Response to an Argument for Utilitarianism

Jacob Williams

Ethics

Dr. Manis

28 April 2009

Stephen Darwall describes an argument for utilitarianism (the theory that the goal of morality is to produce the greatest achievable total happiness among all people combined) which initially seems promising. I believe that this argument's premises are plausible, but that it is incorrect to infer utilitarianism from them. I hope to identify where the flaw in the argument lies, and to show the general way in which the argument would need to be modified in order to successfully derive a normative moral theory from it.

To begin, let us see the argument. The most central notion of utilitarianism seems to be hedonism, the idea that only happiness has intrinsic worth. Darwall considers and rejects one interpretation of this idea – "that pleasure is the only thing it makes sense to aim at as an *end*"[1] - but notes that "some version of hedonism may be far more plausible as a theory of what is good *for* a person."[2] The first premise, then, is that only happiness can be good for a person. (Darwall makes some remarks earlier in the chapter regarding how he thinks "happiness" should be defined so as to make this claim most reasonable[3], but such specifics do not impact the discussion in this paper; Mill's definition, "pleasure and the absence of pain,"[4] is sufficient as long as we remember that "pleasure" and "pain" might need to be given somewhat nuanced meanings.) Using this premise, Darwall presents the following argument:

It is broadly agreed . . . that impartiality is at the center of our conception of morality. Moral judgment, by its very nature, aims to be impartial. It is natural to think, therefore, that the moral point of view is one of impartial, equal concern for all persons. When we are concerned for an individual, we naturally want her good; we want things to go well for her. And if a person's good is her happiness, as we are assuming, then in being concerned for her, we must want her to be happy. So if the moral point of view is a perspective of impartial, equal concern for all, then it must entail an equal desire for the happiness of all. From this the utilitarian concludes that the moral point of view involves a desire for the greatest overall happiness, counting everyone's happiness equally. What is good from the moral point of view is the greatest amount of good for

1   Stephen Darwall, *Philosophical Ethics* (Boulder, Co.: Westview Press, 1998), pp. 127. Italics original.
2   Ibid., pp. 127. Italics original.
3   Ibid., pp. 124-125.
4   John Stuart Mill, *Utilitarianism*. Ed. Steven M. Cahn, *Exploring Ethics* (New York: Oxford University Press, 2009), pp. 92-93.

individuals – that is, the greatest overall happiness.[5]

The core of this argument seems to be as follows:

(1) The only good for anyone is her own happiness.

(2) The moral point of view holds impartial and equal concern for everyone.

(3) Concern for someone entails desire for her good.

(4) Due to (1) and (3), concern for someone entails desire for her happiness.

(5) Due to (2) and (4), the moral point of view desires[6] each person's happiness equally.

(6) Due to (5), the moral point of view desires the greatest overall happiness.

If we accept the premises (1, 2, 3), (4) and (5) seem fairly certain. However, the move from (5) to

(6) is problematic; although it may seem reasonable at first glance, I think there are reasons to

doubt its validity.

The root of the problem is that the first premise contains an ambiguity which propagates

down to step (5), and is then suddenly (and invalidly) resolved in the conclusion. The most

straightforward (but obviously not the intended) way to interpret (1) would be to take

"happiness" to be a specific thing that a person either has or does not have (as opposed to

something a person may have to a degree). If it were interpreted this way, then the "desire for [a

person's] happiness" referred to in (4) would be a desire that can be either met or not met – it

cannot be just met to a degree. (5), then, would imply that the moral point of view contains a set

of equal desires for the happiness of individuals, and each of these desires can either be met or

not met (but cannot be just met to a degree). Since presumably any conflict between these desires

would be resolved by making the choice which satisfies the greatest number of the desires, this

reasoning would let us conclude that the moral point of view desires to make as many people

5   Darwall., pp. 127.
6   For convenience I am going to speak of the moral point of view as having desires; the intended meaning is that a
     person adopting the moral point of view would have those desires.

happy as possible; that, however, is of course not what is intended by (6). By "overall happiness" the utilitarian means the sum of the *quantities* of every individual's happiness. If the argument is to yield any conclusion about the moral point of view's desire regarding this overall happiness, the fact that happiness can be had to a greater or lesser degree must be born in mind throughout the whole argument, not suddenly introduced at the end.

Therefore, it would seem expedient to interpret (1) as meaning "the only good for someone is her own *greatest* happiness." Then (4) can be clarified as "concern for someone entails desire for her greatest happiness" and (5) as "the moral point of view equally desires each person's greatest happiness." For this to help us, we will have to treat desires for a person's greatest happiness as desires which can be fulfilled to a greater or lesser extent (otherwise, (5) will lead us to the conclusion that what the moral point of view desires is for as many people as possible to achieve their greatest happiness, and is not at all satisfied by helping a person achieve any lesser level of happiness – implying, for example, that given a choice between a world containing one million people who are ecstatic but just slightly less happy than they could be, and a world containing one person who is as happy as he could be and any number of people who are as unhappy as they could be, we would be obliged to choose the latter). But the argument will still fail to prove or even render probable its conclusion. To see why, let us consider a specific situation.

