between justifying a belief -- giving evidence of its truth -- and showing that it would be good to adopt it.

³⁴ As is made clear by his discussion of it in Mill, A System of Logic, Bk. VI, Chap. XII, Sec. 6, VIII:949-52.

³⁵ Note that even in the presence of contrary evidence, the content of one's desires or visual experiences will be on board as potential counterevidence.

required evidence for our judgments, we would lack the capacity to make the judgments in the first place. In the presence of the relevant experiences, though, we have both the concepts and the required evidence -- "not only all the proof which the case admits of, but all which it is possible to require."³⁶

V. The "Proof" Again Part II. From Each to All, with Kant in Tow

Needless to say, even if Mill can legitimately invoke the analogy between vision and desire, and so appeal to desire in establishing desirability, he is a long way from having given any argument at all for thinking that the *general* happiness is desirable. Mill's first step in offering such an argument, of course, is his claim that, as a matter of fact, people desire their own happiness -- and so see their own happiness as good. This fact (assuming it is one) means that we can rely on this view being available to people when one offers them an argument for thinking that the general happiness is desirable. In the same way, the fact that someone has a certain sort of visual experience means that she sees her car as red, and we can rely on this view being available to her in offering her arguments about the color of other things, or of the car, should it turn out to be someone else's.

Significantly, beginning with the claim that each person desires her own happiness, Mill infers that *each person's* happiness is a good *to that person*, and the general happiness is a good *to the aggregate of all persons*. Thus, as Mill presents the case, there are two proprietary aspects of the situation: first, the happiness in question is supposed to be someone's; second, the value it is supposed to have is a value to someone. As it turns out, though, the conclusion Mill ultimately wants to establish is proprietary along neither dimension: he wants to show that happiness, *no matter whose*, is valuable *simpliciter*, not just valuable *to* one person or another or even *to everyone*.

That Mill is out to show that happiness, no matter whose, is valuable, is clear; "the happiness which forms the utilitarian standard of what is right in conduct, is not," Mill notes, "the agent's own happiness, but that of all concerned."³⁷ That he also is trying to show that it is valuable *simpliciter*, as opposed to valuable to each and every person, is much less clear. Indeed, it is tempting to think he has not successfully responded to egoists and so not defended the general happiness as valuable as an end in the relevant way until he has shown that it is a good to each person.

³⁶ When he writes this, in the third paragraph, Mill is just summarizing the case for thinking that happiness is good, not yet that it is also the only thing good, as an end. However, the stronger conclusion is supposed to rest simply on a redeployment of the same argument form. See Mill, Utilitarianism, IV, 9.5.

³⁷ *Ibid.*, II, 18.2.

So it is worth noting, first, that Mill explicitly denies that the general happiness is, in fact, a good *to each person*. Then, second, it is worth trying to figure out why he denies it. The denial is clear and comes up when, in a letter to Henry Jones, Mill writes:

[W]hen I said the general happiness is a good to the aggregate of all persons *I did not mean that every human being's happiness is a good to every other human being*, though I think, in a good state of society and education it would be so. I merely meant in this particular sentence to argue that since A's happiness is a good, B's a good, C's a good, &c., the sum of all these goods must be a good. [emphasis added]³⁸

Here Mill is doing three things: (i) saying that he was not trying to show that the general happiness is a good to each and every human being, (ii) expressing his view that things would be better if the happiness of others were a good to each, and (iii) claiming that if each person's happiness is a good (note: a good *simpliciter*, not a good *to that person*) then the happiness of each, taken together, must be a good as well. Appreciating all three points is important to understanding Mill's view, but for now I will focus on the first claim.

What is it for something to be a good *to a person*? Why does Mill think that his defense of utilitarianism can succeed without showing that the general happiness is a good to each person? The answer to the first question is, in broad outlines at least, pretty simple. On Mill's view, and plausibly, for something to be good *to a person* is for it to be a good that the person gets, a benefit she receives, enjoys, or secures. Thus, if happiness is a good, for it to be a good *to a person* is for that person to get it, that is, be happy. If she is concerned about the welfare of another, and that other person's happiness is (as Mill supposes) a good, then when the other person is happy, that is a good to the person who is happy, and a good to her as well insofar as she too is made happy. If, alternatively, she is unconcerned about the welfare of another, but (again as Mill supposes) that other person's happiness is nonetheless a good, then when the other person is happy, that is a good to the person who is happy, even as it is not a good to the one who is unconcerned.³⁹

³⁸ John Stuart Mill to Henry Jones, June 13, 1868, reprinted in Mill, Collected Works, XVI:1414.

³⁹ Roger Crisp appeals to the passage just quoted from Mill's letter to Henry Jones to argue that Mill does not mean to show that the general happiness is an appropriate end for each individual. According to the interpretation I am offering, however, Mill *is* trying to give each of us an argument for accepting the view that the general happiness is valuable as an end. That the letter does not deny this becomes clear once attention is paid to the difference between showing that something is a good and showing that it is a good to someone. To do the latter is, on Mill's view, to show that it will contribute to *that person's happiness*. In a better world, Mill notes, people's concerns would be

There is more than a little difficulty in the offing, though, when it comes to making sense of what it is for someone to receive, enjoy, or secure a benefit. It is not merely a matter of a person getting something that is good, since that might happen in a way that leaves the person herself no better off, even as, perhaps, she is as a result surrounded by more that is good. For her to benefit, for what she receives to be a good to her, it seems she must herself be better off thanks to its presence. Mill's view, I think, is that this happens exactly when the good either partially constitutes, or brings, happiness (or pleasure) to the person in question.⁴⁰ In any case, that something is a good *to* someone presupposes (on Mill's view) that it is good *simpliciter* – so if, say, happiness is a good to someone who has it, that presupposes the value of what it is she gets.

