The Humean Theory of Motivation Reformulated and Defended

Neil Sinhababu
National University of Singapore

David Hume's famous dictum that "Reason is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them" (2.3.3) has inspired a wide variety of theories concerning motivation.[1] The version of the Humean theory of motivation that I will defend here stands among them and consists of two propositions:

> **The Desire-Belief Theory of Action [DBTA]**: Desire is necessary for action, and no mental states other than a desire and a means-end belief are necessary for action.[2]
>
> **Desire Out? Desire In! [DODI]**: Desires can be changed as the conclusion of reasoning only if a desire is among the premises of the reasoning.

I will refer to philosophers who reject at least one of these propositions as anti-Humeans.

This formulation of the Humean theory is stronger than the one that Michael Smith presents in *The Moral Problem*.[3] Smith does not include any principle analogous to DODI. The effect of adding DODI to DBTA is

1. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1978).

2. DBTA is formulated so that one may still be a Humean whether or not one holds that desires to engage in immediate bodily movements—for example, a desire to move my hand right now—can cause action without the assistance of a means-end belief. The notion of "means-end belief" intended here is broad enough to include constitutive means as well as causal means.

3. Michael Smith, *The Moral Problem* (Oxford: Blackwell Press, 1994).

Philosophical Review, Vol. 118, No. 4, 2009
DOI 10.1215/00318108-2009-015

to make desire an essential precondition of practical reasoning. If we interpret Hume's dictum above as proposing a fundamental role for desire in grounding all motivation, we may regard formulations of the Humean theory that include DODI as truer to its spirit. If DODI is left out, and desires can be changed through reasoning without any input from antecedently existing desires, reason seems to be passion's master rather than its slave.

Opponents of the Humean theory sometimes attack it by trying to offer counterexamples. They present situations in which it seems that human agents deliberate and act in a manner contrary to the Humean theory. W. R. Sorley[4] and J. G. Schurman[5] claim that the Humean theory cannot provide adequate explanations of how we feel when we are motivated by the feeling of obligation. Stephen Darwall presents a situation in which he claims that someone generates a new motivation through reasoning without an antecedent desire.[6] Thomas Scanlon offers cases in which he claims that agents bracket the motivational force of their desires, preventing some of them from motivating action.[7] John Searle argues that the Humean theory is unable to explain how we deliberate and act in cases of *akrasia*.[8]

We can understand these proposed counterexamples as challenges to the explanatory power of the Humean theory. If Humeans cannot provide satisfactory explanations of how we deliberate and act in all cases, we will have reason to reject their theory in favor of one with sufficient explanatory resources to handle the cases in which they fail. The Humean theory offers us the attractive promise that a simple explanation invoking only desire-belief pairs for motivation will be sufficient to account for all cases of action. If this promise cannot be kept, we will have reason to go to theories drawing on a more expansive set of explanatory resources—perhaps theories according to which beliefs about our reasons are capable of causing action or generating new motivational forces without any assistance from desire.

I will show that the Humean theory can provide satisfactory explanations in all of the above cases that anti-Humeans have offered as

4. W. R. Sorley, *Moral Values and the Idea of God* (Cambridge: Cambridge University Press, 1919).

5. J. G. Schurman, "The Consciousness of Moral Obligation," *Philosophical Review* 3 (1894): 641–54.

6. Stephen Darwall, *Impartial Reason* (Ithaca, NY: Cornell University Press, 1983).

7. T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

8. John Searle, *Rationality in Action* (Cambridge, MA: MIT Press, 2001).

counterexamples. In fact, considering these cases shows us that the Humean theory offers us a better overall explanation of how we deliberate and act than its competitors do. The elegant and powerful explanations that Humeans can offer in the cases described by anti-Humeans show us why we ought to accept the Humean theory.

To develop these Humean explanatory stories, I will first discuss the properties of desire. Understanding the many kinds of effects that desire has on our psychology is essential to giving detailed explanations of how we deliberate and act. After setting out the features of desire, I will build explanations of what happens in the situations that the anti-Humeans present as counterexamples.

But first, I should point out the modal modesty of the version of the Humean theory that I will defend and make clear that it nevertheless retains its metaethical significance. Some Humeans regard the Humean theory as a conceptual truth about action. On their view, all possible agents have Humean psychologies. I regard the Humean theory merely as a truth about actual actions performed by human beings and about all the actions that humans are psychologically capable of performing.[9]

Even if the Humean theory is only true within the space of human psychological possibility, its truth will be significant for debates in metaethics. Consider the puzzle that Michael Smith presents in *The Moral Problem*, where he describes the tension between cognitivism, internalism, and the Humean theory. The puzzle runs as follows: If cognitivism is true, moral judgments are beliefs. If internalism is true, moral judgments have intrinsic motivational force. But if the Humean theory is true, there are no beliefs with intrinsic motivational force. So if cognitivism, internalism, and the Humean theory are all true, there can be no moral judgments. To avoid this consequence, at least one of these three positions must be abandoned. Smith then proposes a way of holding cognitivism, internalism, and a weak version of the Humean theory that does not include anything like DODI. But the stronger version of the Humean theory that I defend here cannot be held in conjunction with both cognitivism and internalism.

The puzzle still remains if one regards the Humean theory as a contingent truth concerning human psychology rather than a necessary truth applying to all possible agents. This is because cognitivism and internalism are not mere possibility claims—they are supposed to hold at least of the moral judgments of human beings in the actual world. Cognitivism is

9. In fact, I think that the Humean theory is not a necessary truth. For reasons of space and focus, I will not discuss nonactual counterexamples to the Humean theory here.

supposed to apply at least to the moral discourse of actual human beings. Internalism is usually regarded as a conceptual truth about the connection between moral judgment and action, and thus must be true in the actual world and beyond.[10] So if we accept the Humean theory as a truth about actual human psychology, we cannot maintain the usual versions of both cognitivism and internalism, which apply to the actual world. Nonhuman agents with non-Humean psychologies permitting moral judgments that satisfy both cognitivism and internalism may still be metaphysically possible. But if the Humean theory is true about actual human psychology, as I will argue, our moral judgments cannot satisfy both cognitivism and internalism. So even a modally modest version of the Humean theory will have great import for metaethics.

The Aspects of Desire

Five aspects of desire are particularly important for the Humean explanations I will offer. When appropriately combined with other mental states, desire motivates action, causes experiences of pleasure and displeasure, directs attention, and can be made more violent (as Hume would say) by vivid representations that one associates with its object. It also comes in what might be called two different flavors—positive desire and aversion. I will go over the various aspects of desire in turn.

Probably the most obvious feature of desire is its Motivational Aspect. In combination with the appropriate means-end beliefs, desires can motivate action. I characterize the Motivational Aspect of desire as follows:

> **The Motivational Aspect**: If agents occurrently desire D, and they occurrently believe that they can bring about D by doing A, they will be motivated to do A. The strength of their motivation will increase with the strength of the desire and the subjective probability that they can bring about D by doing A. If at any time there is some action that they are the most motivated to do, they will initiate that action.[11]

10. William Frankena defined internalism in terms of whether it was "logically possible for an agent to have or see that he has an obligation even if he has no motivation, actual or dispositional, for doing the action in question." See "Obligation and Motivation in Recent Moral Philosophy," in *Essays in Moral Philosophy*, ed. A. I. Melden (Seattle: University of Washington Press, 1958), 10.

11. I talk about "initiating" rather than "performing" actions to deal with cases like that of agents who desire to push a button and believe they can push the button by moving their hand but do not know that they are paralyzed. Such agents cannot perform the action of pushing a button but can initiate it.

The reason for talking about "motivation" as well as action itself is to develop an account that explains what happens when we have many different desires and beliefs in play, and we have to choose between different options. Not every occurrent desire-belief pair results in action—some motivate actions that are inconsistent with actions motivated by stronger desire-belief pairs. In these cases we act on the stronger pair and not the weaker one.

Desire is also connected to pleasure and displeasure. This is not the connection that the psychological hedonists suggested—all sorts of things other than pleasure can be the objects of our desires. The connections between pleasure and desire have to do with our experiences upon having some vivid sensory or imaginative representation that we associate with what we desire, or when we believe that the object of our desire is more or less likely to be achieved.

> **The Hedonic Aspect**: If agents occurrently desire D, increases in the subjective probability of D or vivid sensory or imaginative representations of D will cause them pleasure roughly proportional to the strength of the desire and the change in subjective probability or the vividness of the representation. Decreases in the subjective probability of D or vivid sensory or imaginative representations of situations incompatible with D will likewise cause displeasure.

We can see evidence of the Hedonic Aspect in the way that surprises that raise the subjective probability of a desired state please us and how our unpleasant surprises are those that suggest that our desires are less likely to be satisfied than we previously thought. And on the side of imagination, it is pleasant to daydream about the things that we desire and that we would act to bring about.

