The Dissertation Committee for Neiladri Sinhababu certifies that this is the approved version of the following dissertation:

# A Treatise of Humean Nature

Committee:

______________________
Brian Leiter, Co-Supervisor

______________________
David Sosa, Co-Supervisor

______________________
Jonathan Dancy

______________________
John Deigh

______________________
Josh Dever

______________________
Allan Gibbard

# A Treatise of Humean Nature

by

**Neiladri Sinhababu, B. A.**

**Dissertation**

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

The University of Texas at Austin

May 2008

# A Treatise of Humean Nature

Publication No._________

Neiladri Sinhababu, Ph.D

University of Texas at Austin, 2008

Supervisors: Brian Leiter and David Sosa

A strong version of the Humean theory of motivation (HTM) that includes two theses is defended here. First, desire is necessary for action, and no mental states are necessary for action other than a desire and an appropriate means-end belief. Second, desires can be changed as the conclusion of reasoning only if a desire is among the premises of the reasoning. Those who hold that moral judgments are beliefs with intrinsic motivational force cannot accept HTM, even as a contingent truth, since HTM implies that no beliefs have intrinsic motivational force. Many of them argue that there are cases where HTM fails to explain how we deliberate. The response is to develop a novel account of desire and show that HTM provides superior explanations even in their cases.

On this account, desire necessarily motivates action when combined with an appropriate means-end belief. Desire necessarily causes pleasure when our subjective probability of satisfaction increases or when we vividly imagine satisfaction, and likewise causes displeasure when the subjective probability of satisfaction decreases or when we vividly imagine dissatisfaction. It is contingently true that desire directs attention towards things one associates with its object, is made more violent by vivid sensory or imaginative representations of its object, comes in the two flavors of positive desire and aversion, and satisfies the second principle above.

This account of desire helps HTM provides superior explanations of deliberation even in the cases that its opponents offer as counterexamples. In response to Darwall's proposed counterexample to the second principle and some 20th century writers discussing the feeling of obligation, it is shown that Humeans can provide superior explanations of agents' emotions in their cases. In Searle's case of akrasia, Scanlon's case of bracketing, and Schueler's case of deliberation, it is shown that Humeans can build the structures of deliberation more simply than their opponents can. Against Korsgaard, it is argued that agents cannot choose the aims for which they act.

## Table of Contents

# Chapter 1 The Humean Theory and the Moral Problem

## Introduction

This initial chapter will begin by presenting a version of the Humean theory of motivation. Then I will examine cognitivism and internalism, the other two components of what Michael Smith calls the “Moral Problem.” Of these three plausible theses, only two can be maintained, and there is much disagreement about which among the three theses should be rejected. Some philosophers have defended a combination of cognitivism and internalism while arguing that the Humean theory should be rejected. I will show that these philosophers need to argue against even the contingent truth of the Humean theory. As I will argue in later chapters, these objections are unsuccessful, and the Humean theory can respond to them in a way that demonstrates its superiority over competing views of motivation. This gives us a reason to reject cognitivist internalism.

## The Humean Theory of Motivation

The central idea of the Humean theory of motivation is that desire is the source of all of our motivations to act. Desire plays an essential role in explaining motivation, and this role cannot be usurped by any other mental state. While Humean theorists generally regard both a desire and a means-end belief as necessary for action, they distinguish themselves from other theorists by giving desire a more important role in motivating action than their opponents do. Furthermore, they hold that these two mental states are sufficient for a psychological explanation of all action.

On a Humean view, desires and beliefs are what Hume called "distinct existences" – the existence of one does not imply the existence of the other. This view is fairly intuitive. Two people with exactly the same beliefs may act very differently if they differ in their desires. And two people with exactly the same desires may act very differently if they differ in their beliefs.

Hume himself goes even farther than this. He not only holds that beliefs and desires are logically distinct existences – that the presence of a belief does not entail the presence of a desire – but that no belief or combination of beliefs is itself sufficient to create any particular desire through a process of reasoning. Since he also thinks that all action is motivated by the combination of a desire and a means-end belief, Hume accepts that processes of reasoning are only able to influence action by providing the means to antecedently desired ends. He expressed this view in *A Treatise of Human Nature*, offering perhaps the most famous slogan of the Humean theory: "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" (2.2.3).

I will defend a version of the Humean theory of motivation [HTM] which consists of both of the following claims:

> **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.

> **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.

Here I will consider two other formulations of the Humean theory – one from Michael Smith, and one from Ralph Wedgwood – and explain why I have formulated the theory as I have.

When he presents his detailed discussion of the Humean theory in *The Moral Problem*, Michael Smith presents and defends two principles that Humeans are committed to. The stronger principle, which he regards as the crucial one, comes from Davidson, and goes as follows:

> **P1**: R at t constitutes a motivating reason of agent A to φ iff there is some Ψ such that R at t consists of an appropriately related desire of A to Ψ and a belief that were she to φ she would Ψ.

For the desire and belief to be "appropriately related," Smith says, is for the agent to mentally put both of them together, so that they can interact to cause action. "Motivating reasons" are to be contrasted with normative reasons. Motivating reasons explain action, while normative reasons justify action. Smith says that "motivating reasons would seem to be *psychological states*, states that play a certain explanatory role in explaining action" (96).

One significant difference between HTM and P1 is that HTM allows views on which actions can be motivated with no assistance from beliefs to count as Humean, while P1 requires that a belief play some role in motivating the action. Consider a situation where an agent has a desire to engage in some immediate bodily action – perhaps, a desire to move his hand. If he moves his hand as a result of this desire, is his action also the result of a belief that were he to move his hand, he would move his hand? A defender of P1 would have to suggest that this trivially true belief is a motivating reason for his action. Nothing in the phenomenology of acting on a desire to engage in some immediate bodily action licenses us to ascribe some explanatory role to trivially true beliefs as part of the process. In order to avoid the consequence that trivial beliefs like this explain immediate bodily actions, a theorist might reject P1 while accepting DBTA and saying that desires alone can sometimes cause action. Such a view would belong naturally within the Humean camp, as desire is given a fundamental role in explaining action – indeed, a more fundamental role than belief.

A defender of P1 might respond to this by arguing that trivially true beliefs have a role in motivating actions driven by a desire to engage in some immediate bodily movement, and that our formulation of the Humean theory should make room for this. One way to test for whether particular beliefs have a role in explaining action is to consider counterfactual situations in which agents lack these beliefs, and see whether action ensues in these situations. So to test whether the belief that by moving his hand he would move his hand plays a role in explaining action, the defender of P1 might suggest that we consider a strange counterfactual situation – the situation in which an agent desires to move his hand, but does not believe that by moving his hand he can move his hand. If beliefs are necessary for action, such an agent would be unable to act. But if beliefs are not necessary, such an agent would be able to act. While it is common for agents to lack belief in complex conceptual truths like the truths of advanced mathematics, it is harder to imagine an agent who lacks belief in the simple conceptual truth that by moving his hand he would move his hand.

What would it even mean to not believe this? One possibility is that the agent would do things that are usually evidence of having a desire to move one's hand, and that do not require any sort of belief – happily daydreaming about moving his hand, for instance. But perhaps such an agent would be unable to engage in behaviors that required such a belief – for instance, actually acting.

One might want to say, instead, that there is something incoherent about an agent's lacking a belief on such a simple topic. If a functional characterization of belief is correct, and if there is no coherent set of inputs and outputs that would license the attribution of such a belief, then this belief would be incoherent.

It is hard to know what to say about these kinds of cases, so I have formulated the Humean theory in a way that is neutral as to whether belief is required for action. The important thing that the Humean theory says about beliefs is that no belief can be sufficient for action, and (as I will discuss further) no belief can generate action through a process of reasoning whose other premises consist entirely of beliefs. The insufficiency of belief for action is part of DBTA. DBTA and DODI together entail that processes of reasoning which have only beliefs as premises will not produce action. Since desires are necessary for action, and since beliefs cannot cause desires through processes of reasoning, no collection of beliefs will be sufficient to generate action through reasoning.

HTM, through DODI, denies the Humean label to views on which desires can be generated, eliminated, or altered in strength through processes of reasoning that do not have desires as premises.  On these views, beliefs might be able to motivate action indirectly, by giving rise to desires through processes of reasoning that involve no other desires.  P1 allows such views to count as Humean.  In this regard, HTM is truer to the historical origins of the Humean theory, and to the central thought that drives it.  If beliefs can create new desires through processes of reasoning that do not include desires as premises, reason is far more than the slave of the passions.  Rather than merely serving and obeying an agent's desires, reason can raise an army of new desires to oppose them.  While all motivation may still involve desire in some way, some kinds of motivation – namely the kinds of motivation caused by desires created through these processes of reasoning – will not have desire itself as their ultimate source.  So a DODI-style constraint on the ability of reason to generate new desires will be part of a properly formulated Humean theory.

In "Practical Reason and Desire," Ralph Wedgwood considers a version of the Humean theory of motivation that consists of three propositions, the first two of which are similar to the two propositions I have presented above as constituting the Humean theory.  They run as follows:

> **1** Whenever one acts with the intention to φ, one's action is motivated (at least in part) by the desire to φ.  (346)

This first component is analogous to DBTA.  Like DBTA, it allows that desire alone may be able to motivate action.  It does not, however, explicitly mention belief.

The Humean theory should be formulated in such a way that mental states other than desire and belief – for example, states of will or intention irreducible to desire and belief

– cannot be necessary conditions for action. DBTA restricts the set of necessary conditions for action to desire and belief, while 1 does not. It is part of the theoretical motivation for the Humean theory to provide a simple and economical explanation of deliberation and action, and the inclusion of more necessary motivational components would thus go against the Humean spirit.

> **2** Whenever a desire is motivated at all, its motivational history includes some further desire. (346)

This component is similar to DODI. The "motivational history" of some mental state, in Wedgwood's terms, includes "not only the mental states that motivated it, but also the mental states that motivated those further motivating states, and so on" (346). If a desire's being motivated is equivalent to its being generated by some process of reasoning, 2 and DODI will be equivalent.

> **3** All motivation must flow, ultimately and in part, from unmotivated contingent desires. (347)

An unmotivated contingent desire is "a desire that one is not rationally required to have" (347).

First I will explain why I do not include anything similar to 3. One problem with including 3 is that it makes the nature of the Humean theory depend on normative questions about the rational requirements on agency. If it turned out to be part of the correct theory of practical rationality that having some particular desire were rationally required – perhaps a desire for pleasure, or a desire to do what one has the most reason to do, or a desire to do the good – then the Humean theory of motivation would have to say that this desire could not motivate action. This would be a strange consequence. While the Humean theory of motivation, when combined with other plausible claims, has implications for various aspects of our normative theorizing, its content and its truth should not be regarded as dependent on normative facts in any way, and thus it should not be defined in normative terms.

If there is anything essential to the Humean theory that 3 captures, it is the idea that the mental states that motivate action cannot be the product of rational requirements that apply to all agents, regardless of what they desire, or whether they desire anything at all. If it is true that all agents to whom a rational requirement applies have the capacity to reason in accordance with this requirement, then agents will have the capacity to produce new desires, no matter what their pre-existing desires are. But Wedgwood's 2, like DODI, already blocks this sort of desire-generation. So it seems that 3 is superfluous.

We can formally represent HTM as it applies to the relation between belief and action in the following way, with B standing for the predicate "is a belief," C standing for the predicate "is capable of causing action by itself," and R standing for the predicate "is capable of causing action through processes of reasoning whose other premises include only beliefs."

**HTM on Belief and Action [HTMBA]**: $(\forall x)\ (Bx \rightarrow \neg\ (Cx \vee Rx))$

## Smith and the Moral Problem

The idea that Humean views about the structure of motivation have consequences for the objectivity of morality and the motivational force of moral judgments is at least as old as David Hume himself. Having argued that reason itself cannot give rise to a motivation, he continues:

> Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason alone, as we have already prov'd, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (3.1.1)

Since moral judgments have an intrinsic power to influence our motivational and affective states, and – as Hume's view of motivation claims – reason cannot do this, moral judgments must not be derived from reason.

Both the additional premise of Hume's argument and the negation of his conclusion are widely regarded as intuitively appealing. The additional premise is that moral judgments "have an influence on the actions and affections." The negation of the conclusion is that the rules of morality are conclusions of our reason. Someone who accepted both of these views, then, might respond to Hume's argument by denying his theory of motivation.

The claims at stake here have been laid out as follows by Michael Smith in *The Moral Problem*:

1 Moral judgments of the form 'It is right that I φ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.

2 If someone judges that it is right that she φ s, then, ceteris paribus, she is motivated to φ.

> 3 An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences. (12)

The first claim is cognitivism. The second claim is internalism. The third claim is a version of the Humean theory of motivation.

If cognitivism is true, moral judgments must express beliefs. If internalism is true, moral judgments carry intrinsic motivational force – they need not be combined with any other mental state to cause action. But if the Humean theory is true, beliefs do not have intrinsic motivational force, since desires are necessary for motivation. The cognitivist thesis that moral judgments express beliefs contradicts the thesis – derivable from the other two propositions – that moral judgments express mental states other than belief. If all three of these theses are true, it seems impossible for any agent to make a moral judgment.

While Smith decribes the three propositions making up the moral problem as "apparently inconsistent," he holds that they are not actually inconsistent, and goes on to present a theory that can reconcile them later in the book (12). According to Smith's view, while actions are always motivated by desire-belief pairs, agents can change their desires through processes of deliberation in which desires do not figure as premises. He claims that "by far the most important way in which we create new and destroy old underived desires when we deliberate is by trying to find out whether our desires are *systematically justifiable*" (158-159). We determine the systematic justifiability of a desire in a process similar to Rawls' method of reflective equilibrium, by "trying to integrate that desire into a more coherent and unified desiderative profile and evaluative outlook" (159). A coherent and unified set of desires, according to Smith, is rationally preferable, and the belief that a particular arrangement of desires would be rationally preferable is capable of changing a rational agent's desires towards that arrangement. Smith describes how our evaluative beliefs generate desires: "an evaluative belief is simply a belief about what would be desired if we were fully rational, and the new desire is acquired precisely because it is believed to be required for us to be rational" (160).

While the concept of desire figures in the contents of the beliefs involved in reasoning one's way towards new desires, no actual desires are among the premises of the reasoning. Smith's account, then, violates DODI and runs afoul of the Humean theory as I have defined it. However, Smith still regards himself as a Humean, as he operates with a weaker formulation of the Humean theory than I do. His formulation, P1, requires only that actions be motivated by desire-belief pairs, and he places no restrictions on how the desires themselves are generated. A stronger formulation of the Humean theory like HTM would rule out the sort of belief-driven desire-creation and desire-elimination that Smith's solution to the Moral Problem relies on. If one regarded HTM as a more suitable formulation of the Humean theory, one might criticize Smith for setting up a defective version of the Moral Problem containing an excessively weak Humean theory, and exploiting this defect to offer a solution.

Having criticized Smith's formulation of the Humean theory in setting up the Moral Problem, I will now look into the other two components of the problem and examine them to see exactly how they should be formulated.  Then I will return to the Moral Problem itself.