Needless to say, if happiness is a good, then the value of one person being happy can be compounded by others caring about her. That way, when she thrives, they do too.⁴¹ That is one reason why Mill thinks that, in light of the value of happiness, things would be better if society and education worked so that one person's happiness was "a good to every other human being." Another is that to the extent people are motivated by what they value as ends (and so as parts of their own happiness), if one person's happiness is a good to another, the second person will more likely be motivated to promote the happiness of the first.

In point of fact, of course, some people are made unhappy when others do well. Even such people (assuming they desire their own happiness) are, according to Mill, committed to the value of other people's happiness, despite the other people's happiness not being a good to them. Yet, for all Mill argues, it is in principle possible that things would be best if one person's happiness were not always a good to another. Competition for goods may well be advantageous when it comes to the production of value. Hence, while Mill

structured so that they would take pleasure in the well-being of others. Until that happens, though, we need to acknowledge that while the well-being of others is in fact good, it does not always contribute to the well-being of an agent in the way that would underwrite saying that the well-being of another is a good *to that agent*. Mill wants to offer an argument that shows that when something makes someone happy, we are all constrained to admit that something good has been produced, whether or not we take it to be good and whether or not we ourselves are benefited by it (which we will be if we care for the other person's welfare, but may not be otherwise). See Roger Crisp, *Mill On Utilitarianism* (London: Routledge, 1997), 77-78.

⁴⁰ Given Mill's views, everything that contributes to the happiness of the person in question will thereby count as good (either in itself or as a means), but not everything that is good will contribute to the happiness of that person. At the same time, what a person values (what is good according to that person) may or may not, on Mill's account, actually be good; whether it is depends upon whether it contributes to someone's happiness -- either to the happiness of that person or someone else's. Also, what might otherwise not be good may be made so by the person valuing it (though only if, in fact, it brings her happiness).

⁴¹ At the very least, they benefit. Of course, if unhappiness is bad, then the costs of one person being unhappy will be compounded by others caring about her.

believes overall value would be advanced by people caring about one another's happiness, that belief assumes not only that happiness is valuable, but also that mutual concern would promote happiness. If mutual concern did not, in fact, promote overall happiness, Mill would oppose it on exactly the same utilitarian grounds he relies on in its defense. His central concern is not with whether people should care about each other's happiness, but with whether happiness is a good in itself, and so desirable as an end.

The last of the three claims Mill makes in the letter -- that if A's happiness is a good, B's is, and C's is, then their happiness taken together is a good as well -- is, as I have mentioned, generally thought to turn on a fallacy of composition. Clearly, if the initial claims were that A's happiness is a *good to* A, B's a *good to* B, and C's a *good to* C, there would be no grounds for thinking that A's happiness, combined with B's, combined with C's, would be *good to* anyone; yet it would still be plausible to think it good to the group, which is all that Mill claims in giving the proof. Whether there would be grounds for thinking that the fact that each person's happiness is good *simpliciter* (as opposed to good-to-someone) shows that the claim that "the sum of all these goods must be a good" depends upon what sort of feature the value of happiness is supposed to be.

Commonly, worries about the move from the value of each person's happiness to the value of the general happiness -- happiness, no matter whose -- focus on Mill's claim that the general happiness is a good to the aggregate of all people. This claim, many assume, is what Mill sets out to establish in order to defend utilitarianism. Yet given Mill's understanding of what it is for something to be a *good to* someone, this claim turns on his already having established what is at issue: that the general happiness is a good in itself. Thus, when Mill writes that "happiness is a good: that each person's happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of persons," the key claim is what precedes the colon: that "happiness is a good." What follows the colon are corollaries of the key claim, not what is primarily at issue.

The key claim, in Mill's eyes, is established, in light of the principle of evidence, by each person desiring her own happiness. Redeploying the same argument form in the rest of the chapter, the claim that happiness (no matter whose) is the only thing desirable as an end is supposed to be established, in light of the same principle, by each person desiring only her own happiness as an end. Analogous corollaries follow: that each person's happiness is the only thing good, as an end, to that person, and that the general happiness is the only thing good, as an end, to the aggregate of all persons.

In thinking about the value of happiness, it is important to mark the difference between what a person might have reason to do and what she might have reason to think is good. Mill seems prepared to acknowledge that a person

has reason to do only what contributes to her own happiness -- what is a good to her. That is, at least in part, why morality calls for sanctions and why it recommends working to make it so that people value the welfare of others (so that when others do well, one benefits also).

Recognizing the difference between what people have reason to do, and what they have reason to think, is especially important when it comes to understanding about what Mill is trying to prove, and what he is not trying to prove, in Chapter IV. On the one hand, he is trying to convince someone who thinks of her own happiness as being the only thing that is valuable that she is mistaken. He is arguing, using that person's own commitments, that happiness, no matter whose, is valuable. That is, he is offering her reason to think that other people's happiness is good. On the other hand, he is not trying to convince people that they are ever *rationally* required to sacrifice their own happiness for the happiness of others -- only that they are sometimes *morally* required to do so. He is addressing those who wonder whether other people's happiness matters that it is valuable, even as he never tries to convince them that it matters to them, that is, that the happiness of others is *a good to them*.

Focusing on the claim that happiness (no matter whose) is a good, rather than that the general happiness is a good to each or a good to the aggregate of all persons, does not, of course, remove the central puzzle concerning the move from the desire for one's own happiness to the desirability of happiness (no matter whose). After all, it sure seems as if a person's desire for her own happiness commits her, at most, to the value of her own happiness, not to the value of happiness *per se*, just as seeing one's own car as red commits one to its redness, but not to the redness of anyone else's car. Still, focusing on the claim that the general happiness is a good does help to bring out that Mill is not trying to argue that the general happiness is a good to each and every person.

