

**Ambivalent Desires and Problem with Reduction [Pre-Print]**

Derek Baker

NOTE: This is the penultimate draft of a paper appearing in *Philosophical Studies*. Please cite that version of the paper.

**Abstract** Ambivalence is most naturally characterized as a case of conflicting desires. In most cases, an agent's intrinsic desires conflict contingently: there is some possible world in which both desires would be satisfied. This paper argues, though, that there are cases in which intrinsic desires necessarily conflict—i.e., the desires are not jointly satisfiable in any possible world. Desiring a challenge for its own sake is a paradigm case of such a desire. Ambivalence of this sort in an agent's desires creates special problems for the project of reducing all facts about an agent's desires to facts about his or her preferences over options. If this reductive project fails, there is reason to suspect that the Decision Theory cannot give us a complete theory of Humean rationality.

Desires conflict. The satisfaction of one almost inevitably demands giving up on the satisfaction of another. Occasionally this conflict becomes acute, and we experience ambivalence, mixed feelings, about an option. One is aware of having two desires that aren't jointly satisfiable.

Most of this conflict depends on some contingency. I desire healthy teeth, and I desire to avoid pain. These two desires conflict, but that's thanks to the current state of dentistry. We could imagine the healthy teeth and the pain coming apart, providing an option I could desire wholeheartedly. My two desires conflict accidentally, or *indirectly*.

But some desires conflict inevitably. Humans like challenges. We become bored and anxious if we constantly avoid them. But activities count as challenges in virtue of obstacles, the risk of failure, or hardships patiently endured. Hence we're often afraid of challenges, and avoid them, and put them off, even while we want them. Here there

aren't two separable objects. There's no possible world where I satisfy my desire for a challenge and my desire for ease.

Here we have desires that conflict necessarily, or *directly*. An agent desires  $p$  and desires  $\sim p$ . (Or he desires  $p$  and desires  $q$ , knowing that  $p$  entails  $\sim q$ .)

If directly conflicting desires are possible, then desires cannot be reduced to the preferences of Decision Theory. For the most explicit endorsement of the reduction of desires to preferences (hereafter called 'reductionism'), see Dreier. Reductionism is also in the background of Jackson, Broome, and Edgington. A fundamentally ambivalent agent will have a desire and an aversion aimed at the same state of affairs. So one attitude will constantly mask the other in the agent's preferences.

Reducing desires to preferences is the first step in providing an attractive theory of the Humean aspect of rationality—i.e., that part of rationality concerned with consistency among practical attitudes. The reductionist argues that desires are consistent when they supervene on Decision-Theoretically consistent preferences. But supervenience fails, thanks to direct conflict, leaving it mysterious what consistency for desires (Humean rationality) might be.

Before beginning the argument, let me clarify terms.

## 1. Desire and Ambivalence

In the philosophical literature 'desire' is a semi-technical term, similar to but not identical with 'desire' in everyday speech. They are what Donald Davidson referred to as 'pro-attitudes,' (1963/1980) or what Gilbert Harman has called 'a want.' (1993/2000) A

desire is an attitude aimed at a state of affairs, where this state of affairs is a goal or end of the agent. An agent desires that *p*.

Since I take the philosopher's notion of 'desire' to be semi-technical, I will make minimal appeal to arguments like, "we would say that this agent desires X." Instead, I will argue that we should attribute desires when doing so is necessary in order to explain an agent's actions, deliberations, or events like the exercise of will-power.<sup>1</sup>

Preferences, on the other hand, take two alternative states of affairs as their content, where the agent would rather the first state obtain than the second. An agent prefers *p* to *q*. In other words, he would take *p* over *q* when those are the options he's presented with.

The only difference between desires and preferences is the sort of content each can take.<sup>2</sup> Desires take a single proposition as a representation of a goal-state, while preferences take two propositions, ranking one state of affairs over another. Reductionists are therefore committed to the view that all of an agent's attitudes representing some state of affairs as a goal reduce to attitudes which rank one state of affairs over another.

## 2. The Examples and Their Interpretation

---

<sup>1</sup> This is not to say that how we use terms like 'desire' and 'preference' in everyday speech is uninteresting, or that it might not have consequences for philosophy of mind or moral psychology. I suspect though, that our use of these terms is too loose and fluid to support any interesting theses of reduction.