## Formulating Cognitivism

The central idea of cognitivism is that moral judgments express beliefs, and that they are thus evaluable in terms of their truth or falsity.  One of the most significant motivations for cognitivism is that it seems like the most natural way to account for the nature of moral thought and discourse.  As Darwall, Gibbard, and Railton say, "A classic problem for noncognitivists is that moral judgments have so many earmarks of claims to objective truth" (16).[1]  Participants in moral discussion take some moral judgments to be true and others to be false, and they often embed moral judgments in larger linguistic expressions, including conditionals, questions, and counterfactual statements, which have close relations to truth and falsity.  Cognitivists think that these features of moral discourse can be explained very well if we accept that moral judgments express beliefs. The connection between moral judgments and the truth is dealt with particularly well by cognitivism, as belief distinguishes itself from many of our other mental states by aiming at the truth.  Ordinary descriptive judgments, which are uncontroversially regarded as expressing beliefs, can be embedded in many of the same contexts that moral beliefs can. If cognitivism is true, this would explain how the same logical apparatus that we use in accounting for the semantics of nonmoral descriptive discourse can be extended to cover moral cases and give us a satisfactory account of embeddings.

[1] "Towards *Fin De Siecle* Ethics: Some Trends"

Recent noncognitivists have tried to show that their theories are capable of matching the success of cognitivism in accounting for these features of moral discourse. They have offered innovative proposals that seek (for example) to explain how moral judgments can be embedded in larger expressions much the way that ordinary descriptive judgments can, even on the assumption that moral judgments do not express beliefs. Cognitivists have responded by arguing that the noncognitivists' semantic proposals are not successful, and that some aspects of moral thought and discourse resist noncognitivist treatments.[2] For reasons of space and focus, I will not give detailed consideration to these issues. Instead, I will consider how cognitivism ought to be formulated.

Mark Van Roojen[3] has divided noncognitivism into two theses: semantic factualism and psychological noncognitivism.

> **Semantic Nonfactualism**: Moral judgments do not express propositions or have truth conditions.
>
> **Psychological Noncognitivism**: Moral judgments do not express beliefs or any other similarly cognitive state.

On Van Roojen's taxonomy of views about moral semantics, acceptance of either thesis is sufficient for noncognitivism.[4] Cognitivists accept the negations of both of these theses.

[2] A characteristic cognitivist response is Bob Hale's "Can There Be A Logic of Attitudes?" (1993).

[3] "Moral Cognitivism vs. Non-Cognitivism", Stanford Encyclopedia of Philosophy

[4] Van Roojen discusses two theories that only accept one of the theses, and he labels both of these theories noncognitivist. One of the theories is moral fictionalism, and the other is the nondescriptivist cognitivism of Timmons and Horgan. Of these, he says, "if the views are coherent this would suggest the two negative theses are logically independent."

Since beliefs are mental states with truth conditions, psychological non-cognitivism and semantic non-factualism are very tightly linked.  It seems generally plausible that if a mental state has truth conditions and is capable of linguistic expression, the linguistic expression of that mental state will have truth conditions.  If this conditional holds, psychological cognitivism will imply semantic factualism.  The logical properties of a linguistic expression should be derived, in some similar way, from the logical properties of the mental state that they express.  In view of this plausible relation between semantic nonfactualism and psychological noncognitivism, and because I am more concerned with psychological issues than linguistic issues, I will regard cognitivism as being the negation of the thesis that Van Roojen calls "psychological noncognitivism," which deals with the psychological states expressed in moral judgment.

Should cognitivism be regarded as holding that moral judgments express only beliefs, or can a cognitivist say that moral judgments express a combination of mental states including belief and some noncognitive state?  Saying that moral judgments express only beliefs, and that they do not express any other mental state, is traditionally regarded as truer to the spirit of cognitivism.  On Van Roojen's taxonomy, theories are termed noncognitivist if they say that moral judgments express collections of mental states including some noncognitive state.  He says that "many non-cognitivists hold that moral judgments' *primary* function is not to express beliefs, though they may express them in a *secondary* way."  Theorists like R. M. Hare, who held that moral judgments had both a prescriptive and descriptive function, are usually regarded as falling within the noncognitivist camp.

It is not hard to see why views on which some additional state is necessary for moral judgment as well as belief are widely regarded as being noncognitivist. Such views will not be able to account for moral discourse using the same logical apparatus that is used for nonmoral descriptive discourse, and which cognitivists regard as sufficient for handling moral discourse. Consider the following argument:

A: Arson is wrong.

B: If arson is wrong, then getting your little brother to commit arson is wrong.

C: Getting your little brother to commit arson is wrong.

Theorists of all stripes will want to say that A and B together imply C. And on a standard cognitivist account where moral judgments express only beliefs, the logic of ordinary descriptive statements that express beliefs can be deployed to account for this conclusion. By *modus ponens*, C follows from A and B.

But if we think that moral judgments express some additional mental state that is not truth-evaluable, and if we think that the truth-evaluability of a linguistic expression is dependent on the truth-evaluability of the mental state it expresses, we will have to regard A and C, and possibly B as well, as expressing some non-truth-evaluable state. The familiar logic of descriptive statements does not account for whatever sorts of logical relations might hold between statements with non-factual content. To account for the relationship between A, B, and C on a theory where moral judgments express combined mental states, we will need to go beyond our logical apparatus for dealing with descriptive statements. The proponent of the theory on which moral judgments express combinations of belief and a noncognitive state will have to bear the burden of developing a logic of non-cognitive attitudes that can explain how inferential relations hold between the non-factual contents of A, B, and C, and also between the mental states they express. This burden is characteristic of noncognitivist theories that attempt to make sense of ordinary moral discourse.

Here I will not inquire into whether this burden can be discharged. Certainly, if noncognitivists are successful in building a logic of non-cognitive attitudes that accounts for the features of moral discourse, the appeal of cognitivism will be diminished, and noncognitivism will offer an appealing solution to the moral problem. My point is just that the project of developing a logic of non-cognitive attitudes is a distinctively noncognitivist project, and a substantial one.

The foregoing discussion leads us to this formulation of cognitivism:

> **Cognitivism**: All mental states whose expressions are moral judgments are beliefs.

Cognitivism can be formally represented as follows, with J standing for the predicate "is a mental state whose expression is a moral judgment," and B standing for "is a belief":

> **Cognitivism**: $(\forall x)\ (Jx \rightarrow Bx)$

## Formulating Internalism

Michael Smith begins his discussion of internalism with the following example:

> Suppose we debate the pros and cons of giving to famine relief and you convince me that I should give. However when the occasion arises for me to hand over my money I say 'But wait! I know I *should* give to famine relief. But you haven't convinced me that I have any *reason* to do so!' And so I don't. (60)

As Smith says, this outburst would "occasion serious puzzlement… *Believing I should* seems to bring with it *my being motivated to* – at least absent weakness of will and the like" (60).

Following the schema of naming introduced by Stephen Darwall in "Reasons, Motives, and the Demands of Morality: an Introduction,"[5] we can call the variety of internalism suggested in Smith's example "morality/motivations judgment internalism". According to this view, whenever an agent makes a moral judgment that it would be right to do something, she will feel some motivation to do that thing. To incorporate Smith's last proviso that I have presented above, this connection between morality and motivations will hold at least in cases where the agent is not subject to weakness of will, depression, or some other kind of state that inhibits her motivation or powers of reasoning. What is important is that there be pathways of reasoning and motivation by which the moral judgment can motivate action, even if occasional defects in the agent's rationality and motivational capacity sometimes interfere. The motivation to do the right thing in cases where an agent makes a moral judgment need not be able to override other motivations in generating action, but some motivation to do actions one judges to be right or to refrain from actions one judges to be wrong must be present.

Traditionally, internalists have held that the connection between moral judgment and motivation is logically necessary. According to William Frankena, who was among the first to formulate it, internalism is the thesis that it is "not logically possible… for an

[5] Darwall only mentions morality/reasons and reasons/motives internalism in his article. These views together imply the truth of morality/motives internalism.

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" (40-41). The versions of internalism that Smith considers all have the same modal strength as the one that Frankena proposes – each of them involve a "conceptual connection" between moral judgment and motivation. (The nature of the conceptual connection differs in Smith's three formulations – in one it is a conceptual truth that moral judgment entails motivation, in another it is a conceptual truth that moral judgment entails motivation unless the agent is irrational, and in a third moral judgment entails having a reason to act.) The role of characters far from actuality like Milton's Satan as examples in the internalist-externalist debate further testifies to the modal strength of the internalist thesis as it is generally defended.

There is no need for internalists to claim that the mental state expressing the moral judgment must itself be the proximal cause of action. Smith's solution to the Moral Problem, for example, involves beliefs modifying an agent's desires, which end up being the proximal causes of action. The possibility that internalists are concerned to reject is that moral judgments might be entirely motivationally inert. If they are motivationally potent, but only in virtue of their power to generate other mental states that can be proximal causes of action, the basic intuition underlying internalism will be satisfied. Looking back at Smith's famine relief example, it would not "occasion serious puzzlement" if he was willing to give to famine relief because his new moral belief had generated a desire in him to give to famine relief, and this desire had motivated him to act.

When internalists say that moral judgments are sufficient for motivation, they do not mean that the mental states expressed in these judgments would have to bring about action without any aid from means-end beliefs. Even if I judge that it would be right to give money to famine relief, I may not act if I do not believe that I have any money. What internalists want of the mental states expressed in moral judgment is that they can play a role similar to the role that Humeans attribute to desire in motivating action, or produce mental states capable of playing this role. It would suffice for the truth of internalism if the mental states expressed in moral judgments, while not themselves being desires, were capable of producing desires through processes of reasoning, and if these desires could be combined with beliefs through further processes of practical reasoning to bring about action.

What internalists cannot allow is that motivational mental states – that is, states like desire – which are not expressed in our moral judgments are necessary conditions of our acting on our moral judgments. (Of course, if noncognitivism is true and moral judgments express desires or similar states, the necessity of desire for moral action will be consistent with internalism.) If motivational states in addition to the states expressed in moral judgment were necessary, the power to motivate action would not be intrinsic to the mental states expressed in moral judgment. It would then be possible for someone to respond in the puzzling manner that Smith describes in his example, without any faulty reasoning, if they merely lacked the desire to do what was right. Furthermore, Smith holds than an agent who acts out of a belief that some particular thing is right to do and a desire to do the right thing, where "do the right thing" is read *de dicto*, is not genuinely acting out of a moral judgment. This, Smith says, is merely a "moral fetish" (76). For an agent's honest action to genuinely issue from a moral judgment, the agent cannot merely desire to do the right thing *de dicto*, believe that the honesty of a particular course of action will make it right, and from these premises derive a desire to do the honest thing. Rather, the agent must be motivated by an intrinsic desire, not derived from any processes of means-end reasoning, which either issues from or is constituted by the moral judgment.

We can formulate internalism as follows:

> **Internalism**: All mental states expressed in moral judgment are capable of causing action, either by themselves or through processes of reasoning whose premises include only beliefs.

Internalism can be formally represented as follows, with J standing for the predicate "is a mental state whose expression is a moral judgment," C standing for the predicate "is capable of causing action by itself," and R standing for the predicate "is capable of causing action through processes of reasoning whose other premises include only beliefs."

> **Internalism**: $(\forall x)\ (Jx \rightarrow (Cx \lor Rx))$

This formulation is not intended to capture the modal status that internalists usually ascribe to their position.

## The Moral Problem Revisited

Now I will return to the moral problem. Here are cognitivism, internalism, and the portion of the Humean theory that deals with belief, together.

**Cognitivism**: $(\forall x)\ (Jx \rightarrow Bx)$

**Internalism**: $(\forall x)\ (Jx \rightarrow (Cx \vee Rx))$

**HTMBA**: $(\forall x)\ (Bx \rightarrow \neg(Cx \vee Rx))$

These theses do not generate a contradiction. To get a contradiction, this fourth thesis should be added, with E representing the existential quantifier:

**Moral Judgment**: $(\exists x)\ (Jx)$

This fourth thesis is simply that mental states whose expressions are moral judgments exist. When added to cognitivism, internalism, and HTMBA, it forms a contradictory tetrad.

Smith describes the three propositions constituting his version of the Moral Problem as "apparently inconsistent" (12). It is because of his solution, which relies on an excessively weak version of the Humean theory, that he thinks that cognitivism, internalism, and the Humean theory are actually consistent. As it turns out, internalism and cognitivism are jointly consistent even with a stronger version of the Humean theory. But this consistency does not offer anything resembling an attractive solution to the Moral Problem, because if one holds all three theses, one is committed to denying that mental states whose expressions are moral judgments exist. This consequence amounts to a *reductio ad absurdum* of the three theses taken jointly.

If one wanted to reformulate the Moral Problem as a contradictory triad instead of a tetrad, one could try to make cognitivism, internalism, and the Humean theory inconsistent by attaching existential commitments to cognitivism or internalism. One would do this by making it part of the cognitivist or internalist thesis that moral judgments exist. But there is good reason not to go this route. If someone says that cognitivism is necessarily true, it would be beside the point to respond that there are possible worlds containing only asteroids and no minds capable of making moral judgments. Such possibilities should not be taken to refute Frankena's claims about the conceptual necessity of internalism either. So we should interpret both cognitivism and internalism as lacking existential commitment.

In the end, the metaethical state of play involving the Moral Problem is much as it originally appeared. Of cognitivism, internalism, and the Humean theory, one must choose at most two. This is not because the three theses are contradictory, but because they together commit one to deny the existence of mental states whose expressions are moral judgments. Humean internalists, then, have to deny cognitivism. Humean cognitivists have to deny internalism. Cognitivist internalists have to deny HTMBA (and thus HTM). It is to this last position, cognitivist internalism, that I will now turn.

## Cognitivist Internalism vs. The Contingent Humean Theory

Whether they accept both theses or just one of them, cognitivists and internalists both hold that their views are true at least about the actual world. They do not merely argue that it is a counterfactual possibility that moral judgments express beliefs or motivate action. Cognitivists argue that we can best account for actual moral discourse if we accept that moral judgments express beliefs. Internalists have traditionally held that connections to motivation are conceptually necessary for a mental state's being such that its expression constitutes a moral judgment, and this implies that moral judgments in the actual world express mental states that can motivate action.

Defenders of the combined cognitivist internalist position who accept that there are actual moral judgments are thus committed to the actual truth of three of the four jointly inconsistent theses in the Moral Problem as I have laid it out. So these cognitivist internalists cannot declare victory after showing that it is conceptually or metaphysically possible for beliefs to motivate action. They cannot merely argue that the Humean theory of motivation is not a metaphysically necessary or logically necessary truth. They need to argue for the falsity of the Humean theory in the actual world. Opponents of the cognitivist internalist position, meanwhile, can argue against the conjunction of cognitivism and internalism by arguing for nothing more than the contingent truth of HTM. If we find convincing evidence for the contingent truth of HTM, cognitivism and internalism can only be jointly true if no actual mental states exist whose expressions are moral judgments. This would be a *reductio ad absurdum* of cognitivist internalism.

