Philosophy and Phenomenological Research 84 (2012): 233-43; penultimate draft; please use published version for citation.

Goodness and Justice[*]

Ben Bradley
Syracuse University

In *Goodness and Justice*, Joseph Mendola defends three related views in normative ethics: a novel form of consequentialism, a Bentham-style hedonism about "basic" value, and a maximin principle about the value of a world. In defending these views he draws on his views in metaethics, action theory, and the philosophy of mind. It is an ambitious and wide-ranging book. I begin with a quick explanation of Mendola's views, and then raise some problems.

I. MAC, Hedonism, and HMP

The book is divided into four parts. In Part One, Mendola explains and defends his version of consequentialism, called "Multiple Act Consequentialism" (MAC). MAC is motivated by two well-known problems for act consequentialism that Mendola creatively ties together. The first problem is Castaneda's paradox of act utilitarianism.[1] Here is the paradox: sometimes a complex act may have the best consequences, even though its parts do not; (A&B) might have better consequences than any alternative to (A&B), while A fails to have better consequences than A's alternatives, and the same for B. This paradox was widely discussed in the 1970s but has received little attention since. The second problem is the free-rider problem. If I can produce some value for myself (or someone else) only by defecting from an important group action, and my participation in that group action wouldn't otherwise matter to the outcome, then act consequentialism allows me to defect. Thus act consequentialism allows for, and even requires, free-riding; many have found this unacceptable.

To tie these two problems together, Mendola argues that complex acts performed by a single individual over time are relevantly like group acts – strictly speaking, they *are* group acts. Here Mendola introduces the notion of an atomic agent, which is a "momentary period" of an agent (p. 33). When an individual performs a complex, temporally extended act, we may think of this complex act as being performed by a number of atomic agents cooperating with one another. Mendola's view here is inspired by Parfit's well-known views about personal identity, in particular the view that the relation between myself now and in the future is not importantly different from the relation between myself now and someone else.[2]

Mendola's strategy is to take what seems to be a natural view about complex individual acts and apply it to all group acts. A natural view to take about the Castaneda paradox is that the consequences of a complex act are (at least sometimes) more important than the consequences of its parts; the complex act seems like a more suitable object of evaluation than its components. For example, if breaking the patient's rib is part of a beneficial surgical procedure, then it ought to be done even if *it* does not bring

---

[*] Joseph Mendola, *Goodness and Justice: A Consequentialist Moral Theory* (New York: Cambridge University Press, 2006). All page references in the text are to *Goodness and Justice*.

[1] Hector-Neri Castaneda, "A Problem for Utilitarianism," *Analysis* 33 (1973), pp. 141-142. Also see Lars Bergstrom, "Utilitarianism and Alternative Actions," *Nous* 5 (1971), pp. 237-252.

[2] Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), Part Three.

about anything good. On this point Mendola hopes all will agree. But if what we think of as a complex individual act is in fact a sort of group act, then the consequences of a group act are sometimes more important than the consequences of its parts when determining what a member of the group should do. Hence, at least sometimes, an atomic agent ought to cooperate with a group even though that atomic agent could bring about better consequences by non-cooperation. And this is so whether the group is composed of momentary periods of a single temporally extended agent, or of multiple distinct temporally extended agents. Hence the connection with the free-rider problem.

But under what circumstances is an atomic agent required not to defect from a group act? Mendola says "one should defect from a group act with good consequences only if one can achieve better consequences by the defecting act alone than the entire group act achieves" (p. 46). I will return to this view in Section II.