<sup>2</sup> In conversation I've found that people frequently take 'desire' to mean something like the folk notion of 'desire', something that has both affective and behavioral components, while they take 'preference' to mean something like 'revealed preference', something purely behavioral. But one can give a behavioral-functional theory of desires (for example, Smith), or give a non-behaviorist theory to preferences (see Dreier). The distinction between desires and preferences is a distinction in the type of content each attitude receives, not a distinction between purely functional attitudes and attitudes with a phenomenology.

In the cases considered below, we can know that the agent wants and is averse to the same state of affairs rather than two contingently connected states, because the desired state of affairs is appealing thanks to the presence of something the agent also finds aversive. The aversive state is partially constitutive of the state desired. So if the state of affairs lost its aversiveness, it would also lose some of its attraction to the agent. Consider cases like:

**The Unpleasant in Art:** We frequently desire to see the tragic, frightening, shocking, melancholy and disturbing in art. Tragic scenes in movies derive much of their power because we want the characters to survive, to succeed, to be happy; and yet at the same time we don't want these things, because the movie would no longer have its tragic aspect if they did. Or think about someone describing an especially harrowing scene from a book, saying that they could barely read it. This can be high praise in the right context.

But the conflict is clearest in horror movies. Imagine desperately wanting to look away from the screen, because the movie is so frightening, but forcing yourself to continue watching, because the movie is frightening. You must desire to look away, otherwise why the need to exercise will power? You must desire to keep watching, otherwise why bother? And if the movie lost the terrifying aspect that you have to exercise will power to endure, it'd be a less interesting movie.

**Challenges:** We often desire challenges for their own sake. But the desire's odd, because an activity qualifies as a challenge in virtue of obstacles, the risk of failure, or

difficulties patiently endured. Agents who have an intrinsic desire for a challenge (as opposed to an instrumental desire for the benefits of completing a challenge) desire some activity precisely because of properties they find aversive.

Consider the following: An amateur runner is contemplating entering a marathon. The idea of a challenge is attractive. The agent desires to run in a marathon because it is hard. At the same time, the agent is aware that he might fail to complete the marathon, and that a great deal of effort will be required. The fact that the marathon is hard makes it frightening.

Again, we can imagine this agent going through the characteristic experiences of internal conflict. He may engage in prolonged deliberation about whether he really wants to take part in the race, may find himself forced to exercise will power in order to stick by his decision to enter—and will certainly require will power to stay in the race, let alone exert himself to a degree that would make his time in the race an achievement. If the marathon lacked risk, and his success were guaranteed, the agent could easily lose motivation to participate; or, if the marathon lacked the painful and tiring components of physical exertion, the race could cease to be of interest to the runner. The risk and difficulty that make the marathon frightening, which the runner must exercise will power to face, are the same elements that make the marathon exciting.

Of course, not all challenges have the same aversive properties. A challenging philosophical problem is not painful in anything like the way that a marathon is painful. Nonetheless, any particular challenge will have some cluster of aversive properties that partially constitute its challenging nature. These are features like pain or frustration that must be endured, the risk of failure, obstacles that must be overcome or evaded, and so

on. So, when someone desires a particular activity thanks in part to the fact that it is challenging, then the desire must depend in part on the presence of some combination of aversive features. These aversive features are necessary for the activity to qualify as a challenge.

In any case, examples of wanting to experience the unpleasant through art and wanting challenges both look like cases of directly conflicting desires, and they both look rational.

#### Objections to My Interpretation of the Cases

One who wishes to insist that direct conflict is irrational has two options. He can say that since the desires presented above conflict directly, they must be irrational. Or he can say that since they're rational, the conflict must be indirect.

Now, if one really plans on insisting that intrinsic desires for challenges are irrational, there's probably not much to say. But let me try an appeal to common sense. Desiring a challenge for its own sake is central to our sense of what makes life worth living. So is the desire to encounter the aversive in art, ranging from sad music to disturbing paintings to scary movies. Given their importance to us as human beings, both sorts of desires deserve a strong presumption of rationality.

But aren't these desires inconsistent? And doesn't that provide an argument that overrides the presumption of rationality? Consistency, after all, provides our basic, uncontroversial grip on what rationality is. (Davidson, 2004.)

Let's begin with a distinction. When we talk about consistency of attitudes, we can be talking about the consistency of their objects—i.e., their contents—or we can be

talking about the consistency of the attitudes themselves. Directly conflicting desires have inconsistent contents. Are the attitudes inconsistent?