If this is the case, however, how is the move from the desire for one's own happiness to the desirability of happiness (no matter whose), supposed to go? A parallel question arises, to bring Kant in now, when Kant turns to defending the principle that people ought to treat not just themselves, but all other rational beings as well, as ends and not merely as means. According to Kant,

The ground of such a principle is this: rational nature exists as an end in itself. In this way man necessarily thinks of his own existence; thus far is it a subjective principle of human actions. But in this way also does every other rational being think of his existence on the same rational ground that holds also for me; hence it is at the same time an objective principle, from

which, as a supreme practical ground, all laws of the will must be able to be derived.⁴²

Kant here is assuming that people all see themselves as valuable as ends, and assuming too that this view is (rationally) grounded by their thinking that "rational nature exists as an end in itself." With these assumptions in place, he is then arguing that the very reason people have for valuing themselves as ends -- that the rational nature they possess exists as an end in itself -- commits them to the value as ends of all who possess a rational nature.⁴³ Similarly, I believe, Mill is assuming that people all see their own happiness as valuable as an end, and assuming too that this view is (rationally) grounded on their seeing happiness as valuable in itself. With these assumptions in place, he is then arguing that the very reason people have for valuing their own happiness -- that the happiness they would enjoy is good in itself -- commits them to the value of happiness no matter who happens to enjoy it.⁴⁴

In both cases, the argument turns on the rational ground people (supposedly) have for their commitments. Hence, it is worth trying to bring out what is distinctive about the sort of grounds Kant and Mill say that people use to underwrite their valuing, whether it be the valuing of their own happiness, their rational nature, or whatever. To bring out what is distinctive, imagine a rich, luscious, moist, chocolate cake (which, I will be assuming, is a lot like happiness, at least in its being desirable). The cake may be predivided into proprietary slices -- one is mine or mine to have, another is yours, etc.-- or the pieces of cake may not be anyone's in particular, even if one person's getting some slice means that no one else can get that slice (and even if some may end up with none).

In the first case, in desiring a piece of cake I might specifically be desiring only my own piece and not someone else's (though of course I might also be desiring yours). In the second case, while I might selfishly desire certain pieces for myself, those pieces are not originally mine and may end up being someone else's. In the first case, but not the second, *that the piece is mine (or mine to have)* may be my reason for desiring it, and I may, in desiring it, be seeing *my* piece or my having the piece as valuable without additionally having, or being committed to, any views about other people's pieces. In the second

⁴² Kant, *Grounding*, 36 (AK 429).

⁴³ Needless to say, one might doubt that all people see themselves as Kant supposes they do, or doubt that those who do are relying on the grounds he supposes. Still, if he is right about these commitments, his argument kicks in directly.

⁴⁴ Needless to say, one might doubt that all people see their own happiness as Mill supposes they do, or doubt that those who do are relying on the grounds he supposes. Still, if he is right about these commitments, the argument kicks in directly.

case, in contrast, that the piece is mine will (by hypothesis) not be in play. Instead, whatever it is about the cake that I am desiring, and so desiring for myself, is something the cake may still have if someone else were to get it.⁴⁵ In seeing the cake as valuable (and so worth getting for myself), I therefore seem to be committed to thinking that if you should get it rather than me, you have gotten something good. Of course, I am not committed to desiring that you get it, nor to thinking that your getting it is a good to me. However, the grounds I have for thinking it would be a good to me, were I to get it, appear to commit me to thinking that you would be getting something good were you to get it.

It is important here to keep in mind the contrast between my desiring the pleasure I might get from the cake and my desiring the cake.⁴⁶ If it is the pleasure I desire (for myself), then my commitment *vis-à-vis* you would be to seeing your getting a similar pleasure as your getting something good. If the cake does not bring you that pleasure, then in getting a piece of the cake you would not be getting what I take to be good about it, and the difference could consistently be seen by me as making all the difference. What I cannot consistently do is see as valuable some feature of the cake, value getting the cake for myself on those grounds, and then deny that when you get it (with that feature), you get something good.

Analogously, if each of us is, in desiring happiness, desiring not merely *our own* happiness, but desiring happiness (for some nonproprietary feature of it) for ourselves, we cannot consistently then deny that when someone else gets happiness (with that feature), they get something good. Of course, again, we are not committed to desiring that someone else get it, nor to thinking that their getting it is a good to us. Nevertheless, the grounds we each have for thinking it would be a good to us, were we to get it, appear to commit us each to thinking that in getting it, someone else would be getting something good. Analogously too, if the grounds I have for thinking of my rational nature as an end are not originally proprietary, if the features they appeal to are features that your

⁴⁵ In this second case there may well be two desires in play, one depending on the other, with the first being, in effect, a matter of seeing the cake as good, and the second being a matter of seeing getting the cake as good because it is the getting of something good. The second desire, of course, does not always follow upon the first, for a number of reasons. For instance, it may be that what one sees as good about the cake is something one cannot get, or it may be that one could get it, but only at the expense of someone else's not getting the cake, where that other person's getting the cake would be better. Note also that the second desire might be present without the first, since all the second requires is that one accept (perhaps as the content of a belief) the value of what it is that one desires for oneself.

