

Allan Gibbard, *Reconciling Our Aims: In Search of Bases for Ethics*.  
Oxford: Oxford University Press, 2008. viii + 216 pp.

Allan Gibbard's latest book contains three concise Tanner lectures, challenging responses from Michael Bratman, John Broome, and Frances Kamm, and a detailed reply by Gibbard to these critics (as well as a short appendix discussing a "Harsanyi-like result," and an introduction by Barry Stroud). Gibbard begins with a sketch of the planning account of normative judgments developed in detail in *Thinking How to Live* (2003), together with reflections on how a proponent of his account of normative judgments might best approach the difficult problem of determining what the proper role for our moral intuitions should be when it comes to answering substantive normative questions. The second and third lectures are devoted to the development of arguments intended to establish that, to the extent we find the Rawlsian contractualist ideal of all of us reaching reasonable and unforced agreement on fundamental moral principles to be an ideal to which we are committed, we ought also commit ourselves to act utilitarianism.

This is a conclusion that Gibbard traces back to John Harsanyi, but he argues for it in his own distinctive way. The first of the two lectures in normative ethics loosely revolves around Harsanyi's first welfare theorem, while the second lecture loosely revolves around Harsanyi's second welfare theorem. Gibbard's arguments are intended to be true to the spirit of Harsanyi without being true to the letter of his theorems, as he makes clear in his reply to Broome, who disagrees with Gibbard in his interpretation of Harsanyi's second theorem and argues that Gibbard's own technical "Harsanyi-like result" really establishes much too little for Gibbard's purposes.

One of the main reasons Gibbard wishes to distance himself from Harsanyi is that he is moved by certain arguments concerning the difficulty of specifying what is good for individuals, as provided in Tim Scanlon's *What We Owe to Each Other* (1998). In particular, Gibbard is convinced that questions about individual welfare can't be answered simply by reference to the satisfaction of individual preferences, even preferences formed in ideal conditions (47). What is to count as being in my interest is something that is itself open to negotiation, rather than something fixed in advance of agreeing upon a social contract (75). This means we can't *begin* by utilizing Harsanyi's second theorem in the way it is standardly utilized to derive utilitarianism (or, more precisely, to derive the summation of individual benefits claim that is at the heart of utilitarianism) from an attractive Pareto condition (if a prospect X is either better for or equally good for each individual than another prospect Y, and X is better for at least one individual, then X is a better prospect than Y) in combination with somewhat more controversial coherence requirements on individual and collective goals. Rather, we each begin with individual goal-scales (scales that specify expected

values for a range of prospects), then consider the possibility of being able to adopt a common goal-scale built up out of all of our individual goal-scales, and how the adoption of such a common goal-scale would be best for each of us (if we are to avoid prisoner's dilemmas and the like), and finally come to accept the Pareto condition and coherence requirements in the process of reshaping our individual goal-scales in line with our common aims (160–63).

In their responses, Bratman focuses on Gibbard's metaethics, as does Broome at first, before Broome turns to consider Gibbard's interpretation and use of Harsanyi's second theorem, while Kamm focuses mostly on Gibbard's normative ethics, although she additionally challenges his claim that the planning account of normative judgments provides for a distinctive approach to thinking about moral intuitions. Gibbard's reply to his critics is roughly half as long as the original lectures. Since the original lectures move quickly over a large territory, and the criticisms provided by his interlocutors are genuinely challenging, this long reply is very helpful—Gibbard does a masterful job here of bringing together the several different aspects of his thinking in a fresh way that helps to animate his overall story of how we might best reconcile our aims, at the same time as providing strong (if rarely decisive) responses to the most important criticisms with which he has been presented.

Gibbard ends his response, and the book, by considering the “extremely difficult” question of how commitments regarding the nature of ethical thinking are related to substantive ethical commitments (185–87). He contrasts his own view that judgments about what ought to be done in various situations are, at base, plans concerning what I would do if I were in those situations (for example, to ask oneself whether Caesar ought to cross the Rubicon is to ask, roughly speaking, whether one plans to cross the Rubicon in the situation where one is Caesar faced with the option of crossing the Rubicon) with the more robust normative realism that would have it that judgments about what one ought to do are attempts to represent a reality independent of human plans and desires. Gibbard himself endorses a weak normative realism since he takes it that plans constitute beliefs that can be true or false (and some such beliefs are true). He contends that on the alternative, robust view, moral intuitions tend to be treated as sacred “visions of how things stand morally,” rather than simply as candidates for guides to action that can be helpful to consider when planning what to do. His example is the intuition that there is a fundamental moral difference between killing and allowing to die, discussed by Kamm in her comments. He suggests that when we consider the fact that it will make no difference to an individual who dies whether she is dead as the result of a killing or as the result of a letting die, the stubbornness of nonconsequentialists may be due to their commitment to robust realism, while the intuition at issue will be much more likely to melt away for someone committed to Gibbard's own planning account of normative judgments (other more fundamental intuitions about what matters will remain, perhaps because it is impossible to plan without them).

Gibbard is careful to note that he doesn't think these connections between the relevant metaethical and normative views are conceptually necessitated (186). Nonetheless, it seems that at least two objections are appropriate at this point. First, there are bound to be counterexamples to the claim that it is a commitment to robust realism that undergirds nonconsequentialist intuitions: for example, Kantian normative constructivists are not robust realists (here I am thinking of Christine Korsgaard, in particular), yet they are typically non-consequentialists who are likely to want to defend the moral significance of the killing/allowing distinction. Second, and more importantly, there is the overlooked possibility that a person committed to viewing "ought" judgments as planning judgments, who is asking herself what guides to action to use when deciding what to do, may find that there is a principled way of reconciling her aims with the aims of other people that doesn't result in a commitment to the act utilitarian principle, as the act utilitarian principle requires her to give up too many of the guides to action with which she is likely to have begun (I am still assuming she rejects robust realism for Gibbard's weaker realism). For instance: she might think she will be able to be a better and more reliable planner by adopting the rule utilitarian principle, in addition to, rather than instead of, many of her ordinary guides to action (an idea that is in the spirit of Brad Hooker, *Ideal Code, Real World*, 2000).