Suppose you – an upstanding person whose desires are completely in line with those prescribed by the moral point of view – must choose either to (a) cause Tom to have one unit of happiness and Jane to have one unit of happiness or (b) cause Tom to have three units of happiness and Jane to have zero units of happiness. Your choice will have no other consequences. According to (6), you want to cause as much total happiness as you can and thus

will choose (b). We can investigate whether (6) follows from (5) by seeing whether you would make the same choice on the assumption of (5). If (5) is correct and should be interpreted as suggested in the above paragraph, you have two equally strong desires that come into play: a desire for the greatest amount of Tom's happiness and a desire for the greatest amount of Jane's happiness. Clearly, the first desire is fulfilled to a greater extent in the second alternative, while the second desire is fulfilled to a greater extent in the first alternative. That knowledge alone obviously does not tell us which choice you would make.

Perhaps if we could quantify the extent to which each desire is fulfilled by each alternative, we could predict that you would take the alternative in which the sum of those extents is greatest. Since the greatest happiness Tom can have in this situation is three units, and the greatest happiness Jane can have is one unit, (a) represents a one-third fulfillment of your desire for Tom's greatest happiness and a complete fulfillment of your desire for Jane's greatest happiness, while (b) represents a complete fulfillment of the former and no fulfillment at all of the latter. One-and-one-third of your two desires are fulfilled by selecting (a) while only one is fulfilled by selecting (b), so you choose (a), and thus (6) cannot follow from (5) since they prescribe different courses of action in the same situation. One possible problem with this reasoning is that it seems odd that giving Jane one unit of happiness is enough to satisfy in full your desire for her happiness, while with Tom the same amount counts only as a partial fulfillment. These are necessary consequences of construing "greatest happiness" to mean the greatest amount of happiness that can be brought about given the contingencies of the situation; perhaps it would be better to interpret "greatest happiness" as being the greatest amount of happiness metaphysically possible for a person to have. Then, however, the "greatest happiness" is infinite in quantity (since it is presumably metaphysically possible for a person to live forever

and to continue to add to her total quantity of happiness throughout her infinite lifespan), and it is not entirely obvious how to quantify the extent to which an achievement of some finite amount of happiness fulfills a desire for an infinite amount of happiness. Further, however "greatest happiness" is construed, it really is not clear that the process of choosing whichever alternative yields the highest number upon calculating and summing the extents to which our desires are partially fulfilled describes how a conflict of desires would or should actually be resolved; if it does not, though, then unless some other means of resolving the conflict can be determined (I cannot think of one), we are left with the conclusion that (5) does not prescribe any choice at all, meaning again that (5) does not entail (6), since (6) does prescribe a choice.

Although I believe that the argument Darwall has presented fails in its final step, the rest of it may still be useful in constructing a different consequentialist ethic. If we know that the moral point of view desires each person's greatest happiness equally, then we have some reason for accepting the normative claim that we should perform any acts which would increase at least one person's happiness without sacrificing anyone else's (so, for instance, it would be morally obligatory to tell a lie that hurt no one and helped someone, or to euthanize a person whose remaining life would be worse for her than being dead and whose passing would not injure the happiness of anyone else). This, of course, fails to tell us how to act in many of the most morally significant situations we find ourselves in. We need some criterion (or criteria) for choosing what to do in these cases. At least one more premise would need to be added to the argument in order for it to yield a theory which provides such criteria. One type of candidate for such a premise would be assertions that the moral point of view also desires some other thing – "justice," perhaps – but (in addition to seeming rather arbitrary) an assertion like that would likely amplify the problem rather than solve it, since the newly-specified desire would probably in some

situations conflict with the already-postulated desire for the happiness of each individual. Instead what we need is a premise which explicitly specifies how a person adopting the moral point of view should reconcile conflicts among the moral point of view's desires.

I do not know what specific premise would be best to use; I can only illustrate the sort of premise that I am referring to. Consider this one: "A person adopting the moral point of view should reconcile conflicts between its desires for the greatest happiness of each individual by choosing the alternative in which the least-fulfilled desires are fulfilled to a greater extent than they are in any other alternative." Combined with (1) through (5) above, this would let you derive a moral theory which essentially prescribes that you act so as to never allow any person to be less happy than the situation requires that some person be. This theory has the advantage of avoiding at least one of Darwall's objections to utilitarianism: he describes a situation involving four people in which you can either subtract two units of happiness from one person and add two units to every other person, or just add one unit to every person; utilitarianism counter-intuitively implies that it does not matter what you choose.[7] This new theory prescribes the latter instead. The theory does have its own potentially counter-intuitive consequences, though; for instance, if you for some reason had the opportunity to give a million moderately-happy people a billion dollars each, but this would cause a person outside this group to suffer minor injury (say, a stubbed toe), it would be wrong to do so. More importantly, the premise does not have anywhere near the intuitive appeal that our three starting premises have (if it has any at all). However, my goal was not to recommend this as a correct theory but only to demonstrate what sort of additional premise could be used to allow the argument to generate a moral theory. If no plausible additional premise could be found, then the argument would have to be abandoned.

---

7   Darwall, pp. 132.