It is this argumentative strategy that I will pursue in the next two chapters. While cognitivist internalists have argued that the Humean theory is unable to account for structural and phenomenological features of actual deliberation, I will respond to them by showing 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 anti-Humeans' cases show why we ought to accept the Humean theory. This gives us reason to think it is correct, and to reject cognitivist internalism.

My arguments will not give us reason to regard the Humean theory of motivation as a conceptually or metaphysically necessary truth. I will concede to my opponents that we can imagine creatures whose deliberative processes and actions are not accurately described by the Humean theory. These, however, would be creatures for whom the psychological laws governing deliberation and action are deeply different from the psychological laws governing the deliberation and action of human beings. If, in the face of my arguments, cognitivist internalists want to stick to their guns and hold that the capacity for a kind of deliberation that violates the Humean theory is indeed necessary for moral judgment, they will be forced to defend the deeply counterintuitive position that while some conceivable creatures might be able to act morally, nothing humans are psychologically capable of could count as a moral judgment.

While humans are universally regarded as psychologically capable of moral judgment, the idea that some creatures lack this capacity because they lack the appropriate psychological features is uncontroversial. Some cognitivist internalists have offered accounts of the features in virtue of which humans and animals differ. Christine Korsgaard, for example, holds that the human capacity for moral judgment rests on the ability of humans to generate new motivational states in a way that is inconsistent with the Humean theory. Animals are incapable of moral judgment, she thinks, because their motivational psychology follows the basic outlines of the Humean theory, and they are unable to generate new motivational states in the way that we can.[6] If Humean views about motivational psychology are correct, human motivation is governed entirely by the

[6] This view is expressed in the 3rd and 4th of her *Locke Lectures*.

laws that Korsgaard regards as the laws of animal motivation. The laws of animal psychology that Korsgaard sketches out adequately characterize the way that humans can generate new desires through reasoning. (This does not mean that humans are psychologically the same as animals – certainly there are differences, and humans are capable of many mental processes that many animals are incapable of.) Combining Korsgaard's claims about the necessary psychological conditions for moral judgment with the psychological laws governing humans would give us the conclusion that humans cannot make moral judgments. This is why Korsgaard's claims about the necessary conditions for moral judgment, which involve commitments to both cognitivism and internalism, should be rejected.

How is it possible that human beings could be psychologically incapable of some processes of reasoning of which some conceptually possible nonhumans are capable? Considering the Korsgaardian position that I have described may be instructive here. It does not seem intuitively implausible that animals are psychologically incapable of some processes of reasoning that human beings are psychologically capable of. After all, the biological hardware and wiring of human and animal brains are different, and the physical structure of a creature's brain determines which psychological laws govern its mind. Perhaps we humans are similarly incapable of some processes of reasoning that higher beings yet, whose brains physically differ from ours, would be capable of. And perhaps we would still regard the distinctively nonhuman processes by which their mental states interacted as processes of reasoning, despite our own inability to reason in these ways. This opens up the farfetched but eye-catching possibility that we could, despite lacking moral agency ourselves, come to learn that other beings with different mental capacities are genuine moral agents.

Since I hold that it is conceptually possible for creatures with non-Humean motivational psychologies to exist and engage in action, I will not be able to argue that the Humean theory is true as a matter of conceptual necessity. Nor will I argue that it is true as a matter of metaphysical necessity, like the identity between water and $H_2O$. In my view, the Humean theory is a correct empirical psychological theory. We will discover the truth of the Humean theory by seeing that it offers simpler and more powerful explanations of our observations than its competitors do.

Unfortunately, very little rigorously collected data from the disciplines of empirical psychology and neurobiology has come to light on the particular questions that would confirm or disconfirm the Humean theory.[7] So I will meet the opponents of the Humean theory on the same methodological terrain that they argue from in presenting their counterexamples. They argue that the Humean theory is unable to deliver successful explanations of the observations that underlie our commonsense folk psychology. I will argue that the Humean theory not only explains these observations adequately, but that it outperforms its competitors in terms of simplicity and explanatory power. Many of the

[7] Part of this has to do with the ways that the particular groups of psychologists to whom one might look for data have focused their research. Cognitive scientists do very systematic work on the operations of the mind, but focus much more on belief and theoretical reasoning than on desire and practical reasoning, while social psychologists often come up with interesting results about motivation, but rarely build their findings into a sufficiently systematic theory to be useful for the purposes of this project. The impression of this that I got while sifting through the psychology literature was confirmed by Art Markman. While some cognitive scientists, like Markman, are focusing on motivation, I have yet to find results from them on the particular questions that would bear on the truth of the Humean theory. For example, I have not found rigorously collected empirical results on when people have experiences of pleasure and displeasure during processes of reasoning, and I have been told that such results are not currently available.

cases that are presented as problematic for Humeans have features that are actually better explained by the Humean theory than by its competitors.

My argument for the contingent truth of the Humean theory will have two stages. In the next chapter, I will set out the conceptually necessary conditions for a mental state's being a desire, as well as the properties that desire has throughout the actual world and throughout the space of human psychological possibility. To understand the empirical predictions that the Humean theory generates, one must first understand the properties of desire and its relations to other mental states. Having set out the properties of desire, I will be in position to construct explanations of how we deliberate, even in the most complex situations, that are rooted in the way that desire motivates decision and action in the simplest situations.

In the third and final chapter, I will consider the objections that anti-Humeans have offered against the ability of the Humean theory to explain the structure and phenomenology of deliberation. The Humean theory I lay out will provide a simple and powerful explanation of many features of how we deliberate and act. While anti-Humeans charge that it cannot satisfactorily explain many actual cases of deliberation, I will argue that its explanations account for all the data in these cases, and defeat the anti-Humean explanations by fitting into a simpler total picture of motivation. This gives us some reason to accept the Humean theory, and to reject the cognitivist internalist position that is in tension with it.

# Chapter 2: What Humeans Say About Desire

## Introduction

This chapter, which focuses on desire, has two major aims. The first is to characterize the way in which desire interacts with other mental states and mental phenomena, including belief, pleasure, imagination, and action. It is from an understanding of these interactions that a Humean model of practical deliberation can be built. Empirical data from psychology and neuroscience will be useful in explaining how these interactions go. The second aim of this chapter is to determine which among these features of desire are necessary. There is much disagreement, even among philosophers who accept a broadly Humean theory, about which relations to other mental states are necessary for something to be a desire, and which relations merely constitute contingent facts about desire. Over the course of this chapter, I will work out an analysis of desire.

In the course of determining what desires actually and necessarily are like, I will consider what previous philosophers sympathetic to Humean views about action have said about the nature of desire and the role of desire in practical deliberation. Many different views of desire – dealing both with its actual properties and its essential nature – have been offered by Humeans. Considering their views will both allow me to find useful components for my own theory of desire and give me opportunities to criticize competing positions. As I survey the views of Humeans present and past, I will also mine their writings for material that will be useful in overcoming the objections to the Humean theory which will be discussed in the next chapter. Some of the philosophers whom I will discuss in this chapter have offered desire-based explanations of phenomena that might otherwise be regarded as troublesome for the Humean theory, and I will take note of these helpful explanations.

Six theses about desire, each of which concerns the way that it interacts with other mental states, will be advanced in this chapter. Each one describes a functional property of desire, and it holds throughout the space of human psychological possibility. That these propositions are true of desire will be defended with reference to common-sense folk psychology as well as psychological and neurobiological research. While there are many other interesting contingent facts about desire, some of which will be discussed in this chapter, I highlight the ones below because some of them are necessary conditions for desire, and some are important for the explanations that are offered in the next chapter. Each formulation below will be presented again in this chapter, once suitable arguments for it have been offered.

> **The Motivational Aspect**: If an agent occurrently desires D, and she occurrently believes that she can bring about D by doing A, she will have a motivational tendency to do A. Her motivational tendency to do A will increase with the strengths of the desire and the belief. If at any time there is some action that she has the greatest motivational tendency to do, she will initiate that action.

> **The Hedonic Aspect**: If an agent occurrently desires D, increases in the subjective probability of D or vivid sensory or imaginative representations of D will cause her 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.

**The Attention-Direction Aspect**: Desiring that D will make an agent more likely to focus her attention on things she associates with D than things she does not associate with D.

**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.

**Intensification By Vivid Images**: When an agent is presented with vivid images she associates with a state of affairs she desires, either in imagination or by her senses, that will strengthen the desire's causal powers. The desire's phenomenal effects increase sharply, and its motivational powers increase substantially as well.

**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.

DODI is one of the two propositions making up the Humean theory of motivation, as it was laid out in the previous chapter.

If all these things are actually true of desire, which are necessarily true? Intuitions are very strong in favor of the necessity of the motivational aspect. Many people also have intuitions in favor of the necessity of the hedonic aspect, though some do not. I will defend the necessity of both of these aspects, and argue that all the others are only contingently associated with desire.

Before I begin, it may be useful here to say a word about the role of appeals to normal conditions in describing psychological states. Overuse of measures like the appeal to normal conditions can lead to a theory that tells us nothing in cases where we rely on it for explanation or prediction. While it seems clear that desire-belief pairs motivate action, an agent may not do A when she desires D and knows that doing A will bring about D, if she knows that doing A will leave her unable to satisfy some stronger desire. Attempts to use some notion of normal conditions to push these cases aside, by limiting the domain of the theory to normal cases and excluding cases involving conflicting desires as abnormal, can seem unsatisfactory. Situations where an agent has conflicting desires ought to be dealt with as normal cases. They are among the cases that we will rely on desire-belief psychology to predict and understand.

Another class of problem cases, however, ought to be finessed with some appeal to normal conditions. These are cases where, for example, an agent desires B and comes to know that doing A will bring about B, but is instantly decapitated by a ninja and never has the time to even try to act. If cases like this are taken as counterexamples to claims that desires have some kind of robust connection to motivation, it will be hard to make any interesting general statements at all about the causal properties of our mental states. Laws of economics, for instance, are not rejected just because they would make false predictions in cases where comets unexpectedly crash into the Earth and kill everyone. We instead regard the comet scenarios as abnormal and outside the scope of the laws. So the claims about the properties of desire which have been presented above, and which will be advanced in this chapter, should all be understood as restricted to some hard-to-characterize set of situations in which psychological explanations are the ones we seek.

### Hume, Passions, and Reason

The first theory I will consider is that of David Hume himself. Hume takes his opponents to be philosophers who talk of the struggle between passions[8] and reason, "give the preference to reason, and assert that men are only virtuous so far as they conform to its dictates" (2:3:2). He opposes them by arguing that "reason alone can never be a motive to any action of the will; and secondly, that it can never oppose passion in the direction of the will."

Elijah Millgram lays out two of Hume's arguments for this conclusion in "Was Hume a Humean?" Hume's first argument proceeds by examining the ways in which reason can operate, and discovering that none of them will be sufficient to motivate action. Hume contends that "the understanding exerts itself after two different ways, as it judges from demonstration or probability; as it regards the abstract relations of our ideas, or those relations of objects, of which experience only gives us information." This distinction between the ways that the "understanding exerts itself" is the distinction between a priori (demonstrative) and a posteriori (probabilistic) judgments. Since these two ways in which reason operates are equivalent to two different ways of arriving at

[8] Like most Hume interpreters, I will regard the terms "passion" and "desire" as equivalent. Context suggests an interpretation on which "reason" refers a faculty that is in charge of dealing with cognitive, belief-like states.

beliefs, it is appropriate to regard Hume as attacking rationalist theories according to which beliefs, without desires, are capable of generating motivation. A priori reasoning alone cannot cause action – in fact, its removal from the particular objects of our actions makes it insufficient to motivate action. Hume argues that "As its proper province is the world of ideas, and as the will always places us in that of realities, demonstration and volition seem, upon that account, to be totally remov'd, from each other." A priori reasoning can only affect action "as it directs our judgment concerning causes and effects" – that is, in virtue of its effects on probabilistic reasoning.

While Hume gives an account of how probabilistic reasoning can affect action, he holds that reasoning of this kind, by itself, is similarly insufficient for motivation. According to his picture, probabilistic reasoning must interact with desire to motivate action. First, when an object may cause us pain or pleasure, this causes us to feel desire or aversion towards it. The motivational emotion, "rests not here, but making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and effect." In making us think of the causes of the objects of our desire and allowing our motivational energies to focus on them, probabilistic reasoning does its entire work. We then pursue the causes of whatever it is that we desire. Hume points out that "the impulse arises not from reason, but is only directed by it." Our passions provide the impulses to action, and without them, the causal connections supplied by probabilistic reasoning would have no effect on action. So probabilistic reasoning has no motivational force of its own.

Hume concludes that since reason only operates in two ways, and neither of these is sufficient to cause action without the aid of the passions, reason is incapable of causing action by itself. Unfortunately, this argument is unlikely to convince anyone who believes that reason can give rise to motivations. The first premise – that the only two forms of reasoning are probabilistic and demonstrative – is too strong. Anti-Humeans may deny this premise, arguing that there is a third form of reasoning – practical reasoning.

Fortunately, Hume has a second argument for the conclusion that reason cannot motivate action. According to this second argument, which Hume presents more quickly, passions are "original existences" that do not represent the world as beliefs do, and which imply "no reference to other passions, volitions, and actions". Since mental states can only be true or false if they aim at correctly representing the world, passions cannot be true or false in the way that beliefs are true or false. And since "Reason is the discovery of truth or falsehood," there is no way for reason to produce or generate passions. Since passions do not have the properties of truth and falsehood, and since reason can only affect our mental states by affecting our attitudes as to the truth-values of propositions, reason cannot generate or eliminate passions.

This argument is better than the preceding one, in that it suggests a unifying explanation of why reason would be limited to the dual functions of making demonstrative and probabilistic judgments, and could not oppose desire in motivating action. Demonstrative and probabilistic judgments concern themselves with beliefs, which represent the world and are evaluable in terms of truth and falsehood, so reason can affect them. Desires do not represent the world and are not evaluable in terms of truth and falsehood, so reason cannot affect them. But there are a few avenues of escape for Hume's opponent.

One possibility is to defend the view that desire represents states of the world – perhaps, that it aims at the good, and that reason can affect desires in virtue of their representing states of the world as good. The view that desire represents states of the world, however, is hard to square with the way that desires can conflict with one another when we have mixed feelings about something. I can consistently have a desire to go to dinner with my girlfriend's parents (I would like to meet them) and have a desire not to go to dinner with them (the idea makes me nervous). Merely reflecting on my desires in such a situation will not be sufficient to make one of them go away. On the other hand, I cannot consistently believe that it is good to go to dinner with her parents, and that it is not good to go to dinner with them. It is difficult to believe something like this in the first place, and if one has somehow fallen into believing a contradiction, reflecting on one's beliefs is often an effective way of eliminating one of the beliefs and resolving it. Contradictions in representative mental states are hard to maintain – after all, the world does not contain contradictions, and these states represent the world. Since we can have stable contradictions in desire, desires cannot be representative mental states.