In Part Two, Mendola defends an admittedly crude Benthamite form of hedonism, which he calls the view that "pleasure is the basic ethical or normative value, and pain the basic disvalue" (p. 105). He says that in calling pleasure the *basic* value, he does not mean to say that it is *intrinsically* good; nor does he mean to say that it is good *for a person* (pp. 108-109). It is not clear to me what basic value is supposed to be, but perhaps it isn't important, since Mendola says that "the best way to engage the hedonism I will defend… is by regard to current discussions of well-being and the good" (p. 109) – that is, by treating hedonism as it is normally construed, as either the view that pleasure is the sole intrinsic good, or the view that pleasure is the only thing that is intrinsically good for a person. One notable feature of Mendola's hedonism is that according to Mendola, pleasure is a feeling (rather than an attitude, as Fred Feldman believes); pleasures are "phenomenally homogeneous" (p. 107).[3] Mendola ably defends hedonism against various purported counterexamples, especially Nozick's famous "experience machine" case.[4] Among other things, Mendola argues that there would not be much difference between finding out you were on a certain sort of experience machine and finding out that Berkeley's metaphysics is correct (p. 119).

Mendola then gives a metaethical argument for hedonism. He very briefly argues against various competing metaethical views, and claims that the correct view is "non-constitutive naturalist cognitivism" (p. 150). According to this view, moral judgments are true or false (cognitivism); certain natural properties are intrinsically normative (naturalism); those properties are not constituted by any other properties (non-constitutivism). The natural, intrinsically normative, non-constitutive properties are pleasure and pain. In order to fully characterize what happens when someone is stabbed to death, we must describe the visceral sensations of the victim; and these sensations cannot be fully described without ascribing badness to them. To know what it feels like to be stabbed is to know that it is intrinsically bad (p. 157). There is "objective intrinsic disvalue" here (p. 157). Furthermore, the pain sensations are unconstituted, because they are phenomenal properties, or qualia (p. 158). Pleasures, on the other hand, have objective intrinsic positive value. The values of pleasures and pains are not cardinal, but ordinal; some sensations are better than others, but not by any amount (p. 165).

[3] For an opposing view, see Feldman, "Two Questions about Pleasure," in Feldman, *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy* (New York: Cambridge University Press, 1997).
[4] Robert Nozick, *Anarchy, State and Utopia* (Basic Books, 1974), pp. 42-45.

In Part Three, Mendola argues that the value of a world is determined by the Hedonic Maximin Principle (HMP). Here Mendola introduces the notion of a "fleck," or momentary bit of experience. According to HMP, one world is better than another if its worst fleck is better; if their worst flecks are equally bad, one is better than the other if its second-worst fleck is better; and so on (p. 213). Mendola gives a very complicated argument for HMP, based on nine principles governing permissible orderings of worlds (pp. 191-222). The argument depends crucially on the claim that hedonic value is ordinal. I will return to the ordinality claim in Section III.

Mendola then attempts to show that HMP is an intuitively appealing principle. As part of this attempt, he argues that temporal segments of lives "should receive the same sort of distributional concern that lives receive" (p. 227), and thus, just as we typically don't think it is justifiable to sacrifice one person's happiness for anothers, we shouldn't necessarily think it is justifiable to sacrifice one person's present happiness for that person's future happiness. Mendola attempts to defend HMP against some obvious apparent counterexamples. For example, HMP entails that a world with one painful broken finger and lots and lots of happiness is worse than a world with no pain or pleasure – even if the person whose finger is broken gets lots of happiness too (p. 264). Mendola gives three replies to this objection, which I will address in Section III. In Part Four, Mendola combines HMP with MAC to show what his comprehensive theory requires of moral agents, and argues that the results square significantly with common sense morality.

Thus Mendola has presented us with a comprehensive picture of the ethical terrain according to which there are objective facts about goodness and rightness, defecting from a good group act is rarely permissible, and what fundamentally matters are pleasure and pain and their distribution among atomic agents. It is in some ways an attractive picture and its parts hang together well. I now turn to filling in some details and offering some criticisms.