⁴⁶ It is also important here to keep in mind the contrast between my desiring *my own* pleasure, which I might get from the cake, and my desiring pleasure (conceived of in nonproprietary terms) for myself. The former puts us back in the first case, where the relevant objects of desire come in proprietary packets, whereas the latter simply shifts from a case where it is the cake that is seen as valuable to one in which the pleasure is.

rational nature has no less than does mine, then in seeing my own as valuable as an end on those grounds, I am committed likewise to the value of your rational nature on exactly the same grounds.

Needless to say, so far nothing has been said in defense of the view that we desire our own happiness on grounds that are nonproprietary in the relevant sense, nor has any case been made for thinking that we view ourselves as ends on grounds that are appropriately nonproprietary. Although I will say something on Mill and Kant's behalf, in defense of the nonproprietary character of the grounds of our evaluative commitments, my concern is not to defend either Mill or Kant's substantive premise concerning our fundamental (and in Kant's case, necessary) evaluative commitments. Rather, my concern is to bring out the general force of the argument form they are offering, not the specific applications they present.

In thinking through Mill's version of the argument, and keeping the cake example in mind, we need to distinguish between two ways people might be thinking of happiness -- one proprietary, the other not -- when they are desiring their own happiness. It might be that they desire their own happiness, where the happiness is being conceived of *as theirs*, as coming already divided up into proprietary packets that either belong to them or are something they will get. Alternatively, it might be that they desire happiness for themselves, where the happiness is being conceived of originally in terms of what it is like independent of who has it. In a similar way, in thinking through Kant's version of the argument, we need to distinguish between two ways people might be thinking of their own rational nature. It might be that they think of their own rational nature as existing as an end, where the rational nature is being conceived of *as theirs*, as coming already divided up into proprietary packets. Alternatively, it might be that the rational nature they think of as existing as an end is being conceived of originally in terms of what it is like independent of whose it is.

These differences make a big difference to the content of the evaluative position people are, from the start of each argument, supposed to have concerning the value of one's own happiness or rational nature. These initial evaluative positions are what Mill and Kant, each in his own way, set out to leverage into a commitment concerning the value of other people's happiness or rational nature. Predictably, how their arguments might go depends a great deal on what the initial evaluative position is supposed to be and whether what is valued is conceived of in proprietary terms or not. With this in mind, three distinct arguments are apparently in the offing, any of which might be attributed either to Mill or to Kant. The first two take the evaluative starting point to be proprietary, whereas the third does not. (For ease of exposition, I focus primarily on Mill's suggestion that the commitments in question would concern

happiness (either one's own, or *simpliciter*); the same arguments could, on Kant's behalf, be made substituting in a commitment to the value of rational nature (either one's own, or *simpliciter*.)

According to one argument, the evaluative starting point is that each thinks "my own happiness is valuable," and so each has reason to think of others -- given that others are in a parallel situation with respect to their own happiness -- that they similarly think that their own happiness is valuable. Our positions, with respect to our own happiness, appear to be perfectly symmetrical. This, it might seem, puts pressure on anyone who would presume to think that he, and his happiness, stand out from the crowd. The symmetry, assuming it is in place, does establish that each person has as much reason as any other to think of her own happiness as valuable.⁴⁷ However, this does not translate into a reason for each to think of the happiness of others as valuable. Indeed, the argument does not even begin to give someone reason to believe that anyone else's happiness is valuable. The argument relies on our being symmetrically situated with respect to our thinking *as we do*, and not with respect to *what we are thinking* -- since we are each seeing our own happiness, and not each other's, as valuable. Thus, the argument depends not at all on the content of the original evaluative premise, and so does not move to a conclusion informed by that content. Rather, it shows that we have to admit that, if others bear the same relation to their happiness that we bear to our own (that of "thinking it valuable"), then they, like us, will have grounds for thinking that their own happiness is valuable that parallel those we have for thinking our own is. Yet granting this is perfectly compatible with denying the value of others' happiness. Therefore, the symmetry considerations do not establish that people are committed to acknowledging the value of other people's happiness or the value of their rational natures. As a result, it is pretty clearly not the argument Mill has in mind. For a similar reason, it is pretty clearly not the sort of argument Kant has in mind either. At most it would establish, with suitable adjustments to the supposed initial evaluative position, that we each have reason to think of others - - given that others are in a parallel situation with respect to their own rational natures -- that they similarly *think of* their own rational natures as ends, whereas the principle Kant is trying to establish would have it that rational nature, no matter whose, *is* an end.⁴⁸

According to the second argument, the evaluative starting point is again each person thinking "*my own* happiness is valuable," but this fact about each person is taken *as evidence*, with respect to each bit of happiness that is valued,

⁴⁷ That is, as much reason as any other if the only reason anyone has is found in her own happiness appearing to her to be good.

⁴⁸ If the rational ground we each have for thinking of our own rational nature as an end has this standing turn, in each case, on the rational nature being *one's own*, then we would each only be committed to holding that for each person, his or her rational nature is an end for him or her.

that that bit is valuable. Each person is seen as having reason to think that the happiness she enjoys is valuable, and reason to think of others -- given that they are in a parallel situation with respect to the happiness they enjoy -- that each person's happiness is such that there is the same evidence available to each for the value of the happiness that another person enjoys as there is for the value of one's own happiness. If happiness is such that every piece of it is desired by someone, then it seems as if, in taking ourselves to have reason to see the bit we value as valuable, we are committed to acknowledging the value of all the rest. Analogously, again with suitable adjustments, the argument in Kant's hands would take as an evaluative starting point each having reason to think of her own rational nature as an end, and reason to think of other's -- given that they are in a parallel situation with respect to their own rational nature -- that each person's rational nature is such that there is the same evidence available to each for the value of another person's rational nature as there is for the value of one's own. If every rational being is such that her rational nature is thought of by her as an end, then it seems as if, in taking ourselves to have reason to see our own rational nature as an end, we are committed to acknowledging that other people's rational natures are likewise to be seen as ends. Unlike the first argument, this one does, if successful, move to an appropriate conclusion. Of course, its success depends on each person having reason to think of her own happiness, or rational nature, as valuable, as an end.