Desire can also direct an agent's attention. Suppose you are hungry, and you walk into a kitchen where there is a coconut cream pie on the table. Your attention will focus on the pie—particularly on the features of the pie suggesting its deliciousness—and not the wood grain on the table or the hum of the refrigerator. Of course, if you have particularly intense desires associated with wood grain or appliances, your attention might focus on these things.

> **The Attention-Direction Aspect**: Desiring that D will make agents more likely to focus their attention on things they associate with D than things they do not associate with D.

As there are things unrelated to desire (like sudden noises) that can direct one's attention, the Attention-Direction Aspect is framed in terms of the

contribution of desire to attention-direction and not as an explanation of all attention-direction. "Things" in the above formulation is to be read very broadly. It includes not only physical objects that we associate with our desires but actual or counterfactual states of affairs that we associate with our desires. Daydreaming (where we direct our attention toward counterfactual states of affairs) and basking in pleasant memories (where we direct our attention toward past states of affairs) can be understood as psychological phenomena in which the Attention-Direction Aspect manifests itself.

Another interesting fact about desire is that it comes in what we might call two "flavors," each with a different emotional profile. While all of our desires exhibit the Hedonic Aspect, the sorts of pleasure and displeasure we feel under changes in subjective probability of satisfaction or under conditions of vivid sensory or imaginative representation are sometimes different. Some desires, like the desire for a delicious meal, give us a delighted happy feeling when we find that we can satisfy them and an unpleasant feeling of disappointment when we discover that we cannot. Others, like the desire not to miss one's flight, give us the pleasure of relief when we find that we can satisfy them and an unpleasant feeling of anxiety or dread when we discover that we cannot. This gives us reason to divide the category of desire into two subcategories, positive desire and aversion.

> **The Two Flavors**: Agents who desire that D either have positive desires that D, or aversions to not-D. The pleasures and displeasures associated with positive desires are delight and disappointment; the pleasures and displeasures associated with aversions are relief and anxiety.

As will be discussed later, people sometimes use "desire" merely to refer to what I am calling "positive desire." I will use "desire" to encompass both positive desire and aversion. That many of our actions are motivated by aversion is thus consistent with the Humean theory.

Hume himself noted the last aspect of desire to be discussed here—the way it can be intensified by vivid sensory or imaginative representations of the desired state. When Hume distinguished between calm and violent passions, he also noted that by varying "the situation of the object," we can "change the calm and violent passions into each other." Making the agent perceive its object with more "force and vivacity" will increase the violence of a passion (2.2.7). Hume offered some wonderful examples of this. He cites the greater violence of passions for recently tasted pleasures and the motivational power of rhetoric that causes its audience to vividly imagine the objects of passion. He discusses the way we imagine things close

to us in time more vividly and how this causes our passions for those things to be more violent. He also offers an involved historical example from ancient Athens. Themistocles thought up a plan to give Athens naval supremacy by launching a secret mission to burn the ships of all the other Greek kingdoms, which were gathered in a nearby port. Since other kingdoms would learn of the plan and take appropriate precautions if he expressed it openly, he only told the Athenians that he had a secret plan that would benefit them greatly. The Athenians had him explain the plan to Aristides alone, whose judgment they completely trusted. Aristides reported back to the Athenians that the plan would be greatly advantageous to Athens but terribly unjust. Upon hearing this, the Athenians unanimously voted against the plan. Hume rejects the view of a historian who claims that this shows the great intensity of the Athenians' desire for justice. As Hume points out, the Athenians were able to conceive of the plan only in the general terms of justice and advantage. The notion of advantage, being a very general idea, is not conducive to vivid imagining. Had the Athenians been presented with the prospect of naval supremacy, which allows for more vivid imagining of things ancient Greeks liked to do, like destroying enemy ships and raiding coastal kingdoms, more violent passions in support of Themistocles' plan would have been incited, and they might well have decided otherwise.

On Hume's view, an increase in the violence of a passion does not bode the same way for its phenomenal, motivational, and attention-directing effects. He says that "Tis evident passions influence not the will in proportion to their violence, or the disorder they occasion in the temper" (*A Treatise of Human Nature*, 2.3.4) and describes how the force of custom can create strong but calm passions. While he accepts that increases in the violence of a passion increase its motivational effects, as in the case of the Athenians, he thinks that the change in the passion's phenomenal effects is greater, and he does not discuss increases in the passion's attention-directing effects. So following Hume, we might say that increases in the violence of passions dramatically increase their phenomenal effects, substantially increase their motivational effects, and do the least to increase their attention-directing effects:

> **Intensification by Vivid Images**: When agents are presented with vivid images they associate with a state of affairs they desire, either in imagination or by their senses, that will strengthen the desire's causal powers. The desire's phenomenal effects increase greatly, and its motivational powers increase substantially as well.

This should be understood as a claim about the immediate effects of vivid images on desires, not their overall effects once they have triggered other psychological processes. This is why cases in which vivid sensory images reduce motivation by producing beliefs—for example, when someone wants to eat a strange fruit but then sees another person become sick after eating it—are not counterexamples to the above principle. Whatever intensification of desire may have arisen from initially seeing the fruit is overwhelmed by the new belief in the bad consequences of eating the fruit. To consider a different case, if I desire to get married and then see a couple in an unhappy marriage or vividly imagine myself unhappily married, perhaps I will be less disposed to pursue marriage. But the principle above can explain this—I have a variety of aversions to a strife-torn marriage, and the vivid image of marital strife intensifies them more than it intensifies my desire for the benefits of marriage. The claim that vivid images intensify desire, then, should be understood not as an all-things-considered claim about their overall effects but as a *pro tanto* claim about immediate effects that may be overwhelmed by other less immediate effects or by immediate effects on opposing desires.

Before moving on to the counterexamples that anti-Humeans have offered, I should discuss a more general objection that some of them present.[12] The Humean theory, they say, is true only on the weak reading of 'desire', where a desire is any mental state capable of causing action in combination with a means-end belief. But when we regard the Humean theory as a substantial and interesting theory, we read 'desire' as something stronger and in fact as something strong enough to make the Humean theory false. The anti-Humeans warn us not to be tricked by an equivocation where the truth of the Humean theory with the weaker notion of 'desire' lends plausibility to a version of the theory with the stronger notion.

I hope to show that the substantial and interesting form of the Humean theory can be defended without recourse to any such equivocation. I have offered an account of the aspects of desire that will not make the Humean theory trivially true, and I will not argue that all conceptually possible agents have desires. Instead, I will show that with the stronger notion of desire, the Humean theory is capable of providing superior

12. G. F. Schueler's *Desire* (Cambridge, MA: MIT Press, 1995) devotes much effort to developing an objection along these general lines. The issue is also raised by Darwall in *Impartial Reason*, Scanlon in *What We Owe to Each Other*, and John McDowell in "Are Moral Requirements Hypothetical Imperatives?," *Proceedings of the Aristotelian Society Supplementary Volume* (1978): 13–29.

explanations in the cases that anti-Humeans regard as counterexamples to it. In several of the cases I will discuss, the emotional experiences that we have while deliberating will provide evidence for the Humean theory. If we did not have these and other experiences predicted by the Humean theory, we would have reason to reject the theory as one that made false predictions. But as it turns out, the Humean theory explains and predicts our behavior, emotions, and other mental states more simply than any of its competitors. This gives us a reason to accept it.

The Feeling of Obligation

A classic objection to the Humean theory of motivation is that it cannot explain the way we are motivated when we act out of a feeling of obligation. This objection to the Humean theory is grounded in a genuine phenomenological datum. Our feelings of obligation differ in significant respects from the other feelings we have when spurred to action, and I will lay out two distinctive ways that we can feel when we act out of obligation. But as I will argue, the Humean theory can deliver superior explanations of how we feel in both of these cases. It elegantly explains the same phenomenological data while committing us to fewer kinds of motivational processes. Its simpler explanation of the phenomenology of deliberation gives us reason to accept it.

Some opponents of the Humean theory have claimed that the feeling of obligation arises intrinsically from some kind of truth-evaluable mental state—for example, the belief that particular moral facts obtain or the judgment that a particular maxim could be a universal law for all rational beings. Since desire is not truth evaluable, this claim stands in contradiction to the Humean theory. Immanuel Kant, with his distinction between the autonomy of the will when it is in accordance with duty and the heteronomy of the will when it is driven by desire, is the most famous representative of this strand of the anti-Humean tradition.[13]

This talk about the "feeling of obligation" is not intended to suggest that there is some psychological state that is sufficient for the existence of an obligation. Neither does it imply that the existence of an obligation is sufficient for the existence of this state. An irresponsible person may be under an obligation and still not experience the feeling of obligation. Similarly, someone may mistakenly believe that he or she is under an

13. Immanuel Kant, *Practical Philosophy*, ed. and trans. M. J. Gregor (Cambridge: Cambridge University Press, 1996).

obligation and experience the feeling of obligation even though no obligation is present. And in many cases, genuine obligations are successfully discharged by people who do not have this feeling at all. I am using the term 'feeling of obligation' to pick out a particular experience or group of experiences that many anti-Humeans rightly regard as different from many of our ordinary experiences of desire and that commonly arise in cases where we act out of obligation.