A second response to Hume's argument, probably more effective, is that reason can affect mental states other than those capable of truth and falsity. A version of this response is implicit in the instrumentalist position, according to which reason can take our desires and beliefs into view and determine whether we have a desire-belief pair that would make an action rational. Anti-Humeans who deny DODI will face some challenges in trying to account for the kinds of belief-caused desire-formation distinctive of rationalism along these lines, however. It is easy to see how one belief can bring another belief into existence via rational processes – if one of the propositions an agent believes implies another proposition that is not yet believed, rational processes may generate a belief in the second proposition. If the truth of the first proposition implies the truth of the second proposition, and if rational belief is responsive to evidence of truth, a rational agent would form a belief in the second proposition. Along similar lines, a belief might eliminate another belief that it contradicted. But if Hume is right that desire has no representative quality, it is hard to see how a process like this could generate or eliminate desires. Why would a mental state that does not represent the world, and thus is not responsive to evidence of truth or falsity, be affected in the same way that beliefs affect other beliefs?

Pointing out that these anti-Humeans face some explanatory challenges, however, is not yet a decisive objection against their view. There may be ways to handle these challenges. For example, one might argue that beliefs in some propositions – for example, the proposition that by doing some action one will fulfill an obligation, that some state of affairs is valuable, or that doing something is in one's interests – are sufficient to generate a desire. Or one might invoke reasons as primitive properties, and say that believing propositions that bear reasons is causally sufficient for the generation of a new desire. If there is no general and informative way of characterizing the propositions that have motivational force or that bear reasons, the theory will be lacking in simplicity.  But perhaps this is just the way that motivational propositions or reasons are.  Sometimes reality is not simple, and complexity in reality often demands inelegance in theory.

While Hume's argument raises difficulties for rationalist theories according to which beliefs can generate desires through a process of reasoning, there is a more direct way to determine whether it is a necessary truth that this kind of desire-creation is impossible. We can try to imagine counterfactual scenarios in which an agent comes to have a new motivational mental state that has all of desire's causal outputs, and in which the generation of this new motivational state is caused by a process of reasoning, all the premises of which are beliefs.  If it turns out that no such mental state could count as a desire, or that it is conceptually impossible for a motivational state to be produced by beliefs through a process of reasoning, DODI is a necessary truth.  However, if there are possible states of affairs where beliefs generate new desires through a process of reasoning, DODI will not be a necessary truth – though it may still hold true in the actual world.

Without further ado, I shall test DODI by presenting the example of the Angels:

> **The Angels**:  When Angels are born, they come into the world with no desires at all.  Whenever they form a new motivational mental state, it is because they have formed the belief that it would be right to engage in some future action.  There is a systematic and unfailing psychological process in them by which the belief that it will be right to perform a particular future action brings about a motivation to perform that action.  Other than this systematic tendency to form new

> motivational states, they have none of the mental processes that one might regard as suggestive of an antecedent desire to engage in right actions. For example, when vividly imagining counterfactual situations where they act rightly, they feel no pleasure, and when vividly imagining situations where they fail to act rightly, they feel no displeasure. But once they come to believe that it will be right to perform some particular future action, their new motivation to perform that action has all the associated features of desire.

The question is: Are the Angels' new motivational mental states, which they form after believing that it is right to do something, instances of desire? There are many different things one might say in response to this example. I will consider three of them, the last of which strikes me as the intuitive response.

First, one might say that the example is impossible, since it stipulates at the beginning that Angels do not have desires but then implies that they do. According to this response, the Angels' new motivational states intuitively count as desires, and it is a necessary truth that a desire cannot be created through a process of reasoning unless an antecedent desire exists. So we must attribute to them an antecedent desire to perform right actions. This response does not seem right to me, however, because the Angels were antecedently incapable of engaging in particular mental processes that are necessary for one to have a desire to do the right thing. The thought of doing particular counterfactual right actions did not please them, and the thought of failing to do so did not displease them. (Arguments for a necessary connection between desire, imagination of counterfactual desired states, and pleasure will be given in the section on Michael Smith and phenomenological theories.) Since the Angels lacked something that is necessary for desire in the beginning, their early mental states cannot count as desires.

Second, one might say that the example is coherent, and that the Angels' new motivational states do not count as desires. Someone who has this intuition most likely accepts the necessity of DODI, regards the Angels as not having desires before they engage in their reasoning, and thus regards the Angels' new motivational states as not being desires. As this is merely an intuition test, there is no arguing if someone genuinely has this intuition. But I will say that what draws me away from this position is the fact that the new mental state has the rich set of features typical of desire. Both in terms of motivation and in terms of the experiences an agent has while imagining counterfactual states of affairs, the new mental state has the features typical of desire. While the circumstances of its generation are unusual, the significance of this fact pales in comparison to the significance of its connections to pleasure and action.

The view that seems right to me, then, is that the example is coherent and the new motivational state is in fact a desire. Without having any antecedent desires, the Angels are capable of generating new desires through a process of reasoning. This would show that DODI, however true it may be in the actual world, is not a necessary truth. It is not part of the concept of desire that desires can only be generated by processes of reasoning if antecedent desires come into the picture.

In the previous chapter, I stated that I would be defending the contingent truth of the Humean theory, not its necessary truth. Now I hope it is clear why I take this position.

Since DODI is at best a contingent truth, the Humean theory can only be contingently true.

### Hume, Calm Passions, and Vivid Imagination

I have already considered one of Hume's claims about the nature of our desires – that they do not represent the world as being a certain way, and that they are thus incapable of truth or falsity. I will now go over one of Hume's major contributions to desire-based explanations of the phenomenology of deliberation – his account of calm passions.

In the same section where he presents the two arguments that reason cannot have any motivational force independent of desire, Hume introduces a distinction between calm and violent passions. While passions of both kinds are capable of motivating action in the same way, there is a phenomenological difference between them. The calm passions, "tho' they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation." Violent passions, on the other hand, are experienced more robustly. Hume remarks that "When I am immediately threatened with any grievous ill, my fears, apprehensions, and aversions rise to a great height, and produce a sensible emotion." In establishing this distinction between calm and violent passions, however, Hume does not mean to assert that each of our passions is fixed in its calmness or violence – by varying "the situation of the object," we can "change the calm and violent passions into each other." To change a calm passion into a violent one, one needs to bring the object of the passion closer to the agent. As Hume says, “The same good, when near, will cause a violent passion, which, when remote, produces only a calm one.”

Hume writes that calm passions can become more violent if their objects are imagined more vividly. As examples in support of a connection between vividness of imagination and the violence of passion, Hume 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 also offers a historical example from ancient Athens. Themistocles conceived 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 if he expressed it openly, he merely told the Athenians that he had a secret plan that would benefit them greatly. They 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 only able to conceive of the plan 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 possibility of naval supremacy, which allows for more vivid imagining, more violent passions in support of Themistocles’ plan would have been incited, and they might well have decided otherwise.

It is not hard to see how a connection between the vividness with which an object is imagined and the violence of the passion involved would explain a connection between the nearness of a passion's object and the violence of the passion. If nearer objects tend to be imagined more vividly – as is plausible – nearer objects would incite more violent passions. Indeed, this is the explanation Hume offers:

> There is an easy reason, why every thing contiguous to us, either in space or time, shou'd be conceiv'd with a peculiar force and vivacity, and excel every other object, in its influence on the imagination. Ourself is intimately present to us, and whatever is related to self must partake of that quality. But where an object is so far remov'd as to have lost the advantage of this relation… its idea becomes still fainter and more obscure (2:2:7).

Hume's associationist psychology is evident in this passage. The vividness with which we imagine something is explained, at least in part, by the closeness with which we associate it with something (in this case, the self) that is immediately present to us.

We can encapsulate Hume's discussion of the calm and violent passions in a single property of desire, as follows:

> **Intensification By Vivid Images**: When an agent is presented with vivid images she associates with a state of affairs she desires, either in imagination or by her senses, that will strengthen the desire's causal powers. The desire's phenomenal effects increase sharply, and its motivational powers increase substantially as well.

Phenomenal and motivational effects are the two kind of effects that Hume has discussed – the "sensible emotion" that the desire has, and its ability to influence our actions as the Athenians were influenced.

Certainly, the overall effect of vivid images can be otherwise than this formulation suggests. If I desire to get married, and then see a couple in an unhappy marriage or vividly imagine myself in such a situation, perhaps I will be less disposed to pursue marriage. But in this case, what is really affecting my motivational tendencies is the vivid image of marital strife, which may intensify any number of my aversions. Cases in which accurate vivid images give an agent new information that weakens her desires – perhaps, seeing someone eat a strange fruit that I had desired, and then vomit – similarly are not counterexamples to the principle listed above. Whatever intensification of my desire may have arisen from initially seeing the fruit is overwhelmed by the intensification of my aversion to vomiting and my new belief in the bad consequences of eating the fruit. The claim that vivid images intensify desire should be understood not as an all-things-considered claim about their overall effects, but as a claim about immediate functional outputs that may be overwhelmed.

On Hume's view, an increase in the violence of a passion does not bode the same way for its phenomenal and motivational effects. He says that "Tis evident passions influence not the will in proportion to their violence, or the disorder they occasion in the temper" (2.3.4) and describes how the force of custom can create strong but calm passions. A violent passion, then, is one with particularly powerful phenomenal effects, while the strength of a passion is demonstrated in its behavioral effects. This is why I have drawn from Hume's discussion of how vivid imagination increases the violence of a passion the conclusion that the phenomenal effects increase more sharply than the motivational effects.

With his distinction between calm and violent passions, Hume is able to explain the same phenomena explained by philosophers who ascribe motivational force to reason and claim that it opposes passion in the direction of the will. Hume argues that what these philosophers call the operations of reason in guiding the will are really calm passions at work. As calm passions "cause no disorder in the temper," their "tranquillity leads us into a mistake concerning them, and causes us to regard them as conclusions only of our intellectual faculties" (2.2.7). While it is true that we are sometimes motivated by mental states that do not cause the same "disorder in the temper" as violent passions, this can be explained better by the category of calm passions than by a motivationally efficacious faculty of reason. The explanation involving calm passions is simpler than the explanation involving a faculty of reason that stands in opposition to desire, as it invokes only one kind of entity where the reason-based explanation invokes two. Anti-Humeans cannot claim here that Hume has invoked two different kinds of entities simply because both violent and calm passions have entered into the explanation. As Hume shows in his discussion of how violent and calm passions can be converted into each other by making their objects more or less vivid to the imagination, his explanation only involves a single kind of mental state. Passions are at bottom the same with respect to their capacities to be calm or violent, though the situations of their objects differ, causing agents with passions for differently situated things to feel differently.

## Nietzsche, Direction of Attention, and the Standpoint of Reflection

The contributions of Friedrich Nietzsche to Humean views about how we deliberate have gone largely unnoticed. But much like Hume, Nietzsche tried to construct desire-based accounts of complex deliberative phenomena. In *Daybreak*, he lists six ways in which we can combat a vehement drive[9] that is tormenting us. We can avoid opportunities for its gratification, gratify it only on a certain schedule, overindulge it until we become sick of it, mentally associate it with something painful, squander our energy on something else, or depress our entire constitution to weaken it. But whichever method we choose, Nietzsche claims that

> in this entire procedure our intellect is only the blind instrument of *another drive* which is a *rival* of the drive whose vehemence is tormenting us: whether it be the drive to restfulness, or the fear of disgrace and other evil consequences, or love. While 'we' believe we are complaining about the vehemence of a drive, at bottom it is one drive *which is complaining about another*; that is to say: for us to become aware that we are suffering from the *vehemence* or a drive presupposes the existence of another equally vehement or even more vehement drive, and that a *struggle* is in prospect in which our intellect is going to have to take sides. (109)

In such a situation, we do reflect on the drive whose vehemence is tormenting us, but we do not typically reflect on the more vehement drive that directs our deliberation. The intellect is, as Nietzsche says, a "blind instrument" of the drive that controls it. As Philip Pettit and Michael Smith have argued in "Backgrounding Desire," the desire that motivates our actions often sits in the background of deliberation, and it need not itself

[9] Nietzsche uses "Trieb," which is usually translated as "drive," sometimes translated as "instinct," and less often as "desire." The term differs from "desire" in having the suggestion of being innate and primitive, as "instinct" is. It suggests a motivational force within the agent that would operate even in the absence of outside stimuli to produce it, like hunger or the sex drive. None of my claims will turn upon the precise meaning of the term – the ability of our motivational states to direct our attention towards things we associate with our goals is what I am trying to point out. I thank Kathleen Higgins and Irene Price for help on this linguistic point.

occupy the foreground and be focused on in deliberation. The kinds of cases Nietzsche discusses provide examples in support of this view.

When one tries to fit the Nietzschean framework outlined above to the phenomenology of deliberation, one notices that the desire that directs deliberation and drives us towards our conclusions has substantial power to determine how we focus our attention. While directing attention is sometimes taken to be a more intellectual process than such rough operations as motivating action and causing pleasure, Nietzsche holds that desire has this kind of power over the intellect. Someone who desires to combat the vehemence of (for example) his sexual lust will focus his attention on things he associates, either positively or negatively, with the object of this second-order desire – for instance, his lust, past moments when he gave in to his lust, the future state of affairs in which he hopes that his lust will be controlled, and methods for controlling it. While other desires can be objects on which we reflect, the desire from the standpoint of which we reflect has the power to direct our attention. It causes us to pay attention to the object of desire and things we associate with it. As Hume put it, desire "casts our view on every side" of its object when we engage in practical deliberation (2.3.2).

The phenomenon of desire directing one's attention is familiar from cases far simpler than the one Nietzsche describes. Consider a simple case where only one desire is in play – perhaps, a case where a hungry person walks into his kitchen and sees a coconut cream pie sitting on the table. His attention will focus on the pie, and not the pattern of the linoleum or the hum of the refrigerator. Aspects of the pie that satisfy his gustatory desires will be the most central objects of his attention – perhaps, the toasted coconut flakes and the creaminess of the cream. The way that one desire focuses our attention on our other desires in the more complex cases of internal conflict is nothing more than an extension of this simple phenomenon.

One effect of Nietzsche's move in the dialectic between Humeans and their opponents is much like the effect of Hume's discussion of calm passions. Anti-Humeans might regard the activities that one engages in when one combats an intense desire as the outputs of a motivational state other than desire. After all, the desire that is the object of reflection feels different from the desire from the standpoint of which we reflect. Since the object of reflection is clearly a desire, and it feels different from the desire that occupies the standpoint of reflection, anti-Humeans might argue that we should regard the latter as a mental state other than desire. Nietzsche claims that despite the different roles of the two mental states within our phenomenology, both are in fact desires. On Nietzsche's view, it is the greater strength of one desire that causes it to seize the standpoint of reflection and hold it against the other.

The picture Nietzsche offers us is one on which complex processes of practical deliberation can be undertaken – as I have said – from the standpoint of a desire. Desiring that D – whether D is that one eat a piece of pie, that one not give in to one's sexual lust, or that one's suffering not be meaningless – causes the direction of attention towards D and towards things that are associated with it. Contemporary neuroscientists like Antonio Damasio have (though in terms quite different from Nietzsche) defended the view that desire has the power to direct the attention of agents in practical deliberation. Damasio hypothesizes that "a somatic state, negative or positive, caused by the appearance of a given representation, operates not only as a *marker for the value of what is represented, but also as a booster for continued working memory and attention*" (198). Somatic markers of value have the motivational and hedonic effects that desires are often regarded as having, and Damasio here embraces the view that they cause us to attend to things they mark as opposed to things which remain unmarked.