A structurally similar argument might be offered in response to someone who sees part of an elephant and wonders whether there is a whole visible elephant before her. If someone or other stands in relation to each part of the elephant in the same way that she stands to the part she takes herself as having reason to think is visible -- because she seems to see it -- then, if she knows of the others' situation, she has reason to think that they have an equally good reason to think of each part they (seem to) see as being visible. If each and every part of the elephant is such that someone (seems to) see it, then she has reason to think the whole elephant is visible even though she herself only sees a part of it.

We can use this elephant example to see the weakness of the second argument. The others do, I think, each have a reason as good as hers (because the same) for thinking that what appears visible to them is visible. This does not mean, however, that the reason they each individually have is equally a reason for her. After all, the force of the reason is supposed to depend, in the elephant case, on the person having a visual experience of the appropriate kind. Each has the same sort of visual experience with respect to some part or other of the elephant, but the person in question does not have that experience with respect to the parts she does not (seem to) see -- thus, whatever evidence *she* has of those parts is not itself visual evidence of the kind in question.

In a similar way, when it comes to happiness, each person may have a reason as good as another's (because the same) for thinking that what appears valuable is valuable. This does not, however, mean that the reason they each individually have is equally a reason for others. While we might grant that we each have evidence concerning the value of our own happiness (thanks to it being such that we each find our own happiness to be good), there is room to worry that the fact that someone else sees something as good (that is, desires it) neither constitutes nor provides evidence for me that it is. In the case of happiness, the force of one's reason for thinking that the general happiness is valuable is supposed to depend on the person addressed by the argument having a desiderative experience of the appropriate kind. Each person has the same sort of desiderative experience with respect to some part or other of the general happiness, but people will not be having that experience with respect to the parts they do not desire -- thus, whatever evidence one has concerning another's happiness is not itself desiderative evidence of the kind in question. The same point carries over to the grounds one might have for viewing one's own rational nature as an end.

Across the board, the argument brings out that, if others stand in the same relation to their own view of the elephant, or their own happiness, or their rational nature, as one stands to one's own, then one is committed to acknowledging that they have the same kind of reason to think as they do as one has for one's own view. However, this argument leaves those to whom it is addressed with no reason, so far, to think that other people's happiness is valuable, nor any reason to think that their rational natures are ends, even when one does have reason to think that of one's own happiness or rational nature. Moreover, even if we assume that other people's experiences can provide us with the appropriate evidence concerning the value of what appears good to them (by assumption, their own happiness), we would have evidence that happiness, no matter whose, is valuable only if everyone actually values (i.e., desires) his or her own happiness. Furthermore, we would somehow have to take account of the fact that some people desire that others not be happy, which, on this argument, would seem to count as evidence that other people's happiness is not good after all.

According to the third argument, the one I think Mill is actually offering, the evaluative starting point is that people (in desiring happiness) are thinking "Happiness is valuable" and, on that basis, wanting it for themselves. It takes this fact not as new evidence that might be added to or balanced against other evidence, but as reflecting an evaluative commitment each person already has on board. Then, in light of that commitment, the argument points out to each person that if what is valuable (on one's own view) is the happiness one wants to enjoy, then when someone else happens to get happiness, one is committed already to the value of what the other person gets. The argument

thus appeals to something each person shows herself to accept in her desiring happiness for herself, and plays out what this implies -- namely, that whether or not one happens to desire that other people be happy, if they have some share of happiness then they have something of value. Analogously, again with suitable adjustments, this third argument, in Kant's hands, starts with people valuing their own rational natures as ends on the grounds that "Rational nature *per se* is an end." It takes this fact not as new evidence, but as reflecting an evaluative commitment each person already (and, according to Kant, necessarily) has on board. Then, in light of that commitment, the argument points out to each that if others similarly possess a rational nature, one is committed already to their standing as ends.

Significantly, this argument (whether in Kant's hands or Mill's) is directed at a broad audience: (i) those who are selfish and act as though they or their happiness is the only thing of value; (ii) those who are skeptical and wonder why they should think someone else matters, given that they do not happen to care; and (iii) those who are committed to morality, but are unclear what is of value. At the same time, while the argument does rely on appealing to people's already-present commitment either to the value of the happiness they hope for or to their standing as ends insofar as they are rational, these suppositions concerning people's commitments are each widely accepted and not at all implausible.

As Sidgwick pointed out, however, this third argument, in depending on the idea that people have such nonproprietary commitments, leaves unaddressed those who hold that at bottom what is of value is *their own* happiness, not happiness *per se*, or that it is *their own* rational nature, not rational nature *per se*, that is an end.⁴⁹ Such people do not have the sort of initial commitment that the argument needs in order to extend beyond the case of the person to whom it is being addressed. Thus, a person with the appropriately proprietary fundamental desire for happiness, or conviction concerning her own rational nature, can perfectly consistently deny the value of other people's happiness and their standing, in virtue of their rational natures, as ends.

Mill and Kant might just have overlooked this possibility. However, I suspect that they thought there was something evidently untenable about having a proprietary evaluative commitment that is not underwritten by an evaluative commitment that is not proprietary. In any case, neither Mill nor Kant takes the possibility seriously, and it is not very difficult to find an argument that might back up their casual neglect of this view.