Why would anyone think that the motivational force that drives us when we have the feeling of obligation springs from a place other than desire? Among the reasons for holding this position is that the feeling of obligation is phenomenologically different from our most common feelings of desire. Given the uniqueness of the feeling of obligation, it may not seem plausible that a reduction of motivation under the feeling of obligation to motivation by desire is possible.

The view that there are important phenomenological differences between desire-driven action and action done out of a feeling of obligation was expressed by some writers on ethics in the earlier part of the twentieth century. According to W. R. Sorley, something about the feeling of obligation is irreducible to our experiences of desire: "In all moral experience there is something which can not be simply identified with pleasure or with desire, but contains a differentiating factor which makes it moral and not merely pleasant or desired" (*Moral Values and the Idea of God*, 64). In "The Consciousness of Moral Obligation," J. G. Schurman defended the irreducibility of obligation and tied it to a cognitivist and internalist position about moral judgment:

> Confining ourselves, then, to the feeling of moral obligation alone, I think it must be said that this feeling is not susceptible of resolution into smaller elements, whether it be surveyed in its earliest or in its later state of development. It is an experience perfectly simple and unanalyzable, like the thought of being, clear to all who are conscious of it, but incommunicable to any one in whom that consciousness is wanting. Though in its nature the sense of moral obligation is an ultimate feeling, it is yet possible to designate the condition of its emergence in consciousness. That condition is the recognition of a moral law, ideal, or end of life. We are so constituted that what we recognize as right for us to do, that we feel we ought to do. (643)

Schurman continues by saying that "Moral obligation is the soul's response to acknowledged rectitude." According to Schurman, the experience of moral obligation is a sui generis feeling that follows the recognition of some kind of moral fact and that is capable of motivating action.

One part of what Schurman and Sorley say cannot be denied—there is a more or less distinctive set of feelings we have in many cases of obligation that is not present in many clear cases where we are motivated by desire. Suppose you have promised your students that you will grade and return their papers by tomorrow, and that you are a responsible person who takes these promises seriously. Just as you sit down to begin a long night of grading, friends of yours come by and say they are going to a party where many of your other friends will be present. Your emotions as you consider the prospect of keeping your promise and grading will be different from your emotions as you consider the prospect of going to the party. If you end up grading rather than going to the party, you may express these differences by describing your choice in terms that do not fit well with the Humean theory—"I'm doing what I have to do, not what I want to do."

These terms—'want to' and 'have to'—are exactly the terms in which John Searle puts his objection to Donald Davidson's inclusion of regarding something as "dutiful" or "obligatory" in the category of pro-attitudes.[14] Searle attacks Davidson, and Humeans generally, for blurring "the distinction between things you *want* to do and things you *have* to do whether you want to or not." Searle continues: "It is one thing to want or desire something, quite something else to regard it as 'obligatory' or as a 'commitment' that you have to do regardless of your desires" (*Rationality in Action*, 170).

It is not sufficient for Humeans to deal with the issue of obligation merely by positing desires to fulfill obligations—perhaps in the above case, a desire to keep promises that is stronger than the desire to go to the party. While this would successfully explain the agent's behavior in the case where he decides to keep his promise and grade the papers, it would fail to explain the phenomenology of decision making. As Schurman and Sorley say, the feeling of obligation is phenomenologically different from the feeling of the desire that motivates him to go to the party. Simply positing another desire can explain his behavior but more needs to be done to explain how the process of making the decision feels.

What exactly are the phenomenological differences between the feelings that arise from the two motivational forces in this example? At present the case is somewhat underdescribed, and I will consider two different ways it could go. Either way, the experiences associated with the two different motivational forces in the case are different from each other.

14. Searle, *Rationality in Action*, 169. Donald Davidson's position that Searle is attacking is in Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).

In the first case that I will consider, the grader seriously considers each of the choices before him and weighs whether to go to the party or keep his promise. As he does this, he will feel different emotions in considering the options before him. On the positive side, going to the party will seem exciting, while the possibility of handing a full set of graded papers back in the morning will generate more muted satisfactions. On the negative side, missing the party and grading will seem boring and dreary, while facing upset students in the morning without papers to hand back will incite anxiety and seem dreadful. I will call this the case of the Tempted Grader.

In the second case, the grader is focused on what he has to do and does not weigh the possibility of going to the party and leaving his promise unfulfilled. He will feel some disappointment at not being able to go to the party, but he will not seriously consider leaving the papers ungraded. While the desire to go to the party pulls at him, grading the papers will seem to have a kind of necessity, and he will not have the experience of weighing one desire against the other. The feeling generated by the motivational force that causes him to stay in the office and grade the papers will be less intense than the feeling generated by the motivational force that pulls him toward the party. But despite its lesser emotional vehemence, the former force will determine the course of his reflection and decision. This is the case of the Unwavering Grader.

Now I will respond to this objection by showing how the Humean theory, using the picture of desire that I offered before, can explain the feeling of obligation. As I will explain, the different emotions in the case of the Tempted Grader are neatly explained by the fact that desire has two flavors. Several different processes are involved in explaining the case of the Unwavering Grader, but they mostly come down to the effects of vivid images on our desires. We may lack vivid sensory or imaginative representations that we associate with failure to fulfill our obligations, or we may consider ourselves reliable moral agents and thus regard violating our obligations as too remote a possibility to vividly imagine. The feeling of constraint that can accompany obligation, moreover, is not unique to obligation but is present in other cases of desire. With an understanding of these factors in hand, Humeans can explain the feeling of obligation.

The motivational state that causes our actions when we act out of a feeling of obligation is not a positive desire to satisfy our obligations but an aversion to not satisfying them. As discussed previously, desire comes in two flavors with different emotional profiles. This explains the emotions characteristically associated with the feeling of obligation. When we

discover that we may not be able to satisfy some obligation we have, we feel anxious rather than disappointed. And if we are freed from an obligation that would be hard to satisfy, we usually feel relieved. While the emotions associated with positive desire are delight in cases of expected or imagined satisfaction and disappointment in cases of expected or imagined failure, the emotions associated with aversion are relief when we expect or imagine avoiding the object of aversion and anxiety when we expect or imagine failing to avoid it. Part of the experience of obligation can thus be explained by regarding the motivational force underlying the feeling of obligation as an aversion to not satisfying obligations rather than as a positive desire to satisfy them.

This is how the case of the Tempted Grader can best be explained. The different feelings associated with each of the options before him—the excitement of thinking about the party versus the duller satisfaction when he thinks about being able to hand back graded papers, the disappointment when he thinks of missing the party versus the anxiety when he thinks of breaking a promise—are accounted for by a view that grounds the different emotions in desires of different flavors.

Opponents of the Humean theory might claim, in response, that their account provides an equally simple explanation. They might invoke two different states with motivational power—a positive desire to go to the party and a belief that it is right to keep the promise and grade the papers, either of which generates the feeling of obligation as it motivates the action, either directly or by generating a desire through some DODI-denying process of practical reasoning. Humeans also have two states with motivational power—a positive desire to go to the party and an aversion to leaving one's obligations unfulfilled. Both of us explain all the phenomenological data. So why is the Humean explanation simpler?

The important thing to see here is that any plausible anti-Humean view will be committed to the existence of both aversion and positive desire as motivational forces. There are basic cases of emotions occurring before we act where positing a positive desire or positing a belief that it is right to perform some action will each fail to deliver good explanations. Consider a case where someone sitting outside is surrounded by bees and decides to move elsewhere to avoid being stung. His experience as he sees the bees will probably not be one of disappointment but one of anxiety, so a positive desire to avoid being stung will not explain his emotions. This is not a case where believing a moral principle is explaining his emotions or his decision.

An aversion to being stung, however, will nicely explain the motivation and the phenomenology. It is hard to see what better explanation could be offered here, and in the absence of such an explanation, I will take the anti-Humean to be committed, just as the Humean is, to the existence of aversion as a motivational force. In using the phenomenology of aversion as part of a reduction of the feeling of obligation, the Humean uses conceptual resources that are already on the table. Both sides need aversion in order to deal with cases like that of the bees. Rather than invoking a new primitive motivational force, the Humean builds a simpler theory by using a motivational force that both sides must allow.

An anti-Humean might think to explain the motivation and phenomenology in the case of the bees by appealing to principles of prudence that the agent accepts, which lead him to move away. But it is still unexplained why prudence creates this particular phenomenology in this situation. In some cases (for example, the case of an investor who thinks he can buy some land that will rise rapidly in value and subsequently learns that the land will not be sold after all), prudential motivation can have the phenomenology of excitement and disappointment that is characteristic of positive desire. So a mere appeal to prudential motivation will not do all the work.