Beliefs with inconsistent contents are obviously inconsistent, but why? One reason is that beliefs play an inferential role, so they're subject to principles like conjunction introduction, *modus ponens*, *tollens*, etc. We should be skeptical that similar principles exist for desires, both because they don't play an inferential role, and on intuitive grounds. For example: Bob desires to date Sally, and desires to date Sue, without desiring that he date both. (Williams, MS.) Some argument on behalf of conjunction is needed, before we make it a principle of desire.

Or consider what looks like a reasonable principle of consistent desires: the instrumental principle. Desiring  $p$  and believing that  $q$  is a means to  $p$  might require, for consistency's sake, that one desire  $q$ . But an agent who believed that  $p$  and believed that that  $q$  is a means to  $p$ , yet failed to conclude that  $q$ , would not demonstrate inconsistency.

If conflicting desires are consistent, then we need a principle instructing agents with conflicting desires to prioritize one over the other—perhaps on the basis of strength, or connection with the agent's sense of self, so that agents can act consistently despite the conflict.

Simon Blackburn seems to assume that no such principle is available when he writes:

Incompatible and therefore unrealizable goals are bad in a way quite analogous to the way in which inconsistent beliefs are bad. The latter cannot represent the world properly; but the former cannot represent how to behave in the world properly: they cannot mate together with beliefs, in the usual belief-desire psychological framework, to direct effective action.  
(1988/1993: 190)

But we can say in response that as long as one desire is privileged over its rival, that desire can direct effective action. The trumped desire won't—but that seems to just be a case of sacrificing satisfaction of one desire for the sake of a more important one. When making choices we do this all the time, even if exactly what it is that allows us to privilege some desires over others is vague.

In summary, we don't have a clear enough idea of what consistent desire amounts to, to say that directly conflicting desires are inconsistent. Any intuition that they must be comes from assuming that consistency of attitudes depends on consistency of contents, or that consistent desire must closely resemble consistent belief. These are bad assumptions. We should admit it's mysterious what consistent desires might be.

So what about the second objection? When we have desires that seem to conflict directly, are there really two states of affairs that are desired that conflict only contingently?

Let's take the case where one state of affairs is desired merely instrumentally, as a means to achieving some end.<sup>3</sup> It's true that we often desire challenges as a means to some benefit. But sometimes people desire challenges for their own sake. Everyone knows the experience of losing interest in a task (even a task with benefits) precisely because it is easy, and the success is guaranteed. If we only ever desired challenges instrumentally, our desire should always increase as the task got easier, and the chance of success increased.

---

<sup>3</sup> Under some versions of the instrumental principle states of affairs constitutive of the states of affairs we desire are themselves desired instrumentally; however, they are desired as necessary means to the end, and not as contingent means. On this sort of view, that one endure pain would be desired as a state instrumental to one's participation in a challenge. But since it is a state necessary to participation in a challenge, the agent would still have directly conflicting (i.e., impossible to jointly satisfy) desires. The aversive state of affairs must be desired as a mere means to an end, and not a constituent of it, if these are really cases of indirect conflict.



Reinterpreting the desires we have when we watch scary movies so as to make them instrumental would be even more inappropriate. Admittedly, an agent might watch lots of scary movies so that he doesn't frighten easily; or because he's jaded, and only extreme terror alleviates his otherwise chronic boredom. But these are not normal cases.

But there are two other ways of partitioning desires.

The first makes the desires in question vicarious. The marathon runner doesn't really want to risk losing. He wishes to experience what it would be like to succeed at a race he takes to be risky. If he could arrange a situation where his success was guaranteed, but he wouldn't know that it was guaranteed, the runner would prefer it to an honest race.

But as soon as it's stated this way it sounds totally implausible. Not every person who seems to intrinsically desire a challenge in fact only desires the *sense* of a challenge, in order to build up suspense so that their victory is that much sweeter.<sup>4</sup>

As for the scary movie, I have no idea what it is to watch a movie vicariously.

In the second sort of partitioning the agent in question is averse to a part, but desires some whole. The runner is averse to pain, say, but desires the whole of a challenging race, of which one part is pain.

But what sort of case are we imagining here? If the agent is genuinely ambivalent, he must be averse to the part by itself (pain), while desiring the whole *including* the part (the race including the pain). But this is just the case of desiring *p* and

---

<sup>4</sup> Robert Nozick's (1974/2001) example of the Experience Machine illustrates quite well that our desires are rarely about subjective states, and more often directed at events out in the world. A runner doesn't want to experience what it's like to win a race; he wants to win it.

desiring  $q$ , when  $q$  entails  $\sim p$ . (I assume the runner knows that a painful race entails pain.) This is still a case of direct conflict.