In valuing something (my happiness or my rational nature, say), there must be something I see as good about it. Whatever that feature is, it cannot be

⁴⁹ Sidgwick, *The Methods of Ethics*, 420.

simply it being mine or my getting it, since obviously plenty of things that are mine, or that I do get, are not valuable at all. However, whatever other feature it might be will be a property potentially possessed by things that are not mine. For instance, if what is good about my happiness (according to me) is how it feels, then I am committed to thinking that this same feeling, if enjoyed by someone else, is good as well; likewise, if what is good about my rational nature (according to me) is the autonomy that it makes possible, then I am committed to thinking that same autonomy, if enjoyed by someone else, is good as well. Of course, this leaves room for the person who thinks that the crucial property is not simply *x being mine* but *x being my happiness* or *my rational nature*. If someone holds such a view, though, it is natural to ask her, What is special about your happiness or your rational nature— if it is not merely that it is yours (which we have ruled out) and not some general feature of it that might be shared by someone else's? The point is not that someone could not hold this view, but that in holding it one would have no account of what is special about one's happiness or rational nature that distinguishes it alone as being of value. Needless to say, that one lacks such an account is not by itself a reductio of one's view, since any account of value has to stop somewhere, with some feature or other that is supposed to account for the value of what has it. Nonetheless, it seems that stopping one's account of value with an irredeemably proprietary property is not at all plausible.

Alternatively, of course, someone might reject altogether the idea that happiness (or anything else) is good *simpliciter* and hold that the important evaluative notion is "good to someone or other." Neither Kant nor Mill ever explicitly considers this option; I suspect Mill, at least, supposed it was not an option precisely because he assumed that for something to be a good to someone requires that it be good *simpliciter*. Interestingly, Moore later does explicitly consider -- if only to reject out of hand -- the suggestion that something might be good to me, or "my own good" without being good *simpliciter* (or, as Moore puts it, "absolutely"). Moore claims that

When...I talk of anything I get as 'my own good,' I must mean either that the thing I get is good, or that my possessing it is good. In both cases it is only the thing or the possession of it which is *mine*, and not *the goodness* of that thing or that possession... The *good* of it can in no possible sense be 'private' or belong to me; any more than a thing can exist privately or *for* one person only. The only reason I can have for aiming at 'my own good,' is that it is *good absolutely* that what I so call should belong to me -- *good absolutely* that I should *have* something, which, if I have it, others cannot have. But if it is *good absolutely* that I should have it, then everyone else has as much reason for aiming at my having it, as I have

myself. If, therefore, it is true of any single man's 'interest' or 'happiness' that it ought to be his sole ultimate end, this can only mean that *that* man's 'interest' or 'happiness' is the *sole good* ... What Egoism holds, therefore, is that *each* man's happiness is the sole good -- that a number of different things are each of them the only good thing there is -- an absolute contradiction.⁵⁰

No such argument can be found in Kant or Mill, and it has its own problems. Nevertheless, it expresses an intuitively attractive view, and I suspect that it captures the sort of considerations that Kant and Mill would have for not taking seriously the thought that, in seeing *x* as good to me, I am doing anything other than seeing *x* as good and hoping to get some of it (or seeing my getting it as good *simpliciter*). So natural is this understanding of people's values, even of the values of those who are obviously selfish, that I think Kant and Mill would understandably (even if, perhaps, not justifiably) treat the alternative as not borne out by the facts of human psychology.⁵¹

In any case, when it comes to appreciating the structure of Mill's argument, the important point is that he takes it that people come to the argument, so to speak, already committed to thinking of happiness *per se* as worth getting. Such a commitment is revealed, he thinks, by people's desires, and specifically by what they desire as ends and not merely as a means to something else. Happiness is at least among the things we desire in this way, according to Mill, and it is also all that we desire in this way. Were he right, then all the evidence, and the only evidence, one could use in determining what is worth while would end up appealing to happiness as the final arbitrator -- and so his argument goes.

Bradley, I think, comes close to seeing the point of Mill's argument when he notes that "[i]f many pigs are fed at one trough, each desires his own food, and somehow as a consequence does seem to desire the food of all; and by parity of reasoning it should follow that each pig, desiring his own pleasure, desires also the pleasure of all."⁵² Mill, of course, never claims that people do desire the general happiness. He is trying to argue from the fact that people desire their own happiness, to their being committed to acknowledging the value

⁵⁰ G. E. Moore, *Principia Ethica*, 98-99. I am not here recommending this as a sound argument, only as one that gives voice, as if they are obvious, to the sort of assumptions I am supposing Mill made without comment.

⁵¹ That is not to say that they would be right. It would be a tricky and interesting challenge to try to settle whether people might ever have the sort of proprietary desire that Moore explicitly rejects as incoherent and Mill implicitly rejects as something we do not have.

⁵² Bradley's *Ethical Studies*, 113.

of the happiness of others. So there is no move in the argument from "people desire their own happiness" to "people 'somehow as a consequence' desire the happiness of all." Rather, from the fact that people desire their own happiness, Mill takes it that they think of their happiness as desirable, that is, as good. A proper understanding of their state, he assumes, is as one of desiring some happiness, because it is good, for themselves. The happiness desired does not, on Mill's view, come prepackaged in proprietary bundles. Instead, it is a kind of thing different people can enjoy in different degrees and hope to get for themselves or others. Thus, the idea is that in desiring happiness for myself, I am thinking of the happiness as good, and (selfishly) wanting it for myself. Should someone else get what I am after, I am committed (by considerations of consistency) to acknowledge that they got something good.