My arguments for the superiority of the Humean theory, in this case and others, will seek to show its greater simplicity and theoretical unity. In the course of explaining some feature of how we deliberate, I will often invoke explanatory resources that an anti-Humean explanation of that particular feature of deliberation will not invoke, such as the two flavors of desire or the way desire can be intensified by vivid images. But these explanatory resources will be ones that the anti-Humean needs for other cases. Denying that aversions exist or that desires can be intensified by vivid images would leave the anti-Humean with no way of explaining why some of our desires generate different emotions than others or why desires for things that we see before us are more violent than desires for more distant things. The Humean, however, never needs to invoke motivation from a sui generis feeling of obligation. The beauty of the Humean theory is that it explains complex cases using a small set of explanatory resources that its opponents cannot deny, while leaving aside some of the resources they use. Thus it fits into a simpler total explanatory picture.

Now I move to the case of the Unwavering Grader. One interesting feature of many (though not all) cases when we act from a feeling of obligation is that we are not moved by what Hume would call a "violent passion." Even in many cases when we act to fulfill our obligations, and

particularly when we satisfy our obligations by refraining from action, the desire that determines our action is less intensely felt—one might say, with Hume, that it "creates less disorder in the temper"—than the desire that fails to cause action. The case of the Unwavering Grader is a case of this kind. How can a Humean explain this?

As Hume himself pointed out, passions become more violent when we have more vivid imaginative or sensory representations of things that we associate with their objects. There are two reasons why sensory and imaginative representations that go along with the feeling of obligation would be less vivid than the representations that go along with ordinary desire. First, the concepts that fit into the contents of our desires when we are motivated by the feeling of obligation are often quite abstract—for example, the concept of morality and the concept of obligation. We may not closely associate concepts at this level of abstraction with things that we can sense or vividly imagine. Second, people who reliably fulfill their obligations often have confidence in their abilities to do so. This makes them less likely to consider and imagine states of affairs in which the object of their aversion is realized and they fail to fulfill their obligations. I will explore these factors in more detail.

The things that we have aversions to when we experience the feeling of obligation are often fairly abstract. A conscientious person, for example, may have an aversion to doing things that are morally wrong. This aversion can affect whether and how he acts by combining with a means-end belief that by engaging in some action or by refraining from action, he would be doing something wrong. (The means in this case could be a constitutive means and not a causal means.) There may not be many sensory images that he closely associates with morally wrong action in the same way that we closely associate food with the content of our desire when we are hungry.

The objects of our aversion in cases where we experience the feeling of obligation are not always this abstract, of course. And in cases where we actually have vivid representations of the object of aversion, the feeling of obligation sheds its typical calmness and becomes unusually violent. An aversion to letting children suffer may grow violent after one sees images of suffering children. Thoughts of disappointed students make the aversion to not grading more violent. And while an aversion to marital infidelity may be a calm passion for a man who is far away from his wife, it can become more violent when he has some sensory experience connected with her—perhaps, when he is talking with her on the phone and hearing her voice, or when he looks into her eyes. However calm an aversion to

violating obligations may be when we are away from those to whom we are obligated, it will usually become more violent when we are looking into their eyes. In these cases, the passion moving us to action gains violence because of the vivid sensory representations that are before us.

The effect of vivid representations is also relevant in the cases of agents who know themselves to be reliable in fulfilling their obligations. These agents usually do not pause to think about possible states of affairs where their obligations go unfulfilled. Common processes that cause people to focus their attention on possible states of affairs related to their desires do not operate in their case, and this prevents them from considering these possibilities in much detail. Since they know themselves to be reliable in fulfilling their obligations, they are confident, come what may, that they will be able to fulfill the particular obligation that they are at the time motivated to fulfill. Possible states of affairs where they fail to fulfill their obligations seem remote to them, and these possibilities are not imagined vividly. In the absence of vivid representations of the states that they are averse to, their passions remain calm.

An example that does not deal with obligation may be helpful in explaining how this works. Suppose you were to present me with a choice where one of the options was very bad and selecting the other option was an easy choice. For example, suppose you offered to give my family $100 in exchange for my jumping out of a fourth-story window. I would not seriously consider jumping out, and I would reject your offer without seriously thinking about how it would be to fall to my death. Given the terms of the choice, the possibility of jumping out of the window would remain very remote, and I would not think about it enough to start vividly imagining the feeling of falling and the horrible impact of my body against the ground. So my desire to avoid an early death would decide my behavior while remaining calm. This is how the experience of decision making often is for people who are used to fulfilling their obligations. They have confidence that they will go forward and do the right thing, so it is not a usual part of decision making for them to look into the abyss and imagine how it would be if they failed to fulfill their obligations. Since they do not entertain vivid imaginative representations of failure, their passions remain calm.

Now consider a case in which I have to seriously consider jumping out the window. Perhaps some billionaire with strange and gruesome preferences appeals to my humanitarian sensibilities by offering to make a $100 million contribution to Doctors Without Borders, conditional on my jumping out the window to my death. Knowing that my self-sacrifice would

save thousands of lives, I must pause to seriously consider the options. As I consider jumping out the window, the vivid imaginative experiences of falling to my death increase the violence of my aversion to dying. This parallels the way that the feeling of obligation goes for people who are wavering between fulfilling their obligations and not fulfilling them and for whom the possibility of defaulting on their obligations must be seriously considered. In cases where one cannot be confident in one's ability to satisfy one's obligations, one seriously imagines defaulting, and the aversion that underlies the feeling of obligation can become violent.

If some version of the "ought implies can" principle is true, this may also help responsible agents maintain their confidence in their ability to do as they ought. While circumstances beyond their control can tear the objects of agents' other desires away from them, the truth of "ought implies can" would prevent circumstances beyond an agent's control from bringing about the situation that these responsible agents are averse to—the situation in which they fail to do as they ought. If, due to circumstances beyond their control, it becomes impossible for them to do something that they otherwise ought to have done, it will no longer be the case that they ought to do it. The situation that they are averse to—the situation where they fail to do as they ought—will not have come about since it is no longer the case that they ought to do the action. The only way they can end up in the situation to which they are averse is if they have the satisfaction of their obligations within their power and still fail. Then "can" will obtain, and "ought" will too. But if they know themselves to be responsible agents, this situation will seem unlikely and remote, and it will not trouble them.

The last consideration that I want to bring up in explaining the feeling of obligation has to do with the way that it feels to pass up the object of one of your desires in order to satisfy a stronger desire that has been involved in the formation of a prior intention. Suppose I have paid a lot of money for a plane ticket to go visit some friends in another state, and my plane leaves on Thursday. If I subsequently learn of an exciting party on Friday, I will not seriously consider missing the flight to go to the party. I will feel that the party is something I am unable to attend, even though I want to. I will be disappointed about missing the party, but I will have no experience of weighing the options. Rather, I will feel as though the situation constrains me, preventing me from getting something good that remains beyond my grasp.

While the case of missing the party to catch my flight is not a case of obligation, it feels the same way, in one important respect, as motivation from the feeling of obligation does. The feeling of being constrained and

unable to get something you want is not unique to cases where the feeling of obligation is present. Rather, it appears in many different cases of desire. In this way, my experience as I am disappointed at missing the party because I have to catch my flight is like that of the Unwavering Grader as he misses his party because he has to finish his grading.

I will now return to this case. In making his decision, the Unwavering Grader was faced with two choices that felt different to him. The emotions connected with grading were less intense than the emotions connected with the party. The cause of the difference in the intensity of these emotions is the difference in the violence of the passions that drive him toward these two things. And the cause of this difference in violence of passion is the difference in the vividness of the images that pass through his mind as he deliberates about the options. Since the concept of obligation does not lend itself to vivid imagining and since the possibility of failure will seem remote if he knows himself to be reliable in fulfilling his obligations, he will not have particularly vivid mental representations of failure to fulfill his obligations. And since he has a prior commitment to grading, backed up by powerful desires, he will not unlock this commitment to weigh going to the party against grading the papers.

I regard the considerations I have laid out as presenting a good reductive explanation of the feeling of obligation, in two of its common forms. The feelings associated with obligation have been reduced to the feelings associated with ordinary desires. The availability of this reductive explanation, which accounts for all the phenomenological data using a simple ontology of motivational states, gives us some reason to accept the Humean theory of motivation and to reject claims that some motivational force other than desire operates on us when we experience the feeling of obligation. Instead of being a problematic case for the Humean theory, the case of obligation shows that Humeans can deliver detailed and illuminating explanations of the phenomenology of decision making.

Darwall and Desires Formed through Deliberation

In *Impartial Reason*, Stephen Darwall presents a vividly illustrated case in which he claims that an agent forms a new desire through reasoning that does not have another desire as a premise. Such a case would be a counterexample to DODI and to the Humean theory. But as I will argue, the Humean theory can offer us as good an explanation of all the features of the case as Darwall's view can, while relying on a simpler ontology of motivational processes. Rather than being a counterexample to the Humean

theory, Darwall's case demonstrates the superiority of the Humean theory over competing explanations.