It only ceases to be direct conflict if the runner's aversion to pain is made more specific: he's averse to pain except when it's part of a marathon. But then there is no direct conflict only because there is no conflict at all. People can have motivational setups like this. But this possibility only helps us maintain the irrationality of direct conflict if we're willing to count any ambivalence towards the hardship involved in a challenge as irrational.

The strong presumption of rationality the examples were entitled to has not been overcome. And the examples are cases of direct conflict. So rational cases of direct conflict are possible.

### **3. Problems for Reduction**

Reductionism claims that facts about desires reduce to facts about preferences. That an agent wants  $p$  reduces to preferring  $p$  to certain potential alternatives. That an agent wants  $p$  more than  $q$  reduces to the agent's preference for  $p$  over  $q$ , and how much more he wants it reduces to facts about which gambles with  $p$  as a potential prize he'd prefer to  $q$ . (Dreier, 1996.)

There are two major theoretical virtues to this position. First, as noted earlier, it's unclear what consistency means with respect to desire. The reductive thesis gives us a non-arbitrary answer: consistent desires supervene on consistent preferences, and preferences are consistent when they obey the axioms of Decision Theory. (Dreier,

1996.) Second, it's sparse. Everything about an agent's desires is accounted for in terms of preferences over options. All the theory postulates is these preferences, and patterns in how the agent prefers.

The theory requires, though, that an agent's ambivalence towards some option reduce to patterns in an agent's preference—i.e., for every case of ambivalence in desire, there must be some set of preferences which *just is* that ambivalence about the option.

Jackson offers the most detailed account of how patterns of preference would underwrite conflicts in desire. I will briefly describe Jackson's account and show that it cannot deal with cases of direct conflict adequately, and then explain why it is highly unlikely that any patterns in an agent's preferences could underwrite cases of direct conflict.

#### Jackson's Account of Conflict and the Objection to Preference Analysis

Jackson accounts for ambivalence, or what he calls internal conflict, by means of distinction between desires for states of affairs, and desires for properties. Desires for states are simple:

**Desire for States Thesis:** An agent desires state of affairs  $p$  in virtue of preferring  $p$  to its alternatives.

So, 'The agent desires the state of affairs of owning a car' is true in virtue of the agent's preference for owning a car over the alternative of not owning a car.

The desire for a property is more complicated. It's a tendency on the part of the agent to desire states of affairs that have the property to otherwise identical states of

affairs that lack it. Considering the case of an academic torn between grading papers and drinking wine with dinner, Jackson writes:

Let  $S_L$ ,  $S_W$  be states of affairs that would obtain if I were to choose, respectively, to mark the logic exercises or to have the wine. Then my preference for the marking derives from the fact that I rank  $S_L$  over  $S_W$ . But this is consistent with my ranking  $S_W$  over  $S_{W\text{-without-the-wine}}$ ; and I do in fact judge  $S_W$  the better for including the wine.  
(1985:111)

So this gives us:

**Desiring Properties Thesis:** The fact that an agent desires some property  $F$  can be reduced to facts about the consistent preference rankings of the agent. An agent desires some property  $F$  in virtue of consistently ranking states of affairs in which  $F$  obtains higher than otherwise identical states of affairs in which  $F$  does not obtain. An agent is averse to  $F$  in virtue of consistently ranking states of affairs with  $F$  lower than otherwise identical states of affairs without  $F$ .

The reductionist can now explain the conflict between desires as a conflict between a desire for some property, and a desire for some state of affairs. With desires for properties added to the account, we can say that the instructor desires the property of having wine, but desires a state of affairs in which he does not have wine. This gives us:

**Conflict Thesis:** In any case where an agent's desires are in internal conflict, this means that the agent desires some property  $F$ , and also desires some state of affairs  $p$ , and  $F$  does not obtain in  $p$ , and there is some alternative to  $p$ ,  $q$ , and  $F$  obtains in  $q$ .

Unfortunately, preference reduction cannot account for cases like desiring a challenge for its own sake, or wanting to experience the unpleasant in art. In the cases of the marathon runner or the scary movie the same property plays a role in making the desired state of affair both appealing and aversive. The agent desires to run in the marathon because it has the property of being a challenge, and is averse because it has the

property of being a challenge. The property of being frightening makes the movie both attractive and aversive to the audience member.