We can summarize the views of Nietzsche and Damasio regarding desire and the direction of attention as follows:

> **The Attention-Direction Aspect**: Desiring that D will make an agent more likely to focus her attention on things she associates with D than things she does not associate with D.

Instead of making a strong claim that desire is necessary or sufficient to determine how the agent directs her attention, it is better to make the weak claim that it increases the focus on its objects. There are many other psychological causes of attention-direction. Sudden movements or sounds can attract one's attention, and by an intentional action one can direct one's attention to something that does not itself figure significantly in one's desires. One may desire to eat some food more than one desires anything one associates with a loud noise, but still be distracted by the loud noise and momentarily not focus attention on the food. "Things" is to be read broadly to include such things as counterfactual possibilities – desire can cause an agent to focus her attention on desirable or undesirable states of affairs as well as physical objects in her environment.

Is it a necessary truth about desire that it directs attention? As cases of flashing lights and loud noises show, directing one's attention towards certain things is not sufficient for having desires for those things. The question of necessity is harder. The relevant counterfactual situations are very difficult (at least for me) to get any imaginative grip on. What would it be like to be a creature that did not direct its attention towards the things it associated with its desires? In particular, how would this interact with desire's capacity to cause pleasure? Many of our everyday experiences of pleasure and displeasure arise from the natural way we direct our attention towards possible or actual states of affairs that we associate with our strong desires. One thing that contributes to the vividness of an imaginative or sensory representation of some state of affairs for us is the intensity with which we direct our attention towards it. The more I am absorbed in thought about some past success, the more pleasure I feel. Since more vivid imaginative and sensory experiences of the states of affairs that we desire create more intense hedonic experiences, it is hard to get a grip on what our inner lives would be like if desire did not direct attention.

Insofar as this is conceivable, though, it does not seem that the ability to direct attention is necessary for desire. We can test our intuitions with the following case:

> **The Absent-Minded**. While the Absent-Minded are sometimes capable of directing their attention towards states they are motivated to bring about, the systematic and automatic connection between motivation and attention-direction that exists in humans is absent in them. When something causes them to fully focus on states of affairs that they are motivated to bring about, their experiences are the same as ours. But in general, they move about and engage in actions while their attention focuses on unrelated matters.

I name them the Absent-Minded because imagining extreme absent-mindedness seems to be the easiest way to get a grip on what agents who did not direct their attention towards things associated with desired states of affairs would be like. When I think about how they feel in these cases, it seems to me that the Absent-Minded are desirers, despite their lack of humanlike attention-direction in other cases.

### Davidson's Pro-Attitudes and the Objects of Desire

According to Donald Davidson, an agent's having a reason for action consists in having a pair of mental states – a "pro attitude towards actions of a certain kind", and "believing that his action is of that kind" (685). On Davidson's view, "the primary reason for an action is its cause" (686). Davidson never suggests that his view about the mental states causally responsible for action applies to a narrower set of creatures than the set of all agents, so it is best to interpret him as making a conceptual point about the nature of action, and not merely as identifying the pair of mental states that happen to cause actions in human beings. But as the claim that all possible actions must be caused by desire-belief pairs implies that all human actions are caused by desire-belief pairs, Davidson's position implies a position on the issue that I am concerned with.

Davidson does not go into great detail in describing pro-attitudes, which comprise a very broad class of motivational dispositions. His brief remarks give good reason to think that pro-attitudes are mental states that can fit under some reading of the term "desire", and which are much like Hume's "passions." Under the heading of pro-attitudes are included "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind" (686). While there is debate about the mental states underlying moral views, aesthetic principles, and some of the other states Davidson describes, the first four states are clear examples of mental states that seem like Humean desires. Davidson positions himself more firmly in the Humean camp when he says that "It is not unnatural, in fact, to treat wanting as a genus including all pro-attitudes as species" (688).

However, "pro attitudes must not be taken for convictions, however temporary, that every action of a certain kind ought to be performed, is worth performing, or is, all things considered, desirable" (686). A conviction is a belief of whose truth one is convinced, and anti-Humeans might want to say that convictions whose content ties them to action in this way can cause actions without the presence of any desire. If Davidson claimed that pro-attitudes were convictions that actions of a certain kind are worth performing, he would have to accept that a pair of beliefs alone – such a conviction, and the belief that the possible action before us is of that kind – could motivate action. Then he would importantly disagree with Humeans about motivation, who say that desires are necessary if humans are to be motivated to act.

In two places I have cited above, Davidson claims that pro-attitudes are directed not towards states of affairs or ordinary physical objects, but towards "actions of a certain kind". This makes it easy to characterize the kind of belief that is involved in the desire-belief pair – it is a belief that the action to be performed is of that kind. As he says, *R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consists of a pro attitude of the agent toward actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property. (687)

Insofar as we want desires to explain motivation, this works well – a pro-attitude towards actions of some kind is enough to explain how we are motivated to act. But if pro-attitudes necessarily make some reference to action, it will be difficult to identify pro-attitudes with desires. While the content of a Davidsonian pro-attitude necessarily involves an action for the agent to perform, the content of a desire need not involve action. Sports fans strongly desire the success of their favorite teams, but it hard to see how performing any actions fits into the content of this desire. This desire is certainly capable of motivating action – football fans often expend much energy in making noisy disturbances when an opposing team has the ball, and do so with the intention of disrupting the other team – but it is not clear why we should specify its content in a way that makes reference to action. Even more commonly than this desire causes action, it causes experiences of delight, disappointment, anxiety, and relief as games and seasons unfold. These are quite simply explained if one takes the contents of desires to be states of affairs that need not involve action. If we regard the agent's attitude as focusing on the state of affairs where his team wins the game or the championship, and not on any action that he might perform, we can neatly explain many aspects of these emotions. Events that increase the perceived likelihood of states of affairs that the agent desires are accompanied by pleasant emotions like delight and relief, while events that decrease the perceived likelihood of these states of affairs are accompanied by unpleasant emotions like disappointment and anxiety.

If desires are for states of affairs that do not necessarily involve action, Davidson’s characterization of the kinds of beliefs that interact with them to cause action will not be right.  Beliefs that a particular action is of a particular type do not interact with desires for states of affairs in the right way.  The beliefs involved need to be means-ends beliefs according to which an agent can bring about a particular state of affairs by engaging in a particular action.  If desires are for a broad range of states of affairs, this will be the best way to explain the performance of an action by its inclusion in the mental contents of the agent.

I have argued that the focus on action that Davidson attributes to pro-attitudes is not a necessary part of the content of desires, wantings, and other mental states.  Should we say here that Davidson has mischaracterized pro-attitudes, or that his category of pro-attitudes is different from the category of desires?  Davidson does intend the category of pro-attitudes to include a long list of mental states beginning with desires and wantings. However, the proviso at the end of his list – that these mental states are to be considered pro-attitudes “in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind” – leaves open the possibility that Davidson might not regard the majority of our desires as pro-attitudes.  Perhaps the only desires that count as pro-attitudes are the ones that make reference to action, a category including many of our instrumental desires, which are derived from other desires and means-end beliefs.  Even if he thinks that wanting is a genus of which all pro-attitudes are species, he might not think that it is a genus of which only pro-attitudes are species. If this is the correct way to read Davidson, he avoids the criticism that his view mischaracterizes the content of desire. However, he has then given overly complex explanation where a simpler one could be offered. Davidson will explain many actions with reference to a desire, a means-end belief, a resulting pro-attitude towards a kind of action, and a further belief that an action is of that kind. On the account that I have offered, a desire and a means-end belief do all the explanatory work in these cases.

There is a third contender for the focus of desire – while we often speak of people desiring particular states of affairs like the Texas Longhorns winning the national championship, it is also common to speak of people desiring various physical objects.[10] We say of Jenny that she wants chocolate, wants a diamond necklace, wants a goldfish, and wants Orlando Bloom. These are concrete objects and not states of affairs. Why should we translate our talk of desiring these objects to talk of desiring particular states of affairs? These cases may even seem more basic than the cases in which we desire states of affairs – after all, desires for food, sex, and simple possessions, often thought of as particularly basic cases, often are described as desires for objects. But there are good reasons to translate this object-talk into states-of-affairs talk. If we stay with object-talk, we will fail to understand exactly what Jenny will try to obtain and be pleased by. Jenny will try to eat chocolate, wear a diamond necklace, keep a goldfish in her aquarium, and make love to Orlando Bloom. She will not try to wear chocolate, eat a diamond

[10] In “Against Propositionalism”, Michelle Montague considers the position that desires are for things of this kind, and not for states of affairs or anything similarly propositional.

necklace, make love to a goldfish, or keep Orlando Bloom in her aquarium. Thoughts of the former states of affairs will please her; thoughts of the latter will not. If we want desire to do useful work in explaining and predicting the behavior of agents, we will have to regard object-talk merely as a convenient shorthand for states-of-affairs talk. It leaves out the essential differences between the ways we wish to interact with the objects, while states-of-affairs talk specifies this clearly. (Part of why object-talk seems so natural to us may have to do with the way desire directs attention. If Orlando Bloom enters Jenny's environment, her desire will cause her to focus her attention on him and not other objects in the area, since he is the thing that is most powerfully associated with her desires. However, when she plans future courses of action, her mind will be directed more towards possible states of affairs where she makes love to Orlando Bloom than possibilities where she keeps him in her aquarium.)

The following formulation, then, seems to characterize the motivational aspect of desire, with the content of the desire being the state of affairs that the agent is trying to bring about.

> **The Motivational Aspect**: If an agent occurrently desires D, and she occurrently believes that she can bring about D by doing A, she will have a motivational tendency to do A. Her motivational tendency to do A will increase with the strengths of the desire and the belief. If at any time there is some action that she has the greatest motivational tendency to do, she will initiate that action.

The technical term "motivational tendency" is brought in here to deal with cases where agents have desire-belief pairs that point them towards options inconsistent with each other. In such cases, agents are motivated to both kinds of behavior, but only one source of motivation results in action. I am not using this term in a way that will neatly explain feelings of motivation that individuals experience, and I am not claiming that our use of the ordinary word "motivation" tracks the product of desire and belief. Both our feelings of motivation and the ordinary use of "motivation" depend on factors beyond the ones in the above formulation. Even if I strongly desire to eat, and believe that by killing my friend and eating him I could satisfy my hunger, I will not feel any motivation, even an overridden feeling of motivation, to kill and eat my friend. And I would not ordinarily be described as motivated to kill and eat my friend. Predictions about the phenomenology of motivation cannot be directly read off of the magnitude of the agent's motivational tendency, in the sense I use "motivation tendency" here. "Motivational tendency" just refers to a tendency to act, which is a function of desire and belief.

**Strawson and the Connection to Pleasure**

In *Mental Reality*, Galen Strawson argues against "neobehaviorism," claiming that one can understand mental states without understanding how they are related to behavior. Strawson does not hold that desires have no necessary connections to behavior, however:

> Any desire has the following property: it is necessarily true that there are beliefs with which the desire can combine in such a way as to give rise to, or constitute, a disposition to act or behave in some way. This is a conceptual truth, true even of desires to change the past and desires for logically impossible things. But if I am rightly sure that I could never do anything about satisfying any of my desires about the weather, or lack any conception of the possibility of doing anything to satisfy my desires, then I am not *now* disposed to act or behave in any way simply

> on account of the fact that I have certain desires. This is especially clear if I am also constitutionally incapable of any sort of action or behavior. (276-277)

Strawson's view, then, is that having a desire does not entail having a disposition to act. However, that one has a desire entails a counterfactual about how one would be disposed to act in circumstances where one had a particular sort of means-end belief. (For Strawson, the truth of such a counterfactual is not sufficient for the existence of a disposition.) Strawson claims that it is possible to have the concept of desire without knowing about the connection to motivation. While he still holds that there are necessary conceptual connections between desire and motivation, he argues that creatures which lack dispositions to act, in his sense of "disposition", can still have desires.

Why not say that having a desire entails having a disposition to act, and that the activation conditions for this disposition include having proper means-end beliefs and the ability to act? The activation conditions for a disposition can be complex, and perhaps they include everything that would be necessary for us to regard a desire as a disposition to act. Part of what drives Strawson away from this position is a general worry about dispositions. As he says, "everything is set to act or behave in certain ways *should certain conditions be fulfilled*. The table in front of me is set to go for a walk, for example, should certain conditions be fulfilled (a radical rearrangement of its subatomic particles). You just have to put enough into the set of conditions" (268). Strawson is worried that if we regard desires as dispositions to act, things that intuitively do not have desires will be regarded as having desires, since many of these things can be dramatically modified so as to generate an agent.

The solution to this problem is just to specify the activating conditions clearly and narrowly enough so that not everything will be regarded as having a disposition to act. Giving something means-end beliefs is a reasonable activating condition for desire. Rearranging its subatomic particles so that it becomes a person is not. Furthermore, one can prevent the ascription of desires to furniture by stipulating that only agents can have desires.

One example that Strawson offers to argue that dispositions to act are not necessary for desire involves a group of non-actual creatures called the Weather Watchers. The Weather Watchers have sensations, thoughts, emotions, beliefs, and desires. They are pleased when their desires are satisfied and disappointed when they are not. However, their desires have no power to move them to action, as they are not capable of behavior, and even lack behavioral dispositions. They are rooted to the ground, motionless, and pass their time observing the weather and other natural phenomena. Some of the phenomena they witness delight them, some frustrate them, and some make them wistful. Strawson's claim is that these creatures are metaphysically possible, and that their lack of motivational dispositions does not make it impossible for them to have desires.

Strawson does not directly say anything about what would happen if the Weather Watchers could somehow be given means-end beliefs according to which they could bring about some desired state of affairs by performing an action. (Such a means-end belief would be false, since they are totally unable to act.) But since he thinks that "it is necessarily true that there are beliefs with which the desire can combine in such a way as to give rise to, or constitute, a disposition to act or behave in some way," it seems that he would say that Weather Watchers would be disposed to act if they acquired these means-end beliefs. As long as this is the case, it seems right that the Weather Watchers are possible. Even if they do not act, they would act if they were modified in the right way, and they have the emotions that go along with desire. So despite their lack of any disposition to move, it does not seem impossible for them to be desiring beings.

Strawson's second example concerns creatures that are capable of behavior, but incapable of any sort of affective experience or emotion. He names these creatures the Aldebaranians, and describes them as follows:

> Consider a race of creatures—the Aldebaranians—that have beliefs, sensations, thoughts, and so on. They are not capable of any affect states at all, but they are capable of entering into states—call them 'M states'—given which, and given that they believe what they believe, they are regularly caused to move in certain ways, and so regularly engage in what looks like purposive behavior. M states, then, may be defined as motivating states that are functionally very similar to states that we normally think of as desire states. They are functionally similar in respect of the way in which they interact with a being's informational states to cause it to move in apparently goal-directed ways. Roughly speaking, specific M states, in combination with specific informational states, lead to specific movements. (281)

Strawson later says that "the Aldebaranians are in fact experiencing beings, though they lack any affect dispositions" (282). The question is whether the Aldebaranians' M states are, in fact, desires. Strawson seems to think that they are, though he never advances a firm opinion on the case. He claims that people's intuitions differ on the question of whether the M states are desires.