One of Darwall's targets in *Impartial Reason* is the "DBR Thesis" ("DBR" stands for "Desire-Based Reasons"). According to this thesis, all of an agent's reasons "'have their source in' that agent's desires" (*Impartial Reason*, 27, quoting Gilbert Harman). Darwall says that previous discussions of this thesis have left it unclear what it means for reasons to have a source in the agent's desires, though he offers several clarifications of what this could mean. Version III of the DBR Thesis runs as follows:

> **[III]** Something is a reason to act only if it evidences the act to promote something the agent desires.

"Reason to act" here has a motivational reading, on which a reason to act is something that explains action (as opposed to something that justifies action). So III is implied by the Humean theory, and counterexamples to III will be counterexamples to the Humean theory of motivation.

Darwall objects to III and to the Humean theory. He opposes views on which "the agent's *current* desires function as a filter that determine which considerations can move him and which cannot." He supports a view on which someone can be "moved by awareness of some consideration, without that being explained by a prior desire" (*Impartial Reason*, 39). While he suggests that a new desire may be attributable to the agent after deliberation has concluded, he does not think that an antecedent desire is necessary for deliberative processes to go forward.[15]

Darwall offers an example in which new desires are formed through this kind of deliberation:

> Roberta grows up comfortably in a small town. The newspapers she reads, what she sees on television, what she learns in school, and what she hears in conversation with family and friends present her with a congenial view of the world and her place in it. She is aware in a vague way that there is poverty and suffering somewhere, but sees no relation between it and her own life. On going to a university she sees a film that vividly presents the plight of textile workers in the southern United States: the high incidence of brown lung, low wages, and long history of employers undermining attempts of workers to organize a union, both violently and through other extralegal means. Roberta is shocked and dismayed by the suffering she sees. After

15. Similar views are held by Thomas Nagel in *The Possibility of Altruism* (Princeton: Princeton University Press, 1978), and John McDowell in "Are Moral Requirements Hypothetical Imperatives?" I focus on Darwall because he tries to offer a vivid example of non-instrumental desire-formation through reasoning.

> the film there is a discussion of what the students might do to help alleviate the situation. It is suggested that they might actively work in promoting a boycott of the goods of one company that has been particularly flagrant in its illegal attempts to destroy the union. She decides to donate a few hours a week to distributing leaflets at local stores. (ibid., 39)

This is a richly illustrated example. Darwall has fleshed it out in detail so as to make Roberta's story and her mental life fit well with our folk understanding of how people think and feel. I quote it at length because many of its parts will be useful in developing the Humean response to Darwall's objection.

According to Darwall, Roberta's decision to join the boycott does not require explanation by the presence of a preexisting desire to relieve suffering. She simply has achieved a vivid awareness of the unfortunate situation of the textile workers, and this awareness will motivate her to act. Darwall allows that awareness of the workers' situation may cause her to form a new desire which she then acts on, but even after allowing this, he comes out as an opponent of the Humean theory as I have framed it. He says of Roberta, "whatever desire she does have after the film seems itself to be the result of her becoming aware, in a particularly vivid way, of considerations that motivate her desire and that she takes as reasons for her decision: the unjustifiable suffering of the workers" (ibid., 40). Since he sees Roberta's process of desire-formation as driven by her accepting particular considerations as reasons and denies that desires stand at the beginning of her reasoning, he must claim that new desires can be generated by processes of reasoning that do not have desires as premises. Then he will be denying DODI and opposing the Humean theory.

But even as Darwall describes the example, there is some reason to think that Roberta came to the film with a desire that people not suffer and that her desire to help the workers was formed through the instrumental processes accepted by Humeans. Consider the shock and dismay that Darwall describes her feeling when she watches the film. Desire, as I have claimed, has a Hedonic Aspect—when people are presented with vivid images of states of affairs that they are averse to or when their subjective probability of desire-satisfaction decreases, they feel displeasure. The Hedonic Aspect of desire is manifested in both of these ways as Roberta watches the film. Darwall describes the vivid way that the film presents the suffering of the textile workers. Given how comfortable Roberta's previous upbringing was and how sheltered she was from the suffering in the world, she is likely to have had an unrealistic view of how happy other

people were. Someone who desires that others not suffer will feel shock and dismay upon discovering that there is more suffering in the world than she thought. Roberta's emotions are evidence of an antecedent desire.

Darwall does not address the way that Roberta's shock and dismay serve as evidence for the presence of an antecedent desire. These emotional responses are typical among people who have suddenly realized that some undesirable situation obtains and who are about to instrumentally form a new desire to address it. What cause of shock and dismay, other than a preexisting desire combined with a sudden realization that a deeply undesirable situation obtains, will so simply explain the phenomena? The Hedonic Aspect of desire neatly accounts for the fact that we are subject to these emotions.

I shall consider the response that a belief that others are being wronged (or some similar belief) could generate both the sentiments and motivation in this case and that we can just as simply explain Roberta's case in these terms. If we were to look at her case without thinking about the ontology of motivational states that Humeans and anti-Humeans are committed to, this view might indeed seem as good as the Humean view. But when we consider the prior commitments of both views, we can see how the Humean explanation of Roberta's case allows us to develop a simpler overall picture. While Humeans can claim to be extending a simple model of action that covers all cases, no similar claim can be plausibly made on the anti-Humeans' behalf.

Consider the cases of hunger, thirst, and sexual lust. Anti-Humeans generally concede that the states grounding motivation in these cases are best understood as desires. It is hard to see how they could do otherwise, considering the ways in which these states are produced and the inability of reasoning to create or eliminate them. So both Humeans and anti-Humeans will be committed to motivation grounded in desire. Against this background, an explanation of Roberta's deliberation and action in which her motivation is grounded in a preexisting desire will be simpler than one that also includes motivational processes not attested elsewhere. And as we come to Roberta's case (and the other cases discussed in this essay), we have no similarly uncontroversial cases of the motivational processes that anti-Humeans invoke. Cases like Darwall's and the others discussed in this essay are supposed to do the work for the anti-Humean that hunger, thirst, and lust do for the Humean. They are supposed to show that we must admit processes of reasoning that violate DODI. But if Humeans can fully explain these cases without bringing in such processes, the Humean theory will have

demonstrated its greater simplicity in accounting for the same data, and thus its superiority.

One might ask why we should be so quick to invoke simplicity as a criterion for choosing between rival explanations of mental phenomena. The mind, after all, is a complex place, and we should not deprive our theories of the resources they need to explain the rich data presented by our behavior and our inner lives.[16] But there is no reason for our theories to be more complex than the data requires. Simplicity prevents us from populating the mind with redundant mental entities and accepting needlessly baroque explanations of people's behavior. Abandoning simplicity would license the attribution of all sorts of outlandish mental states that contribute nothing to explanatory power. We can avoid the pitfalls of both oversimplification and undersimplification if we make sure that our theories account for all the relevant motivational and phenomenal data and then choose the simplest theory that satisfies those constraints. To argue against the Humean theory, anti-Humeans should either find some data that it lacks the resources to explain or construct an alternative account that equals or exceeds it in simplicity.

Darwall has some arguments against the view that Roberta formed her new desire in a way consistent with DODI, starting from an antecedent desire for others not to suffer. His first argument against the view that Roberta had an antecedent desire begins by imagining a way for her to instrumentally form a new desire out of an antecedent desire to avoid suffering. He then argues that Roberta need not have formed her desire in this way and that it is more plausible to say that Roberta could form her new desire otherwise. He describes the way that she would generate her new desire instrumentally as follows: "She had some such general desire as the desire to relieve suffering prior to seeing the film, saw this as an opportunity, and formed the desire to relieve this suffering, as part of an Aristotelian practical syllogism" (ibid., 40). As Darwall says, "This need not be what happened."

Upon reading Darwall's example, one certainly does not imagine Roberta seeing an "opportunity" in the plight of the textile workers to satisfy a previously held desire. If the Humean theory claimed that she regarded the workers' misfortunes as an *opportunity* to engage in suffering-relieving activity, with the positive attitude that connotes, that would be a serious strike against it. But that is not how the Humean theory that I have constructed would treat it. Typically, people acting to relieve

16. I thank two anonymous reviewers at the *Philosophical Review* for raising this issue.

suffering desire that others not suffer. They will be satisfied whether or not the suffering is relieved by their own actions, as long as it is relieved. The desire to relieve suffering through one's actions—a desire that would cause one to see the suffering of others as an opportunity, much as someone with a desire to eat pizza sees the presence of a pizza as an opportunity—is usually not such a large motivational force.[17] We can see this in how people concerned about the suffering of others are generally quite satisfied to see some third party intercede and alleviate the suffering. Roberta's case seems like a normal one in this regard, and this has implications for how she would feel on discovering that other people were suffering. The new information that people are suffering reduces the subjective probability of desire-satisfaction rather than increasing it, and this produces unpleasant emotions like shock and dismay rather than excitement at the presence of an opportunity. She is averse to the suffering of others and that is why she is unhappy at the sight of their suffering rather than pleased by an opportunity to engage in an action that she desires to perform.

