Utilitarianism and Preference Change

BRIAN BARRY
London School of Economics and Political Science

For the purpose of this article, I shall take it that at any rate one form of utilitarianism (so-called preference utilitarianism) entails the following proposition: one state of affairs is better than another if in one of them there is somebody whose wants are satisfied more and nobody whose wants are satisfied less. Utilitarianism goes beyond this and prescribes for cases where satisfying the wants of one person comes into competition with satisfying the wants of another, telling us to maximize the aggregate amount of want—satisfaction. I wish here to leave on one side this aspect of utilitarianism, and focus on the simple idea that, other things being equal, it is better for a person's wants to be more satisfied than less. This is an idea whose attraction extends beyond the ranks of utilitarians, and it may seem at first sight that it is an idea with which it would be hard to disagree. Nevertheless many philosophers do reject the proposition that want—satisfaction is a good thing.

An obvious objection is that the satisfaction of wants for what is self-destructive can scarcely be valuable. (The problem with wants whose satisfaction is harmful to others is not at issue.) But it may well be that objections of this nature can be met by reformulating the criterion so that only informed and well-considered desires count. I think that this strategy holds out good hopes of success but I shall not pursue the point here. For my object in this article is to assess a more fundamental line of criticism that has come to the fore in recent years and has found favour with a number of distinguished writers. If it is valid, this new objection would have the implication that the whole idea of attaching value to want—satisfaction is flawed, in that it entails a highly unappetizing doctrine about the formation of preferences.

The objection to be considered runs as follows. If we say that want—satisfaction is what matters, this commits us to saying that people must, as far as it lies within their power, modify their preferences continually in the light of what opportunities are available to them to maximize the prospects of satisfying their wants. John Rawls expressed the idea, in an influential article entitled 'Social Unity and Primary Goods', in this way: people who accept the value of want—satisfaction must be 'ready to consider any new convictions and aims, and even to
abandon attachments and loyalties, when doing this promises a life of greater overall satisfaction ... !

Now I should say immediately that Rawls has always seemed to me to have a quite implausible notion of the ease with which people can change their preferences (not to mention their beliefs), and that his whole theory is weakened to the extent that it depends on the assumption that people are responsible for their own preferences. Given the actual difficulty of bringing about a change in one's preferences—as distinct from simply choosing something different while one's preferences remain the same—we would perhaps not observe any dramatic change in the way people behaved if they were to adopt the prescription offered by Rawls. Nevertheless, the suggestion that somebody who embraces want—satisfaction as a value must on pain of irrationality accept the Rawlsian prescription seems to me a very disturbing one for any partisan of want—satisfaction. But does attaching a value to want—satisfaction really entail a rational obligation to modify one's desires so as to make them as easily satisfied as possible? I wish to suggest that it does not.

I have not until now set out formally the argument that I want to criticize. Let me begin, therefore, by doing that. The argument is of beguiling simplicity. It can be put as follows. Let us take as our starting point the postulate that more want—satisfaction is better than less. How can you increase your level of want—satisfaction? There are two possibilities. One is to change your environment so that your existing wants can be satisfied more fully. The other is to change your wants so that they can be more fully satisfied within the existing environment. (Or you could, of course, combine the two strategies.) If we take seriously the postulate that what matters is want—satisfaction, we have no way of saying that one of these strategies is in principle better than the other. The choice between them is purely a matter of efficiency: assuming that changing the environment and changing your preferences are both costly, each should be pursued up to the point at which the prospective benefits outweigh the costs.

What is wrong with this argument? What is wrong with it is, I believe, that the postulate is misstated. The intuitively attractive proposition about want—satisfaction is that it is a good thing for the wants that I actually have to be satisfied. This can be stated, as I have stated it, by saying that want—satisfaction has value. But this formulation ought to be understood as simply another way of expressing the first proposition. It is an error to suppose that the second version is a more abstract principle that implies the first as a special case.

I can illustrate the fallacious move with a familiar example. Suppose

that someone is a supporter of a certain football team: he roots for it, he wants it to win, he is pleased when it does and downcast when it doesn’t, and so on. We might then imagine a philosopher saying: ‘Ah, what you want is that the team you support should win. But, given that that is your aim, it is quite irrational of you to support the collection of losers that you do support. You should switch to supporting some team with good long-run prospects of winning.’

The response would obviously be that this was to misunderstand the situation. What is entailed in being a fan of some team is that you identify with the fortunes of that team, so of course you want it to do well. To say that you want the team you support to win is correct, but only against the background assumption that there is some given team that you support. It would be an obvious mistake to suppose that the statement could be detached from the background conditions under which it makes sense and used to generate a prescription about how to select a team to support.