Strawson comes to the conclusion that understanding the connection between desire and affective experience is essential to having the concept of desire, while one can have the concept of desire without even having the concept of motivation or action. He imagines a Weather Watcher philosopher who lacks the concept of action, but has concepts of all the affective states tied to desire. He believes that this Weather Watcher philosopher could still possess the same concept of desire that we do. It seems, however, that Strawson thinks an Aldebaranian philosopher could not have the concept of desire. In Strawson's view

> the primary linkage of the notion of desire to a notion other than itself is not to the notion of action or behavior but rather to the notion of being pleased or happy or contented should something come about (or at least to the notion of ceasing to be unhappy or discontented should it come about) and to the distinct but correlative notion of being unhappy or discontented or disappointed should it not come about. (280)

Strawson does not clearly explain what a "primary linkage" is, and, it is hard to know exactly what he means. He seems to be claiming that the causal connection between desires and affective states is somehow more essential to our concept of desire than the causal connection linking desire to action or behavior. He also seems to hold that both are metaphysically necessary, which is puzzling, as metaphysical necessity does not seem to come in degrees. Perhaps even the necessary connection that Strawson accepts – that desire is necessarily able to combine with belief to cause a disposition to act – is being presented as less significant than the causal connection to affective states.

To make the examples of the Weather Watchers and the Aldebaranians more effective in getting at the issues Strawson is talking about, we should say a little more about the cases. So I will clarify one counterfactual feature of the Weather Watchers case, and set up the Aldebaranians case in a similar way. Suppose it is true of the Weather Watchers that if they could somehow be brought to the false belief that they could satisfy their desires through action, they would try to engage in action. And suppose it is true of the Aldebaranians that their lack of affect is due to the combination of a lack of imagination and unchangingly true gut-level beliefs about the future. They lack affect only because they do not have the imaginative capacity to allow for pleasant daydreams and moments of horror at imagined future calamities, and because they lack the changing beliefs that are necessary for delight at unexpected satisfactions and disappointment at unexpected failures. But if they could be given the capacity for imagination, or induced to have changing beliefs about the future, they could have the same affective experiences that humans do in these cases. Now there is intuitively no problem with saying that both creatures have desires, since the relevant counterfactuals obtain in each case. Just as Weather Watchers are not disposed to act due to their lack of means-end beliefs, in Strawson's narrow sense of disposition, Aldebaranians are not disposed to have affective experiences due to their lack of imagination and belief-change. But at least in the way Strawson uses "disposition," claims about the necessity of dispositions to act or feel emotions are stronger than claims about some necessary counterfactual tie to action or emotion, and the latter claims hold while the former do not.

As for whether philosophers of their kind who lacked certain important concepts – action for the Weather Watcher, and affect for the Aldebaranian – could share our concept of desire, it seems to me that both cases are on the same footing. They could successfully refer to desire, just as people successfully referred to water even before anybody knew about $H_2O$. By use of a reference-fixing description, Weather Watchers, Aldebaranians, and humans all could refer to the same psychological property, even if they had different beliefs about that property. (This does not presuppose that “desire” has the same sort of natural-kind semantics as “water” does. Number terms are usually not taken to have the same semantics as “water”, but one could use reference-fixing descriptions to pick out numbers too.) Weather Watchers and Aldebaranians could not successfully give a conceptual analysis of desire, but the ability to give a successful analysis is not necessary for concept possession.

What if we had set up the cases the other way? What if even the addition of an appropriate means-end belief could not move the Weather Watchers to action, and neither imagination of alternative possibilities nor changing beliefs about the objects of their M states could cause the Aldebaranians to feel any hedonic emotions? In this case, it seems to me that the Weather Watchers would be impossible (since they are stipulated to have desires, but they cannot be motivated even if the appropriate means-end beliefs are somehow forced into their heads). Similarly, since the Aldebaranians are incapable of any affective states even when the appropriate counterfactuals hold, it seems to me that their motivational states are not actually desires, but some other kind of mental state. Weather Watcher and Aldebaranian philosophers would then be referring to something different than we do, and than each other do, when they consider the mental states of their kind.

## Smith’s Objections to Phenomenological Theories and his Dispositional Account

In *The Moral Problem*, Michael Smith presents some objections to phenomenological theories of desire. The particular phenomenological view that he devotes the most time to is the "strong phenomenological conception" of desire.

> **Strong Phenomenological Conception (SPC)**: "Desires are, like sensations, simply and essentially states that have a certain phenomenological content." (105)

Smith thinks that some of his objections to strong phenomenological conception will apply to varieties of the weaker phenomenological conception as well.

> **Weaker Phenomenological Conception (WPC)**: "Desires are like sensations in that they have phenomenological content essentially, but differ from sensations in that they have propositional content as well." (108)

The conception of desire that I will defend in this chapter will not be a version of SPC. But depending on what it means to "have phenomenological content essentially," it may count as a version of WPC. If having phenomenological content essentially means that the mental state itself is essentially constituted in part by some phenomenological state, then I will not be defending a version of WPC. But if having phenomenological content essentially means that the mental state necessarily can produce a certain phenomenology, under some activating condition, then my view is a variety of WPC. In what follows, I will interpret Smith as meaning the latter thing, and arguing against views where some phenomenological states are necessarily connected to desire. Every argument he presents that would apply to phenomenological conceptions under the former reading applies to them simply because they are phenomenological conceptions under the latter reading. So even if I am misreading his conclusions, I will be addressing the key parts of his arguments.

I will now consider Smith's objections to both SPC and WPC. While there is no reason to reject the intuitions that underlie some of his arguments, these intuitions can be accommodated by a theory according to which it is necessary for a mental state's being a desire that it generate particular phenomenological effects under particular activating conditions.

Smith's first argument against the strong phenomenological conception involves agents with unconscious desires. He begins this argument by claiming that "it is plausible to hold that a subject is in pain if and only if she believes that she is in pain," and holding this to be an instance of the more general thesis that "a subject is in a state with a certain phenomenological content if and only if she believes herself to be in a state with that content." If this general claim about the epistemology of phenomenological content is true and the strong phenomenological conception is correct, "it is plausible to hold that a subject desires to φ if and only if she believes that she desires to φ" (105).

The general thesis about phenomenology is not as plausible as Smith thinks – it is a familiar phenomenon from ordinary experience that the content of our phenomenological states often outstrips our beliefs about our phenomenological states. Looking out the window as a car passes by, I may have an experience of blackness because of the car's black tires, but form no belief that I am seeing black, because I am not paying any particular attention to the tires. In the specific case of pain, agents who are focused on things other than their pain (people in life-or-death situations and athletes engrossed in intense competition, for example) may not form beliefs about their pain, because their attention is trained on other things. In all of these cases, it is likely that the people involved would form beliefs about being in pain if these beliefs became relevant enough to their situation to warrant the focusing of attention on states of the world that often precedes belief-formation. But there is no reason to ascribe beliefs about their sensations to them antecedently. Similarly, even if having a desire is nothing more than having a particular phenomenological content, there is no reason to think that all those who desire that P will believe that they desire that P.

Smith follows the aforementioned discussion of the phenomenology of reflection with the following example, in which an agent acts out of an unconscious desire:

> Suppose each day on his way to work John buys a newspaper at a certain newspaper stand. However, he has to go out of his way to do so, and for no apparently good reason. The newspaper he buys is on sale at other newspaper stands on his direct route to work, there is no difference in the price or condition of the newspapers bought at the two stands, and so on. There is, however, the following difference. Behind the counter of the stand where John buys his newspaper, there are mirrors so placed that anyone who buys a newspaper there cannot help but look at himself. Let's suppose, however, that if it were suggested to John that the reason he buys his newspaper at that stand is that he wants to look at his own reflection, he would vehemently deny it. And it wouldn't seem to John as if he were concealing anything in doing so. However, finally, let's suppose that if the mirrors were removed from the stand, his preference for that stand would disappear. (106)

Smith takes this to be a case where John desires to see his own reflection, but does not believe that he desires to see his own reflection. As I have just argued, this should not in itself be a problem for SPC. Cases where agents have particular phenomenological contents, but lack beliefs that they have those contents, are common enough.

What would be a problem for SPC (and WPC, for that matter) is if John's desire to see his own reflection never had any phenomenological effects. And it is not clear, in the above example, that John's desire would be this ineffectual. It is plausible that in the situation where the mirrors were removed from the stand, John would look where the mirrors used to be, fail to see his reflection, and suddenly feel a twinge of dissatisfaction that he could not correctly explain. Perhaps he would explain his dissatisfaction in some other way, or perhaps he would not even think about it enough to come up with an explanation at all. If this is how things would actually go, SPC can handle Smith's proposed counterexample, and if it is plausible that this is how things would go, Smith's counterexample should not persuade us to abandon phenomenological theories of desire.

Now I will move on to Smith's second argument against SPC:

> Let's suppose we grant that desires are like sensations in that they essentially have phenomenological content. Even so, it must be agreed that they differ from sensations in that they have, in addition, propositional content (Platts, 1981: 74-7). Ascriptions of desires, unlike ascriptions of sensations, may be given in the form 'A desires that p', where 'p' is a sentence. Thus, whereas A's desire to φ may be ascribed to A in the form 'A desires that she φs', A's pain cannot be ascribed to A in the form 'A pains that p'.

Smith here assumes that sensations lack propositional content. This is not an uncontroversial claim. Philosophers like Michael Tye have held that sensations do in fact have propositional content. On a view like Tye's, the visual sensation of blue has the propositional content that there is a blue object in front of one, and the sensation of pain has the propositional content that I have sustained some sort of bodily damage.

Could Smith describe the contrast between sensations and desires in a way that is compatible with sensations having propositional content? Perhaps he could say that sensations and desires differ in that the propositional content of sensations, if they have any, is limited to their phenomenal content, while desires have a propositional content that goes beyond their phenomenal content. He could say that this difference is reflected in the difference between the way we talk about pains and the way we talk about desires, with that-clauses introducing the propositional content of desires while no that-clauses can introduce the propositional content of pains. (While people are sometimes pained that P, this is not generally true of pains. I can be pained that my friend misunderstood my motives, but the "pained that" locution is not applicable to purely physical pains that are not associated with any propositional attitudes – for instance, the pain of being stung by a bee.) If it seems strange that some propositional contents cannot be introduced by that-clauses while others can, that is really a problem for Tye and not for Smith.

In fact, there is an important difference between pains and desires that underlies the way we talk about them. For a mental state to be a desire, it is essential that it be able to cause action in combination with a suitable means-end belief. This distinguishes desires from pains and other mental states whose essential nature is limited to their phenomenological content, and which need not cause action. The propositional content of a desire explains what kinds of actions it will cause, and without such a content it would be impossible to explain the causation of action. An agent who desires that D will do A when she realizes that she can bring about D by doing A. This is why it is significant that we talk about a desire that D, while not talking about a pain that P. Smith, then, is right to reject SPC. On a strong phenomenological conception, desires lack a kind of content that is essential to their nature as mental states that cause action.

Smith then tries to parlay this objection to SPC into an objection to WPC. He charges that versions of WPC "in no way contribute to our understanding of what a desire as a state with propositional content is, for they cannot explain how it is that desires have propositional content" (108). Supplementing Smith's argument with the above reflections about the ability to cause action being essential for desire, we can develop out of this remark a good argument against versions of WPC where phenomenological content alone is essential to desire, and propositional content is merely a nonessential characteristic of actual desires. While such theories might be able to account for the propositional content of actual desires (since it is contingently true on these theories that desires have propositional content), they will not be able to account for the necessity of propositional content for desires. Since the motivational feature is essential to desire, and propositional content is necessary to account for the motivational feature, these versions of WPC must be rejected.

But there is a version of WPC that is not touched by this part of Smith's objection. In this version, both propositional content and counterfactual connections to phenomenological content are essential for desire. The view that was suggested at the end of the discussion of Strawson and the Weather Watchers, on which both connections to action and connections to pleasure are necessary for desire, is such a species of WPC. So Smith offers a series of counterexamples against the necessity of phenomenological content:

> Consider, for instance, what we should ordinarily think of as a long-term desire: say, a father's desire that his children do well. A father may actually feel the prick of this desire from time to time, in moments of reflection on their

> vulnerability, say. But such occasions are not the norm. Yet we certainly wouldn't ordinarily think that he loses this desire during those periods when he lacks such feelings. Or consider more mundane cases in which, as we would ordinarily say, I desire to write something down and so write it down. As Stroud points out, in such cases 'it is difficult to believe that I am overcome with emotion… I am certainly not aware of any emotion or passion impelling me to act' ; rather 'they seem to me the very model of cool, dispassionate action' (Stroud, 1977, 163)" (109)

How should a phenomenological theorist respond to these counterexamples? Certainly, having a desire at some time does not by itself require that one have any particular experiences at that time. But this does not defeat all phenomenological theories. Perhaps desires necessarily cause experiences, but only under particular circumstances. This would be much like the way that they only cause action when activated by an appropriate means-end belief.

The distinction between occurrent and latent desires provides one part of the explanation of why we do not always feel our desires. Even our strongest desires do not register in our experience at times when we are focusing our attention on unrelated things. At any given time, most of our desires are latent, and for a desire to affect our experience, something must happen to make it occurrent. For instance, we may have to direct our attention towards something we associate with the object of desire. But this distinction cannot do all of the work. Sometimes, even as we engage in action motivated by a desire, we have no emotions or other internal experiences connected with it. Smith's example of writing something down seems to be such a case. Since the desire is occurrent in this case and succeeds in motivating action, we will need more than the latent/occurrent distinction to defend WPC.

A good phenomenological theory, then, would not make having any particular experience a necessary condition even of having an occurrent desire, by itself. Instead, it might say that an agent with a desire must have particular experiences when some activating condition is present. This is true to the phenomenology of desire. Our desires do not always impact our phenomenological states, but they can do so under particular conditions. One important feature of desire is the systematic way that desires cause us to feel pleasure and displeasure as we have belief changes and sensory or imaginative representations dealing with the desired states of affairs.

> **The Hedonic Aspect**: If an agent occurrently desires D, increases in the subjective probability of D or vivid sensory or imaginative representations of D will cause her 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.

I characterize the hedonic aspect only in terms of rough proportionality because some creatures' levels of pleasure might not increase continuously, but in slight and discrete levels. This is not a reason to say that these creatures lack desires. Whether or not humans have desires should not turn on the question of whether their desires increase in strength discretely or continuously. My formulation also leaves room for the possibility that agents with very weak desires might not feel any pleasure upon imagining their satisfaction.

A phenomenological conception incorporating this hedonic feature would deal quite well with the examples Smith presents. In the case of the father, Smith brings up one particular phenomenological effect of desire – the way the father would feel "the prick of this desire" as he reflects on his children's vulnerability. If reflecting on his children's vulnerability involves having an imaginative representation of the bad things that could happen to them, the account suggested above would be able to explain why desire has phenomenological effects in this situation, but not in other situations. And while there usually is no substantial pleasure or displeasure in the act of writing down a trivial note, there is often some mild frustration in trying to jot down a note and suddenly discovering that your only pen is out of ink. When this happens, one's subjective probability of satisfaction declines sharply in a moment, and one feels the attendant displeasure of frustration.