Darwall criticizes the Humean interpretation of Roberta's desire-formation in another way as well. He says that a desire "includes dispositions to think about its object, to inquire into whether there are conditions that enable its realization" (ibid., 40). Darwall is right to say that desire can make agents think about its object—this is part of what I have called the Attention-Direction Aspect of desire. The "inquiring" that Darwall talks about can be reduced to a combination of attention-direction toward things we associate with the objects of desire and interested thoughts about how to attain these objects if our attention happens to settle on a means to our end. But here Darwall sees the opportunity to develop an objection to the Humean theory: if this sort of thought and inquiry is a necessary condition for desire, why is thought and inquiry about how to relieve others' suffering so absent from Roberta's mental life before seeing the film?

Here it is important to look at the conditions of Roberta's upbringing. Her environment, Darwall says, offers her "a congenial view of the world and her place in it" (ibid., 39). Stimuli that would activate a latent preexisting desire that others not suffer, then, are largely absent from

17. Perhaps Darwall merely meant a desire that others not suffer or, more specifically, an aversion to others' suffering, when talking about a "desire to relieve suffering." And perhaps he did not intend "opportunity" to have the positive connotations that it does. Since the point is worth making, I will go with his actual use of the words and apologize for possibly being uncharitable.

Roberta's early environment. Furthermore, it does not seem that she is presented with any vivid images of suffering or any reliable plan for how she could act to avert it—"She is aware in a vague way that there is poverty and suffering somewhere, but sees no relation between it and her life" (ibid., 39). In the absence of these factors, nothing brings her desire that others not suffer to the forefront of her mind.

A parallel to Roberta's situation may be useful here. Like most people, I have a strong desire that my mother not come to harm. But this desire does not usually motivate me to inquire into means for promoting its realization, and most days pass without my thinking about whether my mother has come to harm or what I could do to prevent anything bad from happening to her. I know that she lives in a safe place, is healthy, and does not take unnecessary risks, so I believe that the likelihood of her coming to harm is quite low. Furthermore, I am not usually presented with vivid images of my mother being harmed. In this, I am like Roberta before she saw the documentary. While my desire (in this case, my aversion to my mother being harmed) is strong, it remains latent because there is nothing to activate it. If something were to change—if I were to learn that my mother was in danger or even if I woke from a bad dream in which she came to harm—my desire would be activated, and it would drive my thoughts.

Part of the point of Darwall's example is that "a person's motivational capacities, in the broadest sense, are not constituted simply by his desires but also by capacities of imagination, sensitivity, and so on" (ibid., 39). I have argued above that Roberta's sensitivity to suffering is best understood as being at least partially constituted by a preexisting desire that others not suffer. And I accept that imaginative capacities and other things beyond desire play a role in determining how people are motivated. Desires are temporarily strengthened by vivid sensory or imaginative representations of their objects, and both belief and desire are necessary for motivation.

But while imagination plays a role in motivation, it is like belief in that its ability to motivate action is dependent on an agent's preexisting desires. An agent with different desires could be motivated in exactly the opposite way by the same set of sensory and imaginative experiences. Consider a man—we might call him Pinkerton—who lacks Roberta's desire that others not suffer and has a little bit of sadism in him as well. His dislike of working-class people manifests itself in a strong aversion to their advancement and in a desire to see the humiliation and defeat of those who stand up against the prevailing economic order. Watching the movie and imagining the workers' situation, he might despise them and come

to support the brutal and repressive tactics of management. Rather than promoting the boycott, he might inquire after summer employment in the South as one of management's anti-union goons. This need not be because of any failure to appreciate the situation of the textile workers—he may see it in his mind just as vividly, and understand the descriptive features of the situation just as well as Roberta does. But the things he perceives will motivate him in a radically different way than they motivate Roberta since his desires are dramatically different from hers.[18]

Scanlon and the Structure of Deliberation

In the first chapter of *What We Owe to Each Other*, Thomas Scanlon offers a new account of desire and two criticisms of the Humean theory—first, that it cannot explain cases in which people act despite "having no desire" to do something, and second, that it cannot adequately explain the structure of deliberation in cases in which agents "bracket" some of their options. My responses will be familiar from the discussion of obligation earlier in this essay. Agents who act despite reporting "no desire" to do something are merely reporting a lack of positive desire and are motivated by aversion. The case of bracketing involves some of the same features as the case of the Unwavering Grader, with the additional point that deliberation is structured by an antecedent higher-order desire. In both of these cases, the pleasure and displeasure that agents feel in the course of deliberation are best explained by the Humean theory.

Scanlon offers an account of "desire in the directed-attention sense. A person has a desire in the directed-attention sense that P if the thought of P keeps occurring to him in a favorable light, that is to say, if the person's attention is directed insistently toward considerations that present themselves as counting in favor of P" (*What We Owe to Each Other*, 39). This account has several things in common with mine. Most obviously, Scanlon regards desire as capable of directing an agent's attention, while I posit an Attention-Direction Aspect of desire. One might also regard the Hedonic Aspect of desire as a way of cashing out how things an agent desires appear to him or her in a "favorable light." Scanlon, of course,

18. I use the example of Pinkerton to argue against a particular anti-Humean view on which motivation can be generated by processes of reasoning that begin with no desires but with the imagination. Anti-Humeans who wish to join me in arguing that imagination cannot play such a role in motivation are entitled to replace desire in the above example with whatever mental states they regard as motivationally efficacious.

cashes out "favorable light" in different terms. But since I hold that desire can focus an agent's attention on things strongly associated with the object of desire, the particular things that count in favor of it will certainly be among the things that attract the agent's attention.

Scanlon and I differ in that he does not regard desires as the motivational forces driving all actions. While he accepts that some urge to act is involved in all cases of motivation, he does not think that desire in the directed-attention sense is always involved. On his view, "it is not the case that whenever a person is moved to act he or she has a desire in this sense: we often do things that we 'have no desire to do' in the ordinary sense, and 'desire in the directed-attention sense' tracks the ordinary notion in this respect" (ibid., 40). In making this claim, Scanlon goes further than Darwall, Nagel, and McDowell, who accepted DBTA while denying DODI. Of course, there is plenty of substantive agreement between all of these opponents of the Humean theory. The difference is only that Scanlon's stronger notion of desire prevents him from seeing a desire in every case of motivation and causes him to deny DBTA.

On Scanlon's own view, motivation is explained not by desire-belief pairs but by the fact that an agent takes particular things to constitute reasons for action. Even in cases when an agent has a desire and acts accordingly, "what supplies the motive for this action is the agent's perception of some consideration as a reason, not some additional element of 'desire' " (ibid., 40–41).

Scanlon offers a case in which someone acts despite having "no desire to do" something—a case where "one must tell a friend some unwelcome news" (ibid., 39). In this case, he says, the characteristic features of desire in the directed-attention sense are missing. It is not hard to see the phenomena that he is pointing to. When one has to tell a friend some bad news, the thought of doing so does not keep occurring to one in a favorable light. One is displeased at the prospect of having to bear the bad news, and one's attention focuses more on how upset the friend will be rather than on the good things about his knowing the truth.

To deal with this case, we need to note that the predominant motivational factors in the case are not positive desires but aversions. While positive desires direct an agent's attention toward things associated with what he wants, aversions direct an agent's attention toward things associated with the object of aversion—in other words, what he does not want. The motivational forces in play in Scanlon's case are most likely a pair of conflicting aversions. The agent may have an aversion to his friend's being in the dark about the bad news (or perhaps an aversion to bad consequences

befalling the friend because he acts without knowledge of the bad news) and also an aversion to the friend's being unhappy. After the agent decides that he is going tell his friend, he focuses on the unpleasant duty before him and the unpleasant features of what he has to do loom large in his mind.

This is generally what it is like when we are faced with two options that we are averse to, and we have to choose the lesser evil. Acting in these cases is unpleasant, as things we associate with the object of our aversions are often close at hand when we act and inflame the violence of the aversion even as they cause displeasure by giving us vivid representations of situations to which we are averse. To overcome the unpleasantness that we feel in these cases, we sometimes choose to focus our attention on our freedom from the even worse consequences that we have chosen to avoid and draw some relief from that. Often, this does not occur in the automatic way that desire causes attention to focus on things associated with its objects but as an intentional decision of the agent to look on the bright side.

If this is a sort of case that the Humean theory gets right, why do people often say in these cases that they have no desire to do the thing in question or that they are doing what they do not want to do? The answer lies in the fact that 'desire' and 'want' are often used to refer only to what I have been calling "positive desire." I have departed from this use of 'desire' in including aversions among the category of desires. Positive desire and aversion have enough in common that it makes sense to bring them both under one term for the purposes of constructing a theory that explains action. They have many similar psychological effects, from their ability to motivate action, to their connections to pleasure and displeasure, to the fact that they can be intensified by vivid images that are associated with their objects. There are, however, slight differences in the emotions associated with them and in the particular way that they direct our attention. Positive desires direct our attention more toward the states of affairs that we act to obtain, while aversions direct our attention more toward the states of affairs that we act to avoid.