Yet even if Mill is wrong about what people in fact desire as an end -- if, say, they value honesty, or respect, or freedom, as ends -- the strategy underlying the argument remains in place, although it would presumably recommend a different conclusion. For the underlying strategy is to rely on the initial evaluative commitments people find themselves with, and show that considerations of consistency based on the content of these commitments carry further, and perhaps unexpected, commitments in their wake.

The very same strategy is in place in Kant's argument, though of course the initial evaluative commitment he relies on differs significantly from Mill's in both its content and its standing as a necessary rather than contingent commitment. According to Kant, recall, "man *necessarily* thinks of his own existence" as "an end in itself" [emphasis added] and in so thinking, Kant holds, the very same rational ground is in play for each person: that "rational nature exists as an end in itself." That is, we all (supposedly necessarily) take it that rational nature, as opposed to *our own* rational nature only, is an end, and it is this commitment we each have that carries with it a commitment (often not acknowledged in action) that all others who have rational natures are likewise ends, and not merely means.

Why should one believe that each person (necessarily or not) thinks of her own existence as an end in itself? Furthermore, even if each person does, why should one believe that the ground each person has is not, in the respects relevant for the argument, proprietary? Mightn't each person think of her own rational nature as an end in itself without thinking that rational nature *simpliciter* shares that status? Needless to say, some argument -- and presumably a pretty hefty one -- is needed here to underwrite attributing to people the initial commitment upon which Kant relies. Nonetheless, with that commitment in place, Kant's argument, just like Mill's, supposes that a proprietary commitment must be grounded on a nonproprietary one, and moves on to show that the content of the latter sort of commitment inevitably extends to a like commitment

concerning the value (in this case, as an end) of others. If the argument's premises are allowed, what the argument shows is that even someone who acts *as if* others are valuable as means only is committed to holding that, in fact, those others are valuable as ends in themselves -- committed by the very view she relies on in holding that she herself is valuable as an end in herself.

Kant's claim that people *necessarily* think of rational nature as an end in itself puts an interesting twist on the argument. Mill, in contrast, apparently thinks of people's commitment to the value of their own happiness as merely a contingent commitment; if Mill is right, then it seems as if one could back out of the commitment to the value of other people's happiness by giving up the initial commitment to the value of the happiness one secures for oneself.⁵³ Whereas, if the commitment is necessarily had then there is no room to back off of the initial commitment in light of its implications.

Interestingly, while this difference matters substantially to how one might respond in the face of the respective arguments, the necessity in Kant's argument does not in itself augment its strength. What Kant claims to be necessary is not the truth of the commitment -- that rational nature exists as an end in itself -- but rather the thinking of it as true. Were the commitment necessarily true, of course, that would, assuming the validity of the argument that extends the commitment from the first-person case to that of others, make the argument as strong as can be. However, to discover (if it is a discovery) only that people necessarily think of themselves in a certain way appears no more grounds for thinking that the implications of what they think are true than would exist if they thought as they did only contingently.

Nonetheless, if it is true that people necessarily think of themselves in a certain way, then from the point of view of each person, the content of that thought is available, and indeed inevitably available, as grounds for them to reach the conclusion it supports. Hence, while the necessity of the belief does not constitute extra evidence for what the content of the belief entails, the necessity of the believing does ensure the availability of the evidence (so thought of from the first-person point of view) needed to establish the conclusion. Each person considering the argument will, in the relevant sense, have the evidence needed for the conclusion of the argument -- and necessarily so -- even though the fact that each person necessarily has the belief the content of which is her evidence is not itself additional evidence. More accurately, it is not additional evidence absent some special account in light of which the fact that we necessarily think as we do is, in the case at hand, grounds for believing

⁵³ Just how contingent the commitment is on Mill's view is a little unclear. He does, after all, claim that "to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility" (Mill, *Utilitarianism*, IV, 10). However, his grounds for this strong claim are less than clear, to say the least.

we are right. There are suggestions of such a special account in Kant's work, but pursuing them would take us away from considering the central form of argument that Kant and Mill share.⁵⁴

This central form of argument starts with an evaluative commitment, to the effect either that happiness is good or that rational nature exists as an end in itself. It then extends this commitment from its initial role -- underwriting the value we each place either on our own happiness or on our own rational nature -- to a new role: underwriting as valuable the happiness of others or the standing of others as ends. The extension, as I have been emphasizing, turns on the initial commitment not being essentially proprietary, on its scope not being in the first instance limited in its application to the person who holds it. Whether Mill was right to assume, and Kant to maintain, that the relevant commitment satisfies this constraint is, of course, highly contentious. If one or the other is right, however, then his argument goes through fallacy-free.

Acknowledging that other people's happiness is valuable, or that other people are ends in themselves, is not yet to accept that one has reason to promote their happiness or to respond to their standing as ends in themselves. As a result, there seems to be room to accept either Mill's argument or Kant's argument and yet still hold that one has no reason oneself to promote the admittedly valuable happiness of others or to respond to the recognized standing of others as ends in themselves.

There are, I think, two sides to Mill's position here. On the one hand, he apparently thinks, or at least is allowing the possibility, that one has reason to do only what is or will provide a good to oneself. Thus, in Utilitarianism's chapter on the sanctions of morality, Mill emphasizes the importance of arranging things so that people find themselves with reason to do as morality demands. On the other hand, he also seems simply to assume that the right thing for a person to do is always whatever is among the best of her options, where the value of the options is seen as not a matter of what is a good to the agent. He therefore seems to share with Sidgwick and Moore the assumption, to use Sidgwick's words, that " ... as a rational being I am bound to aim at good generally -- so far as it is attainable by my efforts -- not merely at a particular part of it."⁵⁵

⁵⁴ In offering the argument I am focusing on, Kant takes the claim concerning people's initial (and supposedly necessary) commitment as a postulate. It is defended in Kant, Grounding, sec. 3, 49-61 (AK. 446-63).