Rule utilitarianism is not discussed at all as an alternative to act utilitarianism by either the author or the respondents. Kamm briefly discusses motive utilitarianism and how she thinks it will fail to provide the right kind of guidance when it comes to the killing/allowing distinction (139–40), but I don't think her argument here is persuasive, especially when we consider other forms of indirect utilitarianism. Rule utilitarianism is a highly relevant alternative to act utilitarianism for two reasons. First, as I just suggested, it might fit together very nicely with a planning account of normative judgments, since it seems much less revisionary in relation to the guides for planning with which we actually begin. Second, we might think that the rule utilitarian principle is a more suitable candidate than the act utilitarian principle for the role of a principle that nobody could reasonably reject (see Derek Parfit, *On What Matters*, forthcoming).

We can illustrate how the rule utilitarian principle might be superior to the act utilitarian principle from a contractualist perspective by further pressing one of Kamm's objections to Gibbard (of course, Kamm herself is not a rule utilitarian). Kamm contends that act utilitarianism fails to respect the separateness of persons and may often direct us to treat people as "mere things," that is, treat them without due moral respect. In his response to Kamm, Gibbard argues that act utilitarianism can provide an account of what it is to respect people, that antecedent talk of treating a person as a "mere thing" is unhelpful, and that to work out what it is to disrespect someone we need to first work out what we owe other people, from a moral perspective (154–55). Gibbard might be right that talk of treating a person as a "mere thing" is unhelpful, and it would also

be unhelpful to be unreflective about “respect” (as though we all start off being clear about what it means). Nonetheless, if it turns out that by “respect” the act utilitarian means something that is very distant from what we start off meaning by “respect” when we begin our efforts to establish a social contract, then it will become very difficult to see how contractualism and utilitarianism could be compatible with each other. After all, we are not meant to be talking about a Hobbesian contractarianism that takes only self-interest, narrowly construed, into consideration (if we were, an additional issue would be that the problem of the free rider would rear its ugly head here).

Suppose one is faced with a credible monster who notices that ten healthy strangers are in need of urgent assistance and who presents one with a choice: either submit one’s child to a brutal, slow death by torture and he will go out of his way to save the ten strangers or, alternatively, stand by and do nothing, letting the ten strangers die. The act utilitarian principle places an intuitively *unreasonable* and intuitively *disrespectful* demand on agents confronted with monsters such as this one. Rule utilitarianism will fare much better because it will not require the same kinds of sacrifices. It is one thing to suggest that the meaning of words like “reasonable” and “respectful” can be best appreciated in the light of a normative ethical theory, but another to stretch the meaning of such terms so far that we don’t seem to be talking about the same things any more. To be fair, Gibbard objects to relying on intuitions in relation to bizarre situations (156). However, I don’t think this objection sits well with a planning account of normative judgments that stretches out to plans for Caesar and the Rubicon. If we approach the world from behind a veil of ignorance (as Gibbard would like us to at a crucial juncture), then it is not clear why we should not consider possibilities involving encounters with monsters like the one I just mentioned. And if we are already beyond the veil of ignorance, it is not simply so-called “strains of commitment” that might prevent one from following act utilitarianism when actually confronted with a monster like the one above. Either behind or beyond the veil, it seems that I can reasonably reject the act utilitarian principle.

I have been taking the planning account of normative judgments for granted. Bratman and Broome provide some good reasons for doubting that it is a correct account of normative judgments. Here are a few of the reasons discussed. Bratman worries about wild contingency plans, the differences that may remain after we attempt to reconcile our aims, and whether such differences can be termed genuine disagreements on the planning account of normative judgments. Weakness of the will strikes Broome as a phenomenon that is often best described as involving a plan to do something that one recognizes conflicts with what one thinks one ought to do, but this description is unavailable to us if planning to do something and judging one ought to do something amount to the same thing. Broome also draws out the extent to which normative judgments are really better described as “plan-like” than as actual plans on Gibbard’s story, partly because it turns out, when Gibbard is being more precise, that they are

actually “okayings” (such as when one okays A and okays B in a Buridan’s Ass situation, where one must choose between A and B), and partly because they are highly idealized (as the Caesar at the Rubicon example illustrates). Broome then asks: why should we view our actual normative judgments as being based in such idealized merely plan-like states, when this is not what they look like to actual agents? Gibbard’s responses to these objections turn on a commitment to judgment internalism (the point of “ought” judgments is to motivate) and a story about how the idealization required by his account goes hand in hand with rational norms that play an essential role in making desires and intentions what they are. He doesn’t have much to say in this book about stubborn, lingering disagreements, but he has written on this topic elsewhere (see the last chapter of *Thinking How to Live*).

Despite all of these criticisms, including the ones I have just added to the pile, I believe that fair-minded readers will be impressed by the multifaceted and sophisticated nature of Gibbard’s overarching vision of the nature and content of our ethical judgments. They are likely to have many remaining questions and nagging concerns (I plan to when in their shoes), but they will hardly wish to deny that their thinking has been richly stimulated by reading this book.

*Daniel Star*

Boston University

*Philosophical Review*, Vol. 119, No. 2, 2010

DOI 10.1215/00318108-2009-042