While the Hedonic Aspect of desire allows the Humean theory to neatly explain the unpleasantness of telling a friend some unwelcome news in terms of the vivid representations of a friend's unhappiness, it is hard to see how Scanlon's theory can do so in a similarly economical way. Why would seeing a reason to act and acting on it be unpleasant? Scanlon could offer a separate explanation of the agent's displeasure, but then the Humean view would be superior on grounds of theoretical unity,

as it invokes one mental state—desire—that does both the motivational and hedonic work. Cases like this, then, are better explained by the Humean theory than by anti-Humean views like Scanlon's.

Now I will turn to a second criticism Scanlon makes of Humean views. Noting that desires are normally understood as having particular weights and focusing on particular objects, he says that the Humean view casts rational decision making as "a matter of balancing the strengths of competing desires. If we take desires, along with beliefs, as the basic element of practical thinking, then this idea of balancing competing desires will seem to be the general form of decision-making" (ibid., 50). Scanlon later says that "reasons for belief do not have the simple structure that the desire model of practical reasoning describes: they do not simply count *for* a certain belief with a certain weight, and deciding what to believe is not in general simply a matter of balancing such weights" and that "reasons for action, intention, and other attitudes exhibit a similarly complex structure" (ibid., 52). This adds up to an objection to the Humean view. Scanlon's objection assumes that Humeans are committed to regarding the weighing of desires against one another as the only way that competing motivational forces interact in practical deliberation. He then argues that competing motivational forces need not interact in this way.

Scanlon describes a kind of decision in which more complex structures than the weighing of competing desires are involved. Many decisions, he says, "involve bracketing the reason-giving force of some of your own interests which might otherwise be quite relevant and legitimate reasons for acting in one way rather than another" (ibid., 52). His example involves the chair of a philosophy department who has strong personal interests at stake in some decision he is making. The chair may put those interests aside in his deliberation and make his decision based on what is good for the department. He does not weigh his personal interests against the interests of the department every time he makes a decision. A model that attributed this kind of weighing to him might predict his behavior successfully, but it would not accurately represent the phenomenology. He does not have the experience of weighing but rather the experience of working toward a goal while he passes by attractive considerations toward which he will not turn. He will notice when he is making a decision that contradicts his personal interests, and he will probably feel chagrined about this. But he is committed to making his decision in the best interests of the department, and he might never seriously think about

whether to make decisions based on his personal interests and against the department's interests.

How might the Humean theory explain how the chair makes his decision? We might begin with the observation that the chair feels obligated to make official decisions in the best interests of his department. In keeping with the account of the feeling of obligation presented previously, this suggests that the chair's motivational states not only include a desire aimed at a personal interest and a desire aimed at a departmental interest but also an aversion to letting personal interests determine official decisions. This aversion aims at the way his own motivational processes are to operate, and thus is a second-order volition.

Previously I considered two cases in which the feeling of obligation is different—the case of the Tempted Grader and the case of the Unwavering Grader. The former case relies on differences between the emotions resulting from desire and aversion to explain why different emotions are associated with the objects of the agent's desires. The latter case relies mainly on the greater vividness of the images associated with the objects of the weaker desire to explain why it still generates stronger emotions than the other.

Scanlon's example seems to have more in common with the case of the Unwavering Grader. Excluding considerations of personal interest is too abstract to lend itself to particularly vivid imagining, unlike (for example) hiring a dear if slightly underqualified friend. If my account of the phenomenology of obligation in the case of the Unwavering Grader is satisfactory, it should go a long way in explaining the motivational processes at work here too. When we consider how strong aversions operate in the absence of vivid images of what is desired, we can understand why the chair feels the negative emotions connected to not hiring his friend more strongly than any positive emotions from the less vivid exclusion of personal interests. And as I will explain, we will still be able to explain why he does not weigh his personal interest against the department's interests or against making the decision unjustly.

One interesting feature of Scanlon's case that builds on the discussion of obligation is the role that the chair's knowledge of his own motivational structure plays in explaining his experiences. People who know themselves to be reliable in fulfilling their obligations often make decisions without weighing the benefits of violating their obligations because they know that their psychology makes these benefits inaccessible, and benefits known to be inaccessible are not weighed in deliberation. The chair will not weigh acting on his personal interest because he knows

that his motivational structure will prevent him from doing so. Conscientious people often say of certain immoral or inappropriate actions that they "couldn't do something like that." It is not that they are physically incapable of performing these actions—they just know that their desires about how their decisions should go make these actions impossible for them. In this sense, deciding on the basis of personal considerations is something that the chair just cannot do. While things we desire sometimes look impossible to us because we know that physical barriers prevent us from attaining them, we may also know that our motivational architecture will not allow us to pursue them.

The chair's strong aversion to acting on his personal interests and his knowledge that he has this strong aversion allow him to bracket some of the considerations when he makes his decision. Desiring to act only in the department's best interests, he initially focuses on the aspects of the decision that relate to the department's interests. But at some point during the decision-making process, he notices that some other thing that he personally desires is at stake. This interest may attract his attention, as the things we desire often do. But he will still regard its object as being unavailable to him. He knows that his motivational structure will prevent him from pursuing it. So it will not even appear to him as a thing to be weighed in deliberation. What we know we cannot attain, we do not weigh.

Scanlon anticipates a Humean response that is like mine. According to this response, the agent's second-order desires are responsible for bracketing. On my explanation, bracketing is driven by a second-order volition—a desire that a particular desire or set of desires be (or not be) effective in moving him to action. Scanlon objects that second-order desires lack the authority to structure deliberation:

> But if second-order desires are really desires, then there is the question of how their second-order character, if it is just a difference in the objects of these desires, can give them the kind of authority that is involved when one reason supports the judgment that another putative reason is in fact irrelevant. My desire to be a person who does not let considerations of personal interest influence his decisions as department chair conflicts in the practical sense with my desire, in this case, to do what will make my life easier. I cannot act in a way that will satisfy both of these desires at once. But they are just two desires that conflict with each other. The introduction of second-order desires therefore does not do justice to our sense that there is a deeper conflict, expressed in the judgment that the reason represented by the latter desire is not relevant. (ibid., 55)

Scanlon's point here is familiar from Gary Watson's response to Harry Frankfurt's account of free will.[19] Why should we suppose that a desire's being higher order gives it any sort of authority over lower-order desires?

It is not clear, however, why we need to invoke authority to explain how the chair deliberates. On a Humean view, higher-order volitions generate the phenomenon of bracketing because of their content, their power, our knowledge of how they will affect deliberation, and how we regard inaccessible things. We need not attribute any special authority to them to explain this.

Here I do not deny that the chair regards the exclusion of considerations of personal interest as authoritative. He surely does, just as graders who have promised to return papers to their students regard those promises as authoritative. Neither do I say that these judgments of authority are false if they are understood purely as normative claims. My account of the feeling of obligation is neutral about what our genuine obligations are, and my account of bracketing is similarly neutral on the question of which judgments have real authority. The point is just that the motivational processes implicated in Scanlon's case can be explained perfectly well without invoking the actual authority of some motivational states and without giving judgments of authority the sort of motivational role that would violate the Humean theory.

Perhaps Scanlon would want to buttress his point by putting the issue of bracketing in terms of which considerations the agent is permitted to weigh in his decision making. The language of permission, certainly, suggests that authority is involved in granting the permission. And the displeasure we feel with ourselves when we seriously consider acting on a bracketed-off consideration—for example, the way that the chair would feel if he imagined acting selfishly—might be taken as a sign of the acknowledged impermissibility of acting in this way.

But the Humean account that I have presented can explain all of this in terms that do not involve authority. If we are forced to consider being the kinds of people who act on bracketed-off considerations, we are thinking about acting so that the antecedent desire that structured our deliberation is not satisfied. (Cases where we think in such a way resemble that of the Tempted Grader more than the Unwavering Grader.)

19. Gary Watson, "Free Action and Free Will," *Mind* 96 (1987): 145–72; Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5–20.

These thoughts are often unpleasant, as the Hedonic Aspect of desire would suggest. And if the antecedent desire is a second-order volition whose object involves our own deliberative processes, our displeasure will be displeasure with ourselves. Scanlon's account leaves it unclear why the acceptance of reasons that drives bracketing would have anything like a Hedonic Aspect, and thus contribute to this displeasure. Scanlon might go the way that I have suggested above—where displeasure results from being aware that one acted impermissibly or from imagining such action. It is important to note that we feel displeasure when we become aware that a strong desire of ours will not be satisfied or when we imagine such a situation. Some of these situations will not be ones in which we become aware that we have acted impermissibly. The Humean theory only needs to invoke one process to explain both the cases where permissibility is at stake and the ones where it is not. Scanlon, meanwhile, must invoke two separate processes—one to explain the bracketing and another to explain the displeasure.