So far, I have dealt with potential counterexamples to the hedonic aspect that are located in the actual world. But if the hedonic aspect is essential to desire, it has to manifest itself in all possible cases. To test for whether the hedonic aspect is necessary, we should consider a case where agents have all other components of desire, but lack the hedonic aspect, and see if we intuitively regard them as desirers:

> **The Neutrals**: The Neutrals are intelligent creatures who are exactly like us, except that their motivational states have no connections to pleasure or displeasure, even under conditions of belief-change or vivid imagination. Like a human, a Neutral would move quickly towards his baby if he saw that the baby was about to crawl into a busy street. But while a human father might have an

> unpleasant experience of fear just as he began to move, a Neutral would not. Though the Neutral's attention would be intensely focused on the baby as he began to move and he would have the same visual and auditory experiences that we do, he would feel nothing unpleasant at all. Even if, in the future, he imagined what could have happened if he had not seen the child in time, he would not feel the unpleasantness of horror in imagining. To an observer, Neutrals are indistinguishable from normal human beings. When you do things to one of them that would make a person laugh or cry, they show the outward behaviors of laughter and crying. But they do not feel the pleasure of laughter or the displeasure of sadness that we usually do when crying.

This case is similar to Strawson's case of the Aldebaranians, since experiences of pleasure and displeasure play a significant role in making our affective mental states what they are. Having offered this example to many people, both philosophers and nonphilosophers, I would have to say that intuitions on the matter of whether the Neutrals have desires differ quite widely, much as Strawson says that intuitions about whether the Aldebaranians have desires differ widely. Many respondents were conflicted on the issue.

I often would ask another question about the case of the Neutrals: do any of their movements count as actions? Without exception, people said that the Neutrals were engaging in action, and no one seemed conflicted on the issue. The difference in the intensity of the responses – certainty that the Neutrals are acting, but lack of consensus about whether they have desires – suggests that there is something more to the concept of desire than the motivational aspect. If it were a conceptual truth that the ability to cause action was necessary and sufficient for a mental state to be a desire, one would expect intuitions about whether the Neutrals act and whether they have desires to go together. The fact that these intuitions do not go together suggests that there is more to the concept of desire than the motivational aspect, and that the Humean theory is not a necessary truth.

A phenomenological conception incorporating the hedonic aspect would also address one of Smith's concerns in this section – developing a plausible epistemology of the propositional content of desire. One way to figure out what we desire is to imagine various counterfactual situations and see how they please us. Seeing how we feel as new information affects the subjective probabilities of various potentially desirable events is also a way of determining what we desire. We are fallible as we do this – we can be misled, among other things, by ill-supported prior beliefs about what we desire, and by wishful thinking about what we want and do not want to desire. Either of these factors may cause us to attend badly to our experiences, and ignore or misinterpret the data that our internal impressions give us.

### Smith's Dispositional Account

I will now consider Smith's own account of the essential features of desire. Smith presents a dispositional account on which the essential property of desire is its direction of fit. He cashes out "direction of fit" as follows: "a belief that p tends to go out of existence in the presence of a perception with the content that not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p" (115). I will set aside the question of whether his characterization of belief is correct, and consider whether he has correctly characterized the essential properties of desire. Is it necessary for a mental state's being a desire that p, that it tend to endure in the presence of a perception that not p? And is this sufficient for a mental state's being a desire that p? I will explore this question via two thought experiments. First, one that will address the necessity of Smith's tendency claim, involving the Fainthearted; and second, one that will address sufficiency, involving the Imaginers.

> **The Fainthearted:** Much like us, the Fainthearted have motivational states directed at many possible states of affairs, which cause them to act when paired with appropriate means-end beliefs. These motivational states are just like our desires in terms of their functional characterization, with one exception: As soon as they have a perception that not-p, their motivations to bring about p vanish with the same swiftness that our beliefs that p vanish in the face of perceptions that not-p. This inhibits their action in many cases where we might act. They still sometimes act, since their past perceptions may not give them any idea whether p or not-p, but they are unable to act after perceiving that not-p. And before perceiving that not-p, they do all the things that we would expect of someone who desired that p – for example, directing their attention towards things they associate with p and feeling pleased when they vividly imagine that p.

It seems obvious to me that the Fainthearted have desires. It is not essential for a mental state's being a desire that it stay in existence even in the presence of a perception that not-p, or any other perception. While Smith's dispositional account may be successful in distinguishing between beliefs and desires, that seems to be because it picks up on a plausible candidate for being an essential property of belief – responsiveness to evidence – that does not correspond to any essential property of desire. Since the Fainthearted have desires, the ability to stay in existence when the agent perceives that not-p is not necessary for a mental state to be a desire.

Now I will consider whether Smith's dispositional account presents us with a sufficient condition for a mental state being a desire.

> **The Imaginers**: The Imaginers spend much of their time imagining situations, both actual and counterfactual. Whenever it happens that they are imagining some situation, and they suddenly perceive that that situation does not obtain, they continue imagining that situation for a long time. But whenever they perceive that a situation they are imagining does obtain, they cease to imagine it. They differ from us in that they never engage in action, no matter what means-end beliefs they form. They also never have pleasure or displeasure of any kind.

Clearly, the Imaginers' imaginative mental states do not count as desires. But if the dispositional account had correctly provided a sufficient condition for a mental state's being a desire, the persistence of these mental states when one perceives the opposite of their contents would mark them as instances of desire. (I doubt that the view that persistence under the perception that not-p is sufficient for a desire that p should really be attributed to Smith. It is a deeply implausible position. But it is not clear that he rejects the position, and so it seemed worthwhile to demonstrate that the position is implausible.)

Smith precedes the presentation of his criterion with "Very roughly, and simplifying somewhat…" so perhaps there is a less rough and less simplified version of his criterion that would escape the objections that have been offered here (115). But it is hard to see how the criterion could be fruitfully modified. What is essential to desire, it would seem, is how it moves us and how it makes us feel, not its persistence under some set of perceptual conditions. Perhaps the way that Smith sets up the dialectic in section 4.6 is, in part, to blame for these problems. A large part of Smith's concern in this section is to distinguish desires from beliefs, and to eliminate the possibility of 'besires' that are both desires and beliefs at once. While his criterion manages to close off this possibility by defining desires in a way that makes them necessarily different from beliefs, it does so more by pointing to essential features of belief than by picking up on essential features of desire.

Two concepts that Smith notably does not invoke in his account of desire are pleasure and action. I take it to be one of the morals of the failure of Smith's account that the necessary conditions for desire must include strong functional connection to one or both of these. In the next section I will consider another account of the necessary conditions for desire that does not make reference to pleasure and action.

## Schroeder's Reward Theory

In *Three Faces of Desire*, Tim Schroeder presents his Reward Theory of Desire:

> **Reward Theory of Desire (RTD):** To have an intrinsic (positive) desire that *P* is to use the capacity to perceptually or cognitively represent that *P* to constitute *P*

> as a reward. To be averse to it being the case that *P* is to use the capacity to perceptually or cognitively represent that *P* to constitute *P* as a punishment. (131)

This view is expressed in an abbreviated fashion later on: “To be a desire is to be a representational capacity contributing to a reward or punishment signal” (168). So what is it to constitute something as a reward or as a punishment? This question is answered by the Contingency-based Learning Theory of Reward.

> **Contingency-based Learning Theory of Reward (CLT):** For an event to be a reward for an organism is for representations of that event to tend to contribute to the production of a reinforcement signal in the organism, in the sense made clear by computational theories of what is called ‘reinforcement learning’ (66).

The basic idea behind reinforcement learning was expressed by Edward Thorndike in 1911:

> Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond. (244)

Reinforcement learning, then, is a process by which behavioral and other psychological dispositions can be acquired, strengthened, or weakened. In many cases, these are dispositions connected to action. For example, a laboratory where I once worked had a metal doorknob that would often give me a mild shock from static electricity when I touched it. I soon developed a disposition to not touch the doorknob. It was very hard to make myself touch it, except through the fabric of my shirt. That I could form this disposition makes it the case that I had an aversion to being shocked – in the language of

RTD, I represented electric shocks as punishments. In the language of CLT, I received a punishment signal whenever I touched the doorknob. Having a desire to not be shocked is simply a matter of having these punishment signals, which changed my behavioral dispositions. A similar thing happens when I acquire the disposition to flick the light switch upon entering my room in the dark. As soon as I do this, I am rewarded with a lighted room. After numerous opportunities for reward, I automatically raise my arm to flick the light switch upon entering the room – I have acquired firm mental dispositions to turn the lights on. I may even do this, habitually, if I know that the power is out and the lights will not turn on. That I represent a lighted room in a way that allows for this kind of reinforcement learning makes it true that I have a desire for illumination.

The learned mental dispositions need not be so directly connected to action. In addition to helping people "learn certain sorts of habits," the process of reinforcement learning can cause "certain sorts of modifications to their sensory capacities" (168). The various patterns of dots that make up Braille letters feel the same to those who have not learned the language. When I pass my finger over the Braille letters in an elevator, it is hard for me to get any precise tactile sensation representing the way that the dots are arranged. It strikes me as surprising that anyone can get such a sensation by touching them – as one must in order to understand the dots as a word. Blind people who have learned Braille, however, can do so. This is because correct sensory discrimination has been repeatedly rewarded in the process of learning Braille. The desire to understand a particular word was satisfied, and perhaps a teacher praised the blind person. So the mental dispositions required for fine-grained perception of Braille letters were reinforced.

The neuroscience of reinforcement learning involves the brain's reward center – according to Schroeder, the VTA/SNpc – changing neural connection strengths. The neural connections that are involved in realizing particular mental dispositions are strengthened when their activation is followed by VTA/SNpc stimulation, which causes the VTA/SNpc to release dopamine.

While what makes something a reward system is its contribution to reinforcement learning, it is contingently true that the human reward system is causally responsible for more than this. Schroeder adduces psychological and neurobiological evidence to suggest that "The neural basis for reward is the normal cause of pleasure and an important cause of motivation" (37). Schroeder presents evidence that "Most euphorigenic drugs directly or indirectly stimulate the VTA/SNpc" and the effects of this stimulation are responsible for pleasure (92). There are direct neural connections by which VTA/SNpc stimulation reaches the PGAC, where (according to Schroeder) pleasure is realized. As for motivation, sufferers of Parkinson disease "lose a very large percentage of the dopamine-producing cells in the SNpc," and this can in extreme cases make them completely unable to move (118, Berridge and Robinson 1998; Langston and Palfreman 1995). Destroying the dopamine-releasing cells projecting from the VTA/SNpc to the motor prefrontal cortex, the home of immediate prior intention, destroys monkeys' ability to keep a prior intention in mind long enough to execute it after a delay (116).

Schroeder does not present RTD as an analysis of desire, but as the necessary and sufficient condition for an agent's having a desire. According to Schroeder, "desire" is a natural kind term, and necessary truths about desires can be discovered *a posteriori* in much the same way that the water/$H_2O$ identity was.  RTD expresses a metaphysically necessary connection between desire and reward, but not a logically necessary one.  This is good, since RTD would be quite implausible as a conceptual analysis of desire.  That desires are linked to the strengthening of mental dispositions is, as Schroeder recognizes, a "commonsense hunch" (51) that happens to be proven true by empirical evidence, and not something essential to the concept of desire.

Even as a claim about how desires are necessarily realized, however, RTD is unsatisfactory.  The realizer that Schroeder has offered fails to satisfy the descriptive component of our concept of desire.  To see this, we should consider the way that speakers' intuitions about natural kind terms will go when they are offered further scientific information, and then consult our intuitions about what would count as a desire.  If "water" really is a natural kind term, speakers who do not yet know that its molecular formula is $H_2O$ should still conditionally assent when given Putnam's Twin Earth thought experiment.  They should accept that if it turns out that the drinkable stuff in rivers and rain has the molecular constitution $H_2O$, and this explains the properties by which we recognize it, the XYZ on Twin Earth is not water.  Likewise, speakers who do not yet believe that desires are actually realized by connections to reward signals should conditionally assent to the proposition that creatures whose mental dispositions are not susceptible to reinforcement learning would not have desires, on the condition that the mental phenomena we associate with desire are actually explained by the reward signals of reinforcement learning.

So we should consider some cases of this kind. The first case will test whether a reward system is necessary for a mental state's being a desire. The second case will test whether it is sufficient.

> **The Unconditionable**: Suppose, as Schroeder's evidence suggests, that there is a mental state of representing something as a reward, which has the power to change the mental dispositions of actual humans, and which also happens to explain motivation and our experiences of pleasure. Suppose further that there are creatures on some other planet who possess motivational dispositions much like ours, but lack the capacity for reinforcement learning. Perhaps these creatures – call them the Unconditionable – are born fully formed and able to do all the things that we humans require a reward system to learn. In particular cases, their phenomenology and behavior is identical to ours, minus whatever immediate phenomenological and behavioral effects necessarily require a reward system. The Unconditionable have a particular conative mental state which plays a major role in explaining their behavior. When they believe that doing A will bring about B, and they have this conative state towards B, they do A. If they fail to attain B, they feel the same emotions of disappointment and frustration that we do. But if they succeed, they feel as excited and happy as we would.

Does their lack of a reinforcement learning system make the conative state that motivates their actions not count as a desire? Ordinary intuition says that the Unconditionable have desires. A reinforcement learning system is not necessary for desire, even if it explains actual motivation and pleasure.

A second example will be helpful in considering the claim that a reward system is sufficient for desiring.

> **The Creatures of Habit**: Assume, as before, that humans have a mental state of representing something as a reward, which has the power to change the mental dispositions of actual humans, and which also happens to explain motivation and our experiences of pleasure. The Creatures of Habit, whose mental lives are largely unlike ours, have the same kinds of reward systems that we do. They regularly engage in habitual, unintentional tics. They have representational states picking out certain states of affairs, and when they happen to tic before one of these states is produced, ticcing behavior of that kind is reinforced. Their reward and punishment systems strengthen and weaken neural connections contributing to these behaviors and other things, changing their mental dispositions just as our reward systems change our mental dispositions. But when they represent B as a reward and believe that they can bring about B by getting A, they never do A (unless A coincidentally happens to be a tic that they have learned). Intentional action that is motivated by a combination of belief and some conative state is entirely foreign to them.

Do the Creatures of Habit have desires, in virtue of the fact that they represent certain states of affairs as rewards? Ordinary intuition says no, rejecting the sufficiency claim. Creatures can have reward systems that reinforce some kinds of behavior, but still not have desires.

It appears that Schroeder's theory gives counterintuitive results in both cases. In the place where he tries to argue that RTD is consistent with common sense, he seems to be unaware that his theory would give the wrong answers about whether certain non-actual creatures have desires. The claim that "To be a desire is to be a representational capacity contributing to a certain mathematically describable form of learning" (168), he says, does not require us to reject any of our prior convictions about desire, and thus could be part of a proper realist account of desire. This claim, however, requires us to reject prior convictions about the possibility of desires in the absence of this system of learning, and the insufficiency of this kind of learning for making it the case that a creature has desires. While Schroeder accepts early on that "Treating 'desires' as a natural kind is not a license to drag up any old entity from the back pages of some journal of neuroscience and proclaim it a desire," (9) this is what he seems to have done.