⁵⁵ Sidgwick, The Methods of Ethics, 382. This is just the assumption Broad takes on, arguing that someone might acknowledge the equal value of others' happiness, but hold that each person "has an obligation to produce good experiences and dispositions in himself, and no such direct obligation to produce them in... anyone else." C. D. Broad, "Moore's Ethical Doctrines," in Paul Schilpp, ed., The Philosophy of G. E. Moore (La Salle, IL: Open Court, 1968), 45.

Put this way, Mill seems simply inconsistent. I think the inconsistency disappears, however, once Sidgwick's view that it is a requirement of rationality that one "aim at good generally" is distinguished from Mill's view that the standard of behavior -- though not specifically the standard of rationality -- is found in the requirement that one take (but, by the way, not necessarily aim at) one of the best available options, where the quality of options is determined by their contributions to the overall good.⁵⁶ Thus, I think Mill is assuming that he and his opponents, and indeed everyone, accepts the view that the right thing to do -- what one should do -- is whatever is best. All the contentious issues, he thinks, have to do with the nature of value, with egoists thinking myopically that only their own happiness is valuable, and others arguing that something distinct from happiness is valuable. Mill's main aim is to argue that those who embrace egoism do so (at least implicitly) on grounds that commit them to acknowledging the value of the happiness of others, and that those who embrace the value of something that appears different from happiness are always (at least implicitly) relying on the very same grounds as their standard, and so are likewise committed to acknowledging the value of the happiness of others (and themselves). Still, Mill consciously makes room for holding, and holds himself, that only what contributes to the happiness of a person is a good to that person, and that what a person has reason to pursue may be limited to what might be good to that person. Of course, when a person successfully does what she has reason to do, it is still open to others to criticize her for not acting morally.⁵⁷

VI. Conclusion

My concern from the start has been to figure out how someone as smart as Mill could have offered an argument that looks as bad as his "proof" does. I have argued that his proof looks so bad -- looks, to tell the truth, as if it is composed of a relentless series of fallacies -- because the real argument of the proof is not what it appears to be. In fact, I maintain, the actual argument Mill offers contains not a single fallacy and, if only the initial premises were true, would be compelling.

In broadest terms, and so the most plausible terms, the argument's strategy is to identify a commitment people are all supposed to share (either as a matter of fact, in Mill's case, or necessarily, in Kant's), and then show that considerations of consistency alone bring further, and often contested (or at least

⁵⁶ As I understand Mill's position here, he is defending happiness as being the "sole end of human action, and the promotion of it the test by which to judge of all human conduct," but seeing morality and rationality as each "parts included in the whole," with neither being identical to the fundamental test of conduct and each being different from the other. On his view, rationality is concerned specifically with what is good to the agent in a way that morality is not. Mill, Utilitarianism, IV, 9.

⁵⁷ In at least some cases, the criticism will also involve the claim that she is failing to care about what is of value. In such a case, the criticism will run, she has reason to value what she does not value, and in failing to value it finds herself without reason to do what she should.

not acknowledged), commitments in the wake of the initial one. It thus begins not by defending the truth of the initial commitment -- an undertaking that would lead to an infinite regress -- but instead by defending the claim that the commitment is available as a premise thanks to what people already accept. As a result, the argument is designed to accommodate well the fact that no evaluative conclusion will follow from purely nonevaluative premises.

In slightly less broad terms, and so a little more contentiously, the argument's strategy turns on the initial commitment being to the value or importance of some consideration that is generic rather than proprietary -- to the value or importance of happiness *per se* or of rational nature considered as such, as opposed to the value or importance of only one's own happiness or one's own rational nature. If, in fact, people only had proprietary commitments, then considerations of consistency would never move people beyond what they are already assumed to acknowledge. However, if the commitment is, as arguments of this form suppose, to the significance of some generic feature, then the argument on hand gets the leverage required to establish that the significance of that feature carries implications for the significance of others. Given the right sort of initial commitment, arguments of this form can provide a person with reason to think that other people's happiness is of value (whether or not the person ends up valuing it), or with reason to think that other people are ends in themselves (whether or not the person ends up treating others as ends). The point of such arguments is to provide each person with reason to accept a certain claim as true -- they aim to justify belief.⁵⁸

In still less broad terms, and so even more contentiously, the argument's strategy turns on particular initial commitments to the value of happiness *per se* in Mill's version of the argument or to the standing of rational nature considered in itself as an end in Kant's version of the argument. Of course, even among those who share the view (perhaps because of the sort of argument that Moore offers) that people are committed to the significance of some generic feature or other, differences concerning just what features might play this role make a huge difference to the conclusions that could be established by the sort of argument used by Mill and Kant. My own hunch -- and a source of my half-heartedness in defending Mill's "proof" -- is that Moore is right that in valuing things, people are inevitably committed to the significance of some generic feature or other, but that Mill and Kant are both wrong in thinking that there is one single such commitment. Perhaps some people are committed to thinking of happiness *per se* as valuable in itself, and

other people are committed to thinking of rational nature, considered in itself, as an end. I suspect, however, that plenty of other people have significantly different commitments -- each of which would invite application of the sort of argument Mill and Kant offer, but not in a way that would lead everyone to a shared commitment.

⁵⁸ Such arguments thus leave open, in an interesting way, what the implications of the given defended standards might be when it comes to what people should value. For all such an argument shows, it might be that the standard defended turns around and requires of people that they value something different than what has been defended as valuable. Indeed, it is a familiar fact about some versions of utilitarianism that they might require rejecting utilitarianism on utilitarian grounds.