In the end, this is a case that shows us the simplicity and explanatory power of the Humean theory. The ways bracketed-off options look to us—both in their seeming inaccessibility and in the displeasure we feel when we are forced to imagine being the unsavory characters who could pursue them—are continuous with the way that our desires shape our thoughts in basic cases. Bracketed-off options look like things we desire that we know we cannot get. Thoughts of being bad enough to pursue them displease us, as thoughts of situations we are averse to generally do. These continuities justify us in accepting an elegant explanation of bracketing that incorporates no motivational states other than desire. They should raise our hopes for a similarly elegant explanation of all deliberative phenomena.

Searle and *Akrasia*

In *Rationality in Action*, John Searle argues that the Humean theory and all theories on which an agent's psychological states are sufficient to explain action will have difficulty in explaining how an agent can be susceptible to *akrasia*. In cases of *akrasia*, an agent's judgment about what to do differs from how he acts, even at the moment of action. I will argue that Searle's account of *akrasia* is unsuccessful and that a Humean account invoking the violence of desires stimulated by vivid sensory or imaginative representations will provide a more successful explanation.

Searle focuses on criticizing Davidson's account of weakness of will, which he regards as part of "a long tradition in philosophy according to

which in the case of rational action, if the psychological antecedents of the act are all in order, that is, if they are the right kind of desires, intentions, value judgments, etc., then the act must necessarily follow. According to some authors it is even an analytic truth that the act will follow" (*Rationality in Action*, 220). The problem Searle finds with these views is that, in tying judgment and action too tightly to their psychological antecedents, they make it impossible to see how judgment and action can come apart. On Searle's own view, the mental states that lead an agent to form an intention (which he takes to be the mental state of judgment in a case of *akrasia*) are not causally sufficient for rational action. There is "a gap, a certain amount of slack between the process of deliberation and the formation of an intention, and there is another gap between the intention and the actual undertaking" (ibid., 231). These gaps are the places where the agent's free will comes in and determines what the agent will intend or what the agent will do.

Searle offers a description of "one way in which *akrasia* typically arises":

> As a result of deliberation we form an intention. But since at all times we have an indefinite range of choices available to us, when the moment comes to act on the intention several of the other choices may be attractive, or motivated on other grounds. For many of the actions that we do for a reason, there are reasons for not doing that action but doing something else instead. Sometimes we act on those reasons and not on our original intention. The solution to the problem of akrasia is as simple as that: we almost never have just one choice open to us. Regardless of a particular resolve, other options continue to be attractive. (ibid., 233–34)

Searle's solution to the problem of *akrasia* is to posit a gap between intention and action in which the agent's free will determines whether he acts. According to Searle, an antecedent psychological state of intending is not sufficient to determine whether the agent will act since agents sometimes act against their intentions as an act of free will. We have an experience of the gap in forming intentions and an experience of the gap in determining whether to act on our intentions. These experiences of the gap mark points at which our free will is active—first in the formation of our intentions and then in our decisions to act on our intentions. It is the latter gap that allows us to contradict our intentions in akratic action.

Searle's account, however, fails to explain what is truly interesting about *akrasia*. The problem of *akrasia* is not merely that we sometimes fail to act on our original intentions because other choices look attractive

to us. The problem is about the unusual psychological processes that are implicated in this failure. We hold fast to our original judgments about what sort of action to perform, affirming them even as we do something else. What needs to be explained in explaining *akrasia* is that our judgments about what to do—which normally run in the same direction as our actions—are somehow overridden without being revised. Searle's view fails to explain why *akrasia* differs at all from normal cases in which an agent changes his mind at the last moment and wholeheartedly decides to do something that he did not plan to do before.

The description of the heroin addict who compulsively takes the drug, which Searle presents early in his book, suggests an alternative way in which he could deal with *akrasia*. Searle regards the case of the heroin addict as an unusual one in which the addict's psychological states are causally sufficient for the performance of the action. If Searle had made his account of *akrasia* generally look like this, with the motivational force of the agent's psychological states overwhelming the force of the agent's will, he would have an explanation of why *akrasia* is different from changing your mind at the last moment. The psychological states would control the agent's behavior, while free will would control the agent's judgment. If the phenomenology of will-driven action was distinguished from the phenomenology of having one's action determined by antecedent psychological states, the experience of akratic action could be explained. There would still be a question of why the akratic agent's will failed in this case while it succeeds in others, and the account would be less simple than Humean views are because it invoked the additional motivational force of free volition, but the view would address the problem. The actual position of Searle's seventh chapter, by contrast, does not even address the real problem of *akrasia*.

Having shown that Searle's proposed solution will not deal with *akrasia*, I will now offer a Humean solution, relying on the details of how vivid images of desired things increase the violence of our desires. I hope to show that the issue of *akrasia* is not a weakness for the Humean theory but a strength. Once we understand how desire interacts with other mental states that produce vivid representations, we will be able to see why we act akratically in the cases where we do and why weakness of will feels the way it does.

I will begin with two ordinary cases of weakness of will. First, a case of *akrasia* at bedtime. I am watching television and I realize that it is 2 a.m. I am tired, and I know that I really should go to bed. Tomorrow morning the Formal Epistemology Workshop begins, and I would like to attend as

much of it as possible so I can learn something about formal epistemology. But the witty dialogue of the *Buffy* rerun and the winsome smile of the red-headed supporting actress have their grip, and even as I tell myself that I really should go to sleep, I stay where I am and keep watching television for another hour.

Akrasia strikes again at 8 in the morning, after my alarm clock wakes me up. Thinking of the workshop, I realize that I should get out of bed and go there. But my bed is warm and soft, and I am still tired, as the previous night's weak-willed television watching prevented me from getting enough sleep to feel fully refreshed. So I lie there comfortably, knowing that I will end up missing the opening session as a result.

Both of these cases—and, I think, all cases of *akrasia*—have some common features. The agent is torn between two different desires, and his environment is such that he has vivid sensory or imaginative representations that relate to the object of one desire. At the same time, he believes that the object of the other desire is in jeopardy but does not have similarly vivid sensory or imaginative representations relating to it. The vivid sensory and imaginative representations increase the violence of the passion whose object they represent. This gives that passion more motivational force and causes significantly more violent emotions. But it does not do quite as much for the violent passion's ability to control the way that the agent directs his attention to various possibilities as he makes his judgment about what he ought to do. Though his calm passion is too weak to overpower the violent passion and determine his behavior, it controls his reflective judgment by directing his attention toward the states of affairs that would satisfy it. As I discussed earlier, vivid images are especially powerful in increasing the motivational and emotional force of a passion, but they do not give an equal boost to all of the passion's effects. The agent's reflection and judgment are controlled by one desire, while his behavior is controlled by the other, and this is why reflective judgment and behavior come apart in cases of weakness of will.

Searle was wrong to treat action against a prior intention as the whole of *akrasia*, but it is an interesting fact about akratic action that it often involves acting in a way that contradicts one's prior intentions. My account explains why this is so often the case. Away from the TV or the comforts of my bed, I am not faced by the vivid images that would activate and excite the desires that eventually drive my akratic actions. So in the calm hours of the afternoon, when I plan my evenings and my mornings, my desire to watch more TV and my desire to stay in bed operate at a lower strength than they would if I were presented with vivid images of television

or the feeling of my bed. As a result, I do not make prior plans to watch TV late at night or stay long in bed. But when actually presented with these sensory experiences, the desires that drive my actions become more violent and control my behavior even as my judgment favors another course of action.

This account also explains why cases of *akrasia* often involve agents acting to attain sensory pleasures, even as they judge against doing so. It is much rarer for agents to akratically pass up a sensory or physical pleasure in favor of a more abstract or remote satisfaction. If the object of some desire is itself a sensory experience (as the experience of a warm and comfortable bed is, especially on a chilly morning), it will register vividly in sensation and imagination, increasing the violence of the desires that are directed toward it. This will make that desire more capable of driving akratic behavior.

The case of *akrasia* offers a striking display of the Humean theory's explanatory virtues. One might have worried that the clear phenomenological difference between the motivational force that controls judgment and the motivational force that determines behavior would force us to accept a less simple theory of motivation. But as we have seen, the Humean theory can account for *akrasia* by appealing to the familiar way that vivid sensory or imaginative representations differentially amplify the effects of our desires. We need not posit a separate motivational faculty of reason, beliefs that can generate desires, or free will. We are in the fortunate position of having a simple theory that does all the explanatory work.[20]

20. For many helpful comments, I thank the members of my dissertation committee at the University of Texas: Brian Leiter, David Sosa, Josh Dever, Jonathan Dancy, and John Deigh. I thank Tim Scanlon and Stephen Darwall for generously helping a young Humean better understand their views. Peter Railton, Allan Gibbard, Nadeem Hussain, Tim Schroeder, and Howard Nye also provided useful feedback. Excellent comments from anonymous referees at the *Philosophical Review* substantially improved the paper. Finally, thanks are due to two of my undergraduate teachers—Raphael Woolf, who guided the early development of these ideas, and Melissa Barry, who first told me to keep pursuing them.