Typically, a person becomes a supporter of some team on the basis of history. Often people will retain through life a loyalty to the local team from the place in which they grew up. And even if they do change their allegiance later on, this is normally explainable by later history. The whole idea of choosing a team to support tends subtly to undermine the value of being a supporter, and calculatingly choosing a team in order to maximize the chance of supporting a winner surely destroys it. We might just as well say that a good general wants the army he is leading to win and then deduce that, to achieve his aim, a good general should always be prepared to change sides if the one he is on looks like being beaten.

Suppose that someone continually changes the team he supports—perhaps even in the middle of a game—so as to support the winning side. We might say that such a person is not merely a fair-weather fan: he is no fan at all. Similarly, someone whose preferences were so labile that they changed continually so as always to be satisfied might well be said to be not a real person at all. It is rational in a quite straightforward way to want to satisfy the preferences I have. But there is nothing in that to suggest that I would be equally rational to want to become some entirely different kind of person who would be more satisfied with the things that would be available to that person than I am now with the things available to me. (These things may or may not be the same in the two cases. We should leave open the possibility that, if I were to become a different kind of person, different things would become available.)

Suppose I were told that if my life were extinguished some other person, with an entirely different character and different tastes, could be brought into existence, and that this person would live a much more
Utilitarianism and Preference Change

satisfying life than I do. Would I have any self-interested reason for welcoming this prospect? Surely not. Why then should it be supposed that the perfectly reasonable wish to satisfy my actual wants must commit me to wishing to turn myself into some entirely different kind of person whose wants would be more easily satisfied than mine are? Such a thorough transformation would be akin to death—it would be, if you like the terminology, the death of my present self—and if I have no reason for welcoming death, even if my replacement would be better off than I am, I have no more reason for welcoming a thoroughgoing transformation. 2

The point here can be extended to utilitarianism. What gives utilitarianism its attraction is the idea that it is a good thing for people to be happy—whether we construe happiness hedonistically or in terms of want—satisfaction. This attractiveness does not carry over to the idea that it is a good thing for there to be happy people. The proposition that it is a good thing for people to be happy does not entail that there is an obligation to bring into existence people who will be happy. Similarly, the proposition that it is a good thing for wants to be satisfied does not entail that there is an obligation to bring into existence easily satisfied wants.

That concludes my argument. But for completeness let me ask why the mistake that I have identified is so easy to fall into. One possibility is that what we have is a simple error in quantification: ‘There exists an x such that it is good for x to be satisfied’ does not entail ‘It is good that there exists an x such that x is satisfied’. 3 But I am inclined to suspect that the explanation goes deeper. The principle that it is a good thing for wants to satisfied, understood in the way that makes it plausible, presupposes the existence of certain wants. The principle is therefore silent on the question of preference formation. By the same token, preference utilitarianism, construed on the same lines, will fall short of being a complete philosophy of life. What, I suspect, motivates the shift from the form of the principle that takes preferences as given to the form that does not is the implicit assumption that no general principle can be worth considering unless it purports to answer all questions of conduct that might possibly arise. If we have a principle that, in its intuitively plausible form, fails this test, it is taken for

2, In the film ‘O God, You Devil’, an unsuccessful songwriter makes a pact with the Devil (George Burns) to make him a success. What actually happens is that he is put into the body of a very successful rock star whose soul has just been repossessed, and he then takes up this rock star’s career. Meanwhile somebody else (so, at any rate, we are urged to conceptualize the situation) takes over his own life where he left it off, living in his house with his wife and so on. The songwriter complains that he has not got what he bargained for, and in my view he is absolutely right. What he wanted was for everything to be the same except that his songs would be hits. What he got was something entirely different from this.

3 I am indebted to G. A. Cohen for this formulation of the point at issue.
granted that what we should do is restate it in a more abstract and comprehensive form, even if in the process it loses its original attractiveness.

I can see no logical compulsion here. We can, it seems to me, perfectly well hold that want—satisfaction is a good thing, in the sense that it is a good thing for wants that exist to be satisfied, without thereby being committed to the view that it is a good thing to have easily satisfied wants. Thus, to be specific, we may feel that it is right to bring up our children with high standards, while recognizing that this will have the result of making them less easily satisfied than they might otherwise have been. Yet we will surely at the same time attach importance to their success in achieving those standards and thus being satisfied. It is, indeed, hard to comprehend how we could think that the cultivation of certain preferences mattered unless we also thought that it mattered to be able to act in accordance with those preferences.

Where does this leave preference utilitarianism? In my view it leaves it in as good a position as it deserves to be, in that it salvages the more plausible interpretation and jettisons the less. It will not give us a criterion on the basis of which we can determine what preferences it is good to have. But why should it?