Can there be a preference-based utilitarianism?*

John Broome
Department of Moral Philosophy, University of St Andrews


* I am grateful to John Skorupski, the editors of this volume and a referee for useful comments.
This forms the basis of the so-called ‘ex post’ school of welfare economics, and it is a solidly preferencist principle.

A bigger loss to preferencism is that the additive principle will have to come from elsewhere. It cannot itself be derived from preferencism as Harsanyi hoped. If a preferencist is to be utilitarian, then the aggregative principle of utilitarianism will have to come from some other source besides preferencism. This need not be a deep blow to preferencism, for two reasons. First, additivity may be derivable by Harsanyi’s own methods, if they are suitably reinterpreted. My *Weighing Goods* develops this idea. The reinterpretation could preserve important elements of preferencism, such as the Pareto principle for histories. Secondly, preferencism could anyhow live happily with an independently derived additive principle. The additive principle is about aggregating the good of different people, whereas preferencism is most fundamentally about the good of individual people. So the two may be coexist independently.

4. Preferencism as an account of individual good

From now on, therefore, I shall concentrate on preferencism as an account of individual good. It is one of several competing accounts that exist within the body of utilitarian thinking. It says:

**Preferencist biconditional.** One history is better for a person than another if and only if the person prefers the one to the other.

Preferencism also says that the determination in this biconditional goes from right to left. The biconditional could be true in an entirely unpreferencist way. A person’s good could be determined in some way independently of her preferences, and then the person could form her preferences by always preferring histories that are better for her to histories that are worse. In that case, the biconditional would be true, but preferences would be determined by good. If a person’s good is to be determined by her preferences, as preferencism requires, her preferences must themselves be independent of her good.

For one thing, this means we have to be careful about the concept of preference we adopt. One concept is the dispositional one: to prefer *A* to *B* is to be disposed to choose *A* rather than *B* when you have a choice between them. This is consistent with preferencism. But the existence of another concept is revealed by this fact: I prefer to get up early rather than waste time lying in bed on Saturday mornings, but sometimes I fail to do so. Evidently I am sometimes not disposed to get up early, but nevertheless I prefer it. I do not prefer it in the dispositional sense, but in some other sense. In fact, I prefer it in the sense that I think it would be better for me. Thinking better is one concept of preference, but it does not suit a preferencist, because a preferencist needs preference to be independent of good. The preferencist must stick to preference as a disposition.

5. Ideal preferencism

The version of preferencism expressed in the preferencist biconditional is too pure for almost everyone. People’s preferences are often hasty, badly thought-out, ill informed, inconsistent and in various other ways defective. Even hard-line preferencists find it implausible that a person’s good should be determined by such defective preferences. Most preferencists rely on preferences that are idealized in one way or another: well informed, settled in a cool hour, made mutually consistent and so on. This gives us:
One history is better for a person than another if and only if in ideal conditions the person would prefer the one to the other.

The notion of ‘ideal conditions’ then needs to be spelt out. However, this improved claim also seems implausible, even before spelling it out. What a person would prefer in ideal conditions might perhaps be good for her in those conditions. But what would be good for her in those conditions might be different from what is good for her in her actual unideal conditions. If you were in a cool hour, a quiet cup of coffee might be good for you, whereas as things are you need a stiff drink. To fix this problem, we have to imagine the person, in her ideal conditions, forming preferences on behalf of herself in her actual unideal conditions.

We get:

\[ \text{Ideal preferencist biconditional} \quad \text{One history is better for a person than another if and only if the person would in ideal conditions prefer the one to the other on behalf of herself as she is.} \]

Let us stick with this form of the biconditional. By good fortune, it cuts through another difficulty that afflicts the original preferencist biconditional. People often have altruistic preferences: they are disposed to make choices on behalf of someone else rather than themselves. These preferences evidently do not determine what is good for themselves. But now we are picking out only the preferences they have on behalf of themselves, so we are ignoring altruistic preferences.

Once again, the determination has to go from right to left. This requirement is now not so easy to secure. Ideal conditions are likely to include the condition that the person thinks about her preference. But preferencists cannot allow her to think about it in a way that presumes a notion of her good. She must not ask herself which histories would be better for her than which, and determine her preferences on that basis. Instead her thinking must presumably proceed something like this. She must represent the alternative histories to herself as accurately as she can, and then just allow herself to end up preferring one or the other. This is not the most plausible model of thinking, but it is the one the preferencist must rely on.

For brevity, from now on the only preferences I shall mention are those a person would have in ideal conditions on behalf of herself as she is. I shall call these ‘ideal preferences’. Even when I simply say ‘preference’, it is to be understood this way.

6. A quantitative concept of good

The preferencist biconditional is not enough for utilitarian purposes. For each person it determines what is better for her than what; it orders things according to their goodness for her. But a utilitarian needs more than an order; she needs a quantitative concept of good. Otherwise, the additive principle could not be applied; we could not make sense of the total of people’s good. We must have a concept of quantities or degrees of good for a person. To cut a long story short, these degrees must be cocardinal. This means that ratios of differences of good must be determinate both for a single person and between people. (In general, it is not enough for differences of good simply to be ordered as greater or less; their ratios must be determinate.) How can this be achieved on a preferencist basis?

Evidently we must have a concept of degree or strength or intensity of preference that is also measured on a cocardinal scale. That is to say, first, the degree to which a person prefers one history to another must be comparable to the degree to which she prefers a third history to