II. Against MAC

Recall that Mendola's criterion of rightness, MAC, is the view that "one should defect from a group act with good consequences only if one can achieve better consequences by the defecting act alone than the entire group act achieves" (p. 46). The idea here is that atomic agents often overlap with larger agents pursuing group acts, and the consequences of both are relevant to what ought to be done. We are supposed to compare two things: the consequences of the group act, and the consequences of the defecting atomic act. More specifically, Mendola says we compare two situations: (S1) a situation where the defecting atomic act is performed, but the other members of the group agent do not constitute such an agent, and (S2) a situation where the individual cooperates and the group agent performs its action (p. 46). If the value of S1 is greater than that of S2, then defecting is permitted; otherwise it is not permitted. So in a typical free-rider case, defection will not be permitted.

To illustrate, suppose Alfred is a member of a group committed to saving lives. Alfred happens to be relaxing in a comfortable chair eating an ice cream sundae when he notices that the other members of the group are jointly saving someone's life. His help is not necessary to save the life, and he is getting some mild pleasure from the ice cream.

Nevertheless, MAC requires Alfred to help the rest of the group. The situation where the group does not save the life, but Alfred eats the ice cream, is worse than the situation where he foregoes the ice cream and helps the group save the life. There is also a second component to MAC, dealing with cases where two group acts overlap. If group act G1 has better consequences than overlapping group act G2, then one ought to defect from G2. Again, this is determined by comparing a situation where G1 is performed and the group agent of G2 does not exist with the reverse situation; if the situation would be better if G1 were performed and the G2 agent doesn't exist than if G2 were performed and G1 doesn't exist, then one should defect from G2. (Alfred's case is really of this sort, since eating ice cream counts as a group act according to Mendola's view, but to avoid confusion I'll continue to refer to Alfred's act as an individual act.)

I have some problems with MAC. First, there is a worry about motivation. Why should we compare (in the case of an individual act overlapping with a group act) a situation where the individual cooperates to a situation where she defects *and the group agent does not exist*? Mendola is led to this comparison in the following way. Take a case like Alfred's, where an individual act and a group act overlap. Direct consequentialist principles require Alfred to act in one way, and require the group of which he is a part to act in a different way. This conflict must be resolved. It cannot be resolved by ignoring the group act – it is a genuine, legitimate action. The same goes for the individual act. What to do? "Neither defecting nor cooperating has a more direct consequentialist rationale, because neither the group act nor the defecting individual act has a more direct rationale. So we have a real conflict here" (p. 44). Thus, we are led to the idea of comparing the consequences of the individual act and the group act.

But this story is problematic. The question here concerns what *Alfred* should do. So long as we are focused on this question, the choices faced by groups of which he is a part seem irrelevant. At least, there is no direct consequentialist rationale for considering the effects of group actions performed by groups of which Alfred is a part, *when we are considering what Alfred should do*. (There would be a rationale for considering the effects of "group acts" that consist only of atomic Alfred-stages doing things, but not for acts that involve agents distinct from Alfred.) This is not to say that group acts cannot be right or wrong based on their consequences, or that group acts are in any way less real. But so long as it is Alfred's individual act that we are evaluating, the consequences of the group act are not *directly* relevant. Contrary to Mendola's assertions, MAC is not a version of direct consequentialism. This is not (yet) to say it is false, but only to say that it faces whatever motivational problems beset traditional indirect consequentialisms.

MAC also faces some problematic cases. I'll start with one that Mendola considers.[5]

> Consider this particular case: Two disasters threaten humanity, Big Disaster, which is that we will all be boiled in oil for the rest of our lives, and Small Disaster, a small but nasty war. If a certain group agent in which I participate takes action, Big Disaster will be averted, though in the normal course of things Small Disaster will occur. But if I secretly defect from my role in that group, I can see to it that Small Disaster is avoided, through an individual project that takes time and hence is a smallish group act, and the big group

[5] Mendola attributes the objection to Torbjörn Tännsjö.