This leads to a problem with the Desiring a Property Thesis. Desiring a property is defined, like desiring a state of affairs, in terms of the agent's preferences. An agent desires the property  $F$  just in case he prefers states of affairs with  $F$  to otherwise identical states of affairs without  $F$ . But this means that an agent who desires and is averse to one and the same property  $F$  will, for a partial state of affairs  $p$ , prefer  $p$ -with- $F$  to  $p$ -without- $F$ , and also prefer  $p$ -without- $F$  to  $p$ -with- $F$ . He will prefer a challenging race to a non-challenging race, and a non-challenging race to a challenging one. His preferences must be inconsistent.

Has the case been oversimplified? After all, it's not really the challenge that our hypothetical runner is averse to. It's the pain, the fatigue, and the possibility of failure. Even if we grant that every challenging race has those properties, there's no reason to think that the entailment goes in the other direction. A race might just be painful and involve the risk of failing without counting as a challenge. But now we can make use of the Conflict Thesis. Let's call the collection of aversive features which *sometimes* make something a challenge, property  $G$ . And let's call being a challenge property  $F$ . So, we can say that the agent prefers the challenging race to the non-challenging, i.e., prefers  $p$ -with- $F$  to  $p$ -without- $F$ . But the agent also prefers, generally, situations in which  $G$  does not obtain. So the agent counts as desiring  $G$ 's absence.  $G$  obtains whenever  $F$  does, so the agent must necessarily be ambivalent. But this is the result we were looking for.

The problem with this route is that it forces the reductionist to collapse two distinct types of motivational situations. An agent may be ambivalent about some state-

of-affairs. He may also have a whole-hearted desire for a very specific state-of-affairs, any deviation from which he finds unpleasant. As an example of the latter, my brother, when he was very young (about six or seven), hated cheese—unless it was melted, in which case he liked it. So let's call the property of eating cheese *G*, and let's call the property of eating melted cheese *F*. But now our way of representing my younger brother's motivations is identical to how we were representing the marathon runner's motivations. All the same, there's no reason to assume my brother was ambivalent about eating something with melted cheese on it. He was just picky.

To return to the scary movie example, an agent may have standing desire not to see gory things, yet still prefer that the movie be gory. This latter preference can have two explanations, however. On the one hand, the agent may find that the gore makes the movie scarier; on the other, the agent may find gore in movies campy and funny. But then we have a problem for the reductionist: the agents will have identical preferences with respect to gore, but only one of them is ambivalent about movie gore.

We can summarize the problem as follows: the Desiring Properties Thesis doesn't allow us to distinguish between cases where an agent is averse to *G*'s that aren't also *F*'s (cheese that isn't melted; gore that isn't in a movie; pain that isn't part of a jog) from simple aversion to *G*'s (cheese; gore; pain).

Now there may be some other way reductionists can account for ambivalence, which would allow us to talk about cases of directly conflicting desires. But I suspect the problem with Jackson's account will generalize. Jackson's account couldn't distinguish certain cases of ambivalence from cases of very specific but wholehearted desire. The problem is that in cases of directly conflicting desires, all those states of affairs at which a

given desire is aimed are also states at which an aversion must be aimed. This means that one of the attitudes will consistently mask its rival in the agent's preferences. This masked attitude (whether the desire or the aversion) will fail to show up in any pattern of preferences, because there is no possible situation in which it could appear unmasked.

In order for the reductionist to get around this problem, he must accept a heavy theoretical commitment. He must deny that an aversion is ever really aimed at the same states of affairs as a desire. This means that desires and aversions are never really in conflict. The conflict, on this view, results from difficulty in forming a suitable means-end belief.

The agent, on this interpretation, has the following preferences: he prefers watching an excellent horror movie to a quiet night alone, but he also prefers a quiet night alone to seeing a really gory scene in a mediocre film. So the options he faces as the particularly gory scene appears on the screen are: (1) a quiet night alone, or (2) a gamble with the possible outcomes of watching an excellent horror movie or else of seeing a lot of gore in a less than excellent movie. What the agent should do here depends on the probabilities he assigns to that gamble. How likely is it that the movie is excellent, and how likely is it that he's simply torturing himself by watching it?

This position is consistent. Unfortunately, it boils down to the claim that desires are never really in conflict with one another. All apparent conflict is really just confusion about the situation one is in—uncertainty about the probable results of one's actions. The scary movie watcher is tempted to look away because he's not sure whether the movie is a good movie or not. Were he better informed, if, for example, he were certain that the movie was truly an excellent one, he would not feel any urge to look away from the

screen, stop the movie, or turn on the lights during the suspenseful or gory scenes. This might be true, but it involves a revisionary account of conflict in desires. A strong argument on behalf of the thesis is needed before we give up on the common sense view of these matters. (It seems a matter of common sense that I could know *The Shining* is an excellent film, yet feel some desire to stop the movie during particularly scary scenes.)