Some of the desiderata that brought Schroeder to accept a theory on which reward systems are necessary realizers for desires are based on bad arguments. He rejects the "Standard theory of desire," on which the power to motivate is necessary for desire and the "Hedonic theory of desire," on which some relation to pleasure is necessary for desire. Instead he accepts RTD, on which desire is necessarily connected to reward. Schroeder regards it as a virtue that a theory presents desire's connections to pleasure and motivation as contingent (albeit firmly grounded in the psychology of actual humans). I will consider his arguments for preferring theories that make these relations contingent, show why these arguments are not good, and then draw two morals from the story. The first moral will be that connections to pleasure and motivation are plausible as necessary conditions for a mental state's being a desire. The second moral will be that the semantics of "desire" differ significantly from the semantics of "water."[11]

First I will consider how Schroeder argues that the ability to motivate action is not necessary for desire. He considers the case of an ancient Greek mathematician who desires that π be a rational number. If we understand desires to merely be dispositions to make the world match their contents, we will not be able to admit the possibility of desires with impossible objects like this one. Since the desired state is impossible, a disposition to have the world match it is impossible. There is nothing it could be to have such a disposition activated. However, Schroeder claims that the desire attributed to the Greek mathematician is a possible desire. This seems intuitively right – we can imagine the mathematician being alternately excited and frustrated as his beliefs about the rationality of π change, and it seems right to attribute a desire for π to be rational to him on this basis. Schroeder then considers the formulation ST2 as a way of capturing the way that the ability to cause motivation is necessary for desire:

> **ST2**: To desire that P is to be so disposed that, if one were to believe that taking action A would be an effective method for bringing it about that P, then one would take A" (17).

To turn this into a claim only about necessity, the issue that Schroeder is ostensibly dealing with in this section, "is to be so disposed" should be replaced with "requires that one be so disposed." This successfully responds to the concern about the Greek

[11] As it turns out, I do not share Putnam's intuition – there are possible non-H2O substances which I regard as water. But since this is the standard example of a necessary *a posteriori* identity, I will continue to use it.

mathematician. But Schroeder sees another problem with it. He considers a case where someone – I will call him the Noninterferer – desires that a committee decide in his favor on some issue without his intervention. This seems like a perfectly possible thing for someone to desire. However, "because of the very nature of the desire it makes no sense to try to act so as to satisfy it" (17). An action performed to satisfy the desire could have no effect but to make its satisfaction impossible.

Schroeder seems to regard the possibility of this desire as a problem for the claim that motivational efficacy is necessary for desire in the fashion described by ST2. But it is not clear why this should be. An agent who realized that acting on such a desire was fruitless would not act, and an (irrational) agent who believed that he could achieve the end by doing something to affect the committee's judgment would act. All of this is just as ST2 predicts. So what is going on here? Schroeder goes on about how it would take a fairly drastic level of irrationality for the Noninterferer to act, and he seems to think this is a problem for ST2. But this should not be regarded as a problem – ST2 can be true even if some desires have contents that make it impossible for them to be acted on.

Interestingly, Schroeder has given an example that could cause trouble for the sufficiency of the kind of connection to motivation discussed in ST2. Cases similar to that of the Noninterferer provide arguments against satisfaction of the following condition, ST2X, being sufficient for desire:

**ST2X**: To desire B, it is sufficient that one do A whenever one believes that one can bring about B by doing A.

ST2X would license the attribution of the following desire as well to the Noninterferer: that the committee act to execute him, without his intervention. This would trivially satisfy the sufficiency condition, since a rational agent will never believe that he can bring about the satisfaction of his desire by any action. So ST2X implies that every agent has all sorts of wacky desires whose contents prevent their activation. However, it is no argument against the necessity of motivation under all belief-activated conditions for desire. Given the placement of the example in the text – under the "Motivation not essential for desire" section heading and before the "Motivation not sufficient for desire" section heading, it appears that Schroeder intends the example to do some work that it is not capable of doing. In any case, he does not offer a successful argument that motivation is not essential for desire.

Now I will consider the supposedly contingent connection between desire and pleasure. Schroeder offers several arguments that this connection is contingent. I will consider the two that I regard as the most powerful and give reasons for rejecting them.

Schroeder's first argument deals with cases of depression.

> Consider a man who has just had a number of powerfully negative life experiences, say, the death of both parents and the loss of a meaningful occupation, and who as a result has become depressed. This man once took great pleasure in many things, including his wife's successes, but now is only slightly pleased by these things. Need we hold that he cares less about his wife now than before? That he has fewer, or weaker desires for her success? Normally, this is not held to be the case. (31)

Cases of depression are troublesome for many theories of desire. According to Schroeder, we want to say that depressed people still desire the same things that they desired when they were not depressed. (Not all people want to say this – some, though perhaps a minority, find it okay to say of depressed people that they simply do not have any desires. If this is actually the right thing to say, depression is far less effective as a counterexample to theories of desire.) But desire seems to lack its characteristic effects in the cases of depressed people. Not only are their experiences of pleasure dulled, but they are less likely to engage in action. People gripped by depression often sit in one place, unhappily, unable even to bring themselves to engage in actions that they believe would pull them out of their depression.

The distinction between latent and occurrent desires could be used to deal with the case of depression. This distinction is one that any theory of desire will have to make use of. In some sense, I can be said to desire that John Edwards become president, even if I am asleep or fully engaged in writing my dissertation. In these cases, my desire for Edwards' victory is latent, but not occurrent. I do not feel my desire, and it does nothing to direct the course of my thought or behavior. Something has to happen to make the desire occurrent – perhaps I glance into my closet, see my blue shirt, and recall that that was what I wore when I met John Edwards. The desire may become occurrent at this point, and now it can move me to action – I may put my dissertation aside for a moment, consider possible ways to advance Edwards' candidacy, and start writing a blog post that will make Democrats aware of his strong support for abortion rights. Thoughts of some object that I associate with a desired state of affairs can make a desire become occurrent, and put me into a position where that desire will affect the direction of my thoughts. Perhaps it will lead me along a chain of thoughts that concludes in action.

In the case of depressed people, the path of latent desires into being occurrent, or at least occurrent with their usual strength, seems to be blocked. Thus they are less disposed to feel pleasure, think out new ways of accomplishing their goals, or act. Since their desires are still in them, but blocked from becoming occurrent, it is right to say that they have those desires, just as it is right to say that a sleeping person has desires for all sorts of things. The hedonic theorist can then pitch his view as an analysis of occurrent desire, but not of latent desire. Since the proposed counterexample involves a case that does not actually fall under the hedonic theory of occurrent desire, it does not defeat the theory.

This distinction between occurrent and latent desire helps to explain some cases where forgetful people do not act. People sometimes fail to act while latently possessing a potentially action-inducing desire-belief pair because these mental states do not become occurrent at the right time. The old-fashioned practice of tying a string around your finger to remind you to do something is intended to counteract this. When you see the string on your finger later in the day, it will most likely occasion thoughts of whatever you were supposed to do, making the relevant mental states occurrent and leading to action.

Schroeder's second objection arises from his theory of pleasure. According to his Representational Theory of Hedonic Tone (RTHT2), "To be pleased is (at least) to represent a net increase in desire satisfaction relative to expectation; to be displeased is to represent a net decrease in desire satisfaction relative to expectation. Intensity of pleasure or displeasure represents degree of change in desire satisfaction relative to expectation" (94). Schroeder regards this as an improvement over the following theory, RTHT1: "To be pleased is (at least) to represent a net increase in desire satisfaction; to be displeased is to represent a net decrease in desire satisfaction. Intensity of pleasure or displeasure represents degree of change in desire satisfaction" (90). RTHT2 has an advantage over RTHT1 because it can account for the way that the intensity of pleasure or displeasure will vary with the epistemic state of the agent. If I see my favorite team win the game by scoring just as time runs out, I will most likely be much happier than I would have been at the moment of victory if they had won by a large margin and the outcome was not in doubt. Unexpected victories are more pleasant than expected ones, and RTHT2 accounts for this.

Schroeder argues that RTHT2 creates an ontological circularity problem for hedonic theories of desire. RTHT2 is incompatible with hedonic theories because "a representation of X is less ontologically basic than X: if pleasure explains desire, then it must be possible to say what desire is without mentioning pleasure, while it will be impossible to say what pleasure is without mentioning desire" (105). Since hedonic theories define pleasure in terms of desire, they cannot accept RTHT2 without being circular.

Of course, this argument will not succeed unless there is good reason to accept RTHT2, and there is not. There are many common experiences of pleasure that cannot be regarded as increases in desire satisfaction relative to expectation, and there are many common experiences of displeasure that cannot similarly be regarded as decreases in desire satisfaction relative to expectation. On the side of pleasure, consider the experience of eating a delicious dessert which you have ordered many times before at a restaurant. You have no doubt whatsoever that the dessert is coming, and you are as certain of receiving it when the waiter takes your order as when the dessert is in your mouth. Yet there is great pleasure in the eating. Similarly, orgasms and good massages that we are certain of experiencing are still very pleasant when they occur. These cases are unlike cases of pleasure induced by drugs and alcohol, where Schroeder could argue that we misrepresent the actual state of the world, as drugs can often cause us to do. It would be strange to claim that we are under a sort of hedonic illusion when we enjoy desserts, massages, and orgasms.

The side of displeasure offers even more powerful cases. Someone who gets a tattoo will feel pain, even though nothing in that moment is reducing her expected desire satisfaction. In fact, she may feel pain while being completely aware that the events of the moment promise to satisfy her desire for bodily illustration. RTHT2 is not a good account of what aversion necessarily is, since things we fully expect can often be painful.

Schroeder's response to these concerns seems ad hoc. He posits two separate systems of gut-level expectations, one of which changes very slowly and is located in the hypothalamus. This would allow him to claim that the agents lack hypothalamic gut-level certainty of getting desserts, orgasms, massages, or tattoos in the cases above. He claims that this system could be used to explain the greater enjoyment of small sensuous pleasures by people who are inured to hardship, and the greater sensitivity to small pains by people who have lived very easy lives. Insofar as these phenomena are real, there are plenty of explanations for them – from closer direction of attention towards the pain or pleasure by those not accustomed to it, to physiological changes like the callusing of skin that make the actual pain less intense. There are also reasons to wonder whether the amount of pleasure and pain felt by the hardened and the comfortable are significantly different, or if they are merely reacting differently to the same sensory experiences. Perhaps those of us who have lived lives of comfort just make more of a fuss when we experience the same amount of pain.

There is a final argument that Schroeder uses to defend the reward theory against both the standard theory (on which ties to motivation alone are essential to desire) and the hedonic theory.

> Because the reward theory places the essence of desire in a phenomenon, reward, which most people link only trivially to a desire (nothing really counts as a reward unless it is wanted), the reward theory allows desires to be independent of the most salient features of desire – motivation, pleasure, felt urges – and so deeply explanatory of them. Both the standard theory and the hedonic theory have a measure of this virtue, but both also give up a measure of it by identifying desire with some of its most familiar phenomena. They render trivial certain explanations that one might have thought were deeper. (178)

The plausibility of Schroeder's claim that reward is trivially tied to desire seems to rest on an equivocation. There is a nontechnical sense of reward in which anything someone desires could (perhaps under certain conditions) be regarded as a reward, and here the triviality holds. But Schroeder's Contingency-Based Learning Theory of reward brings in a technical sense in which "reward" picks out those things that trigger reinforcement learning. As the cases of the Unconditionable and the Creatures of Habit make clear, the connection between desire and this sense of "reward" is not trivial.

Furthermore, there are connections between behavior and pleasure (on one hand) and desire (on the other) which should be trivial. It is plausible that a mental state's being a desire entails some connections to pleasure, perhaps when imagining what is desired. It is even more plausible that a mental state's being a desire entails that it can interact with beliefs in a particular way to cause motivation. A hybrid of the standard theory and the hedonic theory, on which desires necessarily have the power to motivate action and to cause pleasure, would leave trivial what needs to be trivialized here.

Entailments need not hold in the opposite direction. If Schroeder wants it not to be trivially true that actions are motivated by desire, or that hedonic sensations are caused by desire, accepting such a hybrid theory would not interfere with his goal. One could hold that the ability to cause pleasure and motivation are necessary conditions for desire, and also hold that pleasure and motivation can be caused by things other than desire. Then it will not be trivial, merely from the fact that someone acts or feels pleasure, that her action or pleasure was caused by a desire.

Now I have finished responding to Schroeder's arguments that neither motivation nor pleasure are necessarily tied to desire. It seems that there is good reason not to think of "desire" as a natural kind term whose semantics closely mirror the semantics of "water." Desire's ability to cause behavior and pleasure are not analogous to the clarity and wetness of water, which are merely part of a stereotype that helps us identify it. In fact, being a cause of pleasure and motivation is essential to desire itself. Far from being necessarily constituted by a hidden essence that explains its outward properties, desire wears its essence on its sleeve. Mental states are usually regarded as being multiply realizable while water is not, and if the essence of desire was to cause pleasure and motivation under certain circumstances, that would explain why it is right to regard desire, but not water, this way. Desire need not be realized on the chemical level by organic macromolecules, on the neurobiological level by neurons, or on the psychological level by a reward system. All it needs to do is cause behavior and pleasure under the right circumstances.

Throughout this chapter, I have been using intuition tests about counterfactual possibilities to get clear about the nature of desire. Except in the cases of the Creatures of Habit and the Unconditionable, which I used to respond to Schroeder's view, I have not made any stipulations about how things are in the actual world. In setting up the cases this way, I have implicitly been assuming that the necessary conditions for a mental state's being a desire can be discovered through conceptual analysis alone. I have no conclusive argument to rule out the possibility that this assumption is false, and that the necessary conditions for desire depend in some way on how things are in the actual world. Theories different in their details from Schroeder's, but which are similar in that the necessary conditions for desire are to be discovered *a posteriori*, will have to have their own hearing and face their own versions of the counterexamples I have offered above.

But the failure of Schroeder's theory gives us some reason to think that these other theories will be unsuccessful. What better way to build such a theory than to take the phenomena that desire is most often taken to explain – motivation and pleasure – and identify desire with whatever actually explains these? In building his account this way, Schroeder mirrors the way that Hillary Putnam discusses the metaphysical necessity of the identity between water and $H_2O$ – since the chemical structure of water explains its outward stereotype, it must be part of the necessary conditions for something being water. The failure of such an attempt makes the entire class of theories on which the necessary conditions for desire are to be discovered *a posteriori* look unpromising. The project of discovering the necessary conditions for desire through *a priori* reflection, then, can be pursued with some degree of confidence.

## Schroeder on Positive Desires and Aversions

Before I finish with Schroeder, I would like to go over some things he says that are interesting and correct about desire. One of these is his distinction between positive desires and aversions. As he points out, we sometimes are averse to something in a way that cannot simply be understood as desiring that the thing not obtain:

> Being averse to something – say, to Adam's lateness – is not the same thing as having a positive desire or appetite