> will go ahead and achieve its goal without me. At least in some cases of surplus cooperation, it seems we should defect from the weightiest group act. (p. 50)

This seems to be a convincing objection to MAC. Mendola has a reply. He points out that if the members of the group action preventing the Big Disaster are like us, they would not blame someone for defecting in order to prevent a small war. Thus, preventing the small war, even though it involves not helping the group preventing the Big Disaster, would not be defecting from that group act after all. In order for an act to count as defecting, the members of the group would have to be willing to complain.

It is probably true that most people would not blame the defector in these circumstances. But why couldn't there be a group of people very much like us, but who are very strict deontologists and regard cooperation as of paramount importance, even when it involves failing to prevent a small catastrophe? And why should it matter to me whether the members of my group have that attitude when I'm deciding whether or not to prevent the small war?

Here is a different sort of example. Suppose you can achieve a great good by defecting from some group act with good (but not great) consequences. But you can achieve that great good only if the group actually does its good group act. MAC directs us to compare a situation where you achieve the great good *and the group act doesn't exist* to a situation where you cooperate with the group and it does its good action. But by hypothesis, there is no such situation. MAC doesn't give us any result at all in this case; it just blows a fuse.

Here is another problematic case. Suppose Beth is an influential member of a group of teenagers. The teenagers are engaged in a group action that has slightly good consequences. However, the action is very fragile; the teenagers tend to get into trouble, but Beth's good influence helps keep them bringing about the small good. Like most teenagers, they are harmless when alone; it is only when acting together that they tend to be destructive (though Beth's influence helps keep the group on the right track). Beth faces the following situation. She can do a kindness for someone, bringing about a medium amount of intrinsic goodness (greater than the amount of goodness the group would produce); but this requires her to defect from the group, which would result in the teenagers wreaking havoc. On the other hand, she could refrain from the small kindness, remain in the group, and help the group bring about the small good. In this situation, it seems clear that Beth should *not* defect from the good group action. But MAC allows her to defect, because the situation where she brings about the benefit of defecting (the medium good), *and the teenagers do not act in concert* (and thus bring about nothing bad), is a better situation than the situation in which Beth and the teenagers act in concert to bring about the small good. The source of the problem is that MAC requires us to make the wrong comparison. In order to get the intuitively correct result, we need to somehow take into consideration what the group act would achieve *without Beth in it*.

This case demonstrates a structural flaw in MAC: it cannot always account for acts that are obligatory in virtue of *preventing* great evils. According to a standard version of act utilitarianism, an act is right iff it maximizes intrinsic value. The rightness of an act is determined by a comparison between the intrinsic values of its consequences and the consequences of its alternatives. If some act is the only available alternative that

will prevent some great evil from occurring (and all else is equal), it is obligatory – not because its own consequences are intrinsically good, but because the consequences of the *alternatives* are intrinsically *bad*. But MAC cannot account for the importance of prevention in a similar way, as the example of Beth demonstrates. When we compare the intrinsic value of a situation where an agent cooperates with a group action with the intrinsic value of a situation where the group action does not take place, we fail to account for the preventive value of the cooperating atomic act; and this preventive value sometimes explains why the cooperating act is required. We might try to account for preventive value by looking at the total impact of an atomic or group act on the value of the world. There are serious difficulties involved with this move, as Alastair Norcross has shown.[6] I will not rehearse these difficulties here, since this is not Mendola's view.

III. Against HMP

Recall that HMP is the view that world W1 is better than world W2 if (i) W1's worst fleck is better than W2's worst fleck, or (ii) the value of W1's worst fleck = the value of W2's worst fleck, and W1's second-worst fleck is better than W2's second worst fleck, and so on. Mendola's argument for HMP is too complicated to capture in this essay, but it depends crucially on *ordinalism*: the view that while some experiences are better or worse than others, they are not better or worse by any amount. Mendola gives no argument for that claim; he merely says that "the relatively simple ordinalism of value levels that I have so far presumed is apparently consistent with my own introspection" (p. 167). What he finds consistent with his introspection are claims such as that no experience of pleasure is precisely twice as pleasant as another; he says "it is massively implausible that the property of being exactly twice as good as something else could present itself in experience" (p. 166). I am not sure exactly what Mendola means by this. Perhaps he means that it never seems to anyone as if one pleasure is exactly twice as pleasant as another. He might be right about this, but it hardly shows that no pleasure is twice as pleasant as another, for it seems possible that we are not perfect judges of exactly how pleasant our experiences are.