Additionally, the case of the marathon runner poses a special problem for this sort of revision. If challenges require some degree of risk, and it seems that they do, then it's unclear that the agent could be certain of a successful outcome, but still regard the race as a challenge.

In any case, rational direct conflict cannot be accounted for in terms of preference. So, unless we are willing to claim that all conflict is a result of confusion about the situation one is in, we're left with an unreduced practical attitude—the desire.

#### 4. Conclusion

Reducing desires to preferences is the first necessary step in offering Decision Theory as a theory of the Humean aspect of practical rationality—the aspect that's concerned only with consistency among attitudes, not substantive values. But if there are practical attitudes, which do not reduce to preferences, we should suspect that there are rules governing those attitudes, and these rules would be among the rules of Humean rationality.

For instance, let's assume reduction works if we restrict ourselves to all-things-considered desires. All things considered, I want to visit the dentist, or I don't; I want to



run the marathon, or I don't. But this means that there are pro-tanto desires which don't reduce to preferences. And presumably there are rules regarding which all-things-considered desires an agent should have, given certain pro-tanto desires as inputs. These rules—rules in excess of those of Decision Theory—would then need to be part of Humean rationality. (For another example of rules of rationality exceeding those of Decision Theory, on the basis of a psychological state that is not a preference, see Pettit 1991/2002.)

Another way of putting this problem: a theory of Humean rationality is supposed to be a theory of consistency for desires. Desires conflict, so it's not obvious what consistency means. Preference reduction led to an attractive answer, but the reduction itself failed. So we're still lacking a complete theory of Humean rationality.

### **Acknowledgments**

Thanks to Lara Buchak, Arudra Burra, Corrine Gartner, Elizabeth Harman, Lloyd Humberstone, Frank Jackson, Mark Johnstone, Tristram McPherson, Alexander Nehamas, Philip Pettit, Gideon Rosen, Michael Smith, and my anonymous referee for feedback, criticisms, and help making points more clearly.

## Bibliography

- Blackburn, Simon, "Attitudes and Contents," in *Ethics*, 1988; reprinted in Simon Blackburn *Essays in Quasi-Realism*, Oxford University Press, 1993.
- Broome, John, "Can a Humean Be Moderate?" in *Value, Welfare, and Morality*, edited by R. G. Frey and Christopher Morris, Cambridge University Press, 1993, pp. 51–73.
- Davidson, Donald, "Actions, Reasons, and Causes," in *Journal of Philosophy*, 1963; reprinted in Davidson, *Essays on Actions and Events*, Oxford University Press, 1980.
- \_\_\_\_\_, "Incoherence and Irrationality," reprinted in *Problems of Rationality*, Oxford University Press, 2004, pp. 189-199.
- Dreier, James, "Rational Preference: Decision Theory as a Theory of Practical Rationality," *Theory and Decision* 40:3, 249-276, 1996.
- Edgington, "On Conditionals" in *Mind* 104:414, 1995, pp. 235-329.
- Harman, Gilbert, "Desired Desires," in Ray Frey and Chris Morris (eds.), *Value, Welfare, and Morality* (Cambridge University Press, 1993), pp. 138-157; reprinted in *Explaining Value and Other Essays in Moral Philosophy*, Oxford University Press, 2000, pp. 115-136.
- Jackson, Frank, "Internal Conflicts in Desires and Morals," *American Philosophical Quarterly*, April, 1985.
- Nozick, Robert, *Anarchy, State and Utopia*, 1974, BasicBooks; section on the Experience Machine reprinted as "The Experience Machine," in George Sher (ed.), *Moral Philosophy: Selected Readings*, Wadsworth, 2001.
- Pettit, Philip, "Decision Theory and Folk Psychology", in Michael Bacharach and Susan Hurley, eds., *Foundations of Decision Theory*, Blackwells, Oxford, 1991; reprinted in *Reason, Rules, and Norms: Selected Essays*, Oxford University Press, Oxford, 2002.
- Smith, Michael, "The Humean Theory of Motivation," in *Mind* 96, 1987, pp. 36-61; reprinted in *Reason, Emotion and Will* edited by R. Jay Wallace (Aldershot: Ashgate Publishing, 1999) pp.3-28.
- Williams, John, "Kinky Desires: Why There Is No Moore's Paradox of Desires," Manuscript.