Consider the following example. As I write these words I am imagining some lines. First I am imagining two lines A and B that are very similar in length, but A is slightly longer than B. Now I am imagining a very short line C next to a very long one D. In my imagining, D does not present itself to me as being exactly ten times longer than the shorter one. A does not present itself as being exactly 1.3 times longer than B. In fact, no matter how long I contemplate these imaginary lines, I find myself unable to discern the precise length relations between A and B, and between C and D. But this cannot possibly establish ordinalism about imaginary lines.[7]

There are difficult and interesting questions about the measurement of pleasure and value, and I do not intend to suggest that it is obvious that a pleasure may be exactly

[6] Norcross, "Good and Bad Actions," *The Philosophical Review* 106 (1997), pp. 1-34.

[7] Mendola says that "phenomenal spatial and temporal intervals may well present themselves with such a structure" – i.e., one spatial interval may appear to be exactly twice as long as another – but that "we aren't considering at the moment spatial and temporal features" (p. 167). Whatever we make of the drawing of this distinction, I find it implausible that one imaginary line must present itself as being exactly 1.3 times the length of another – at least, it is no more plausible than a pleasure presenting itself as 1.3 times more pleasant than another.

twice as pleasant or twice as good as another. But whatever we say about these difficult questions, it is hard to believe that Mendola's introspection does not tell him that some pleasures are *much more* pleasant than others, or that some pleasures are *only a bit* more pleasant than others, just as A seems only a bit longer than B and D seems much longer than C. Surely pleasures *do* present themselves in experience in these ways. And if they do, ordinalism is counterintuitive. In any case, Mendola does not provide much of an argument for ordinalism.

Since ordinalism apparently leads to HMP, if HMP is implausible, we have good reasons to think ordinalism must be false, whether it is intuitive or not.[8] And HMP does have counterintuitive implications. As mentioned above, HMP entails that a world that is full of happiness save for one experience of pain – a crushed finger that hurts for an hour – is worse than a world with no pain or pleasure, even if the person whose finger is hurt gets lots of happiness too (p. 264). Mendola recognizes that this result is counterintuitive. He has three lines of defense.

First, he claims that "our intuition is misled by the lack of responsibility presumed in this case… So instead assume that you must torture an innocent by breaking his finger against his will in order to bring the other things into their happy existence… That… is not confidently commanded by commonsense moral intuition" (p. 264). This response is very unconvincing. First, why should our intuition be misled in this case? Why should we be unable to judge the values of worlds without assuming any responsibility for a choice between them? And even if our intuitions about values of worlds in non-choice situations are suspect, why think that changing the example in the way Mendola suggests results in a *more accurate* judgment about the values of worlds? Why is it better to think of examples where pain is brought about as a *means* to pleasure, rather than as a side-effect? Surely everyone would judge it better to bring about the great happiness with the short pain as a side effect. Finally, even those who find Mendola's response convincing here would presumably not find it convincing in other, relevantly similar cases. For example, consider two worlds that contain exactly the same amount and distribution of pain, but different amounts and distributions of pleasure. W1 contains a million and one mild pleasures distributed (more or less) evenly among a thousand people. W2 contains a million very intense pleasures distributed (more or less) evenly among a thousand people, plus one very tiny (very tiny < mild) fleck of pleasure. W2 contains a great deal more pleasure, but W1 counts as the better world according to HMP.[9] This is absurd. And I take it that it becomes no more palatable if we imagine that we have a choice between allowing a thousand people to have small amounts of pleasure, or bringing about much greater pleasure for all by causing a very tiny pleasure for one.

Mendola's second reply to the argument really consists of two separate points. First, he claims that the example is unrealistic. Nobody would ever face a choice about whether to bring about a world with lots of pleasure and a one-hour pain or a world with no sentience. One way in which the example is unrealistic is that it presupposes *certainty* about what will happen in each outcome, which we never actually have. And without that certainty, common sense does not provide a clear answer about what is to be done.

[8] In fairness, Mendola admits that he is not certain that ordinalism is true (p. 170). But if it is false, his primary argument for HMP is undermined.

[9] The language of "mild" and "very intense" may itself be a problem for Mendola given his ordinalism, but this could be avoided by using more specific descriptions of the pleasures.

But this reply is weak. The supposition of certainty does not seem to me to undermine the force of the intuitive judgment at all. While we rarely if ever have certainty about what the outcome of a choice will be, it is easy enough to *imagine* having such certainty. Furthermore, certainty is not crucial to the case at all. Suppose you face a choice between two alternatives acts, and you lack certainty about what will happen given each act. If you do A1, it's 99% likely that there will be lots and lots of pleasure for everyone, and a one-hour pain in one person's finger. If you do A2, it's 99% likely there will be no more sentient life at all. Is it so hard to figure out what to do? (It depends, of course, on what happens in the other 1% of cases. But that 1% would have to be very bad in A1, and very good in A2, for it to make sense to choose A2.) Certainty is a red herring here.

The second component of Mendola's second reply is this: "When there is sudden physical and painful danger to someone, it is intuitively absurd to pause and worry about how their lives are going as a whole. The intuitively pressing nature of pain, which we also considered in the last part, undergirds the most salient interaction of the three structural elements of HMP" (p. 266). I cannot see how this is relevant to the argument. Does the fact that pain strikes us as intuitively pressing support the claim that it's better to have no sentience at all than to have one hurt finger along with lots and lots of pleasure? Furthermore, this reply does nothing to counter the revised version of the argument that involves only pleasant experiences. Very mild pleasures do not present themselves as pressing problems.

Mendola's third reply is most interesting. He claims that, despite what is commonly assumed, it would actually be better if there were no sentience. At least, any realistic world, where people are getting some pain, would be worse than no sentience at all. "Our real world is in fact worse than nothing, even according to much less pessimistic principles than HMP" (p. 269). While I doubt the real world is worse than no world, I won't argue with that pessimistic claim here. I will merely point out that HMP has implications for worlds other than the real world. It entails that a world with a one-hour pain in someone's finger, plus lots and lots of pleasure for everyone, is worse than no world. A world with just one finger-pain is not anything like the real world. It is fantastically better. So even if Mendola is right about the real world, or if he is right that common sense supports the claim that the real world is worse than no world, this is irrelevant to the argument at hand.

I've argued that MAC and HMP have unacceptable implications. In closing I note two things. First, there is a great deal of interesting material that I haven't mentioned. It is am impressive and complicated book that requires careful study. Though I've targeted what I take to be some of the weaker points of the book, some of his arguments are more convincing. In particular, I think Mendola shows that hedonism has more going for it than is often supposed. Second, I should point out that Mendola's commitment to HMP is somewhat halfhearted. On pp. 224-5, he suggests that the sort of extreme ordinalism he tentatively endorses may be false, and that value may have a structure that is somehow between ordinalism and cardinalism (what that might be, he does not know). And if this is true, he says, the truth about the values of worlds may lie between HMP and "total utilitarianism" (according to which the value of a world is determined by summing the values of the pleasures and pains in it). Mendola seems more committed to hedonism than to HMP. So Mendola might not be all that unhappy to

discard HMP in favor of some other axiological principle, so long as it is compatible with hedonism.[10]

[10] Thanks to Elinor Mason, Eric Moore, and Ernesto V. Garcia for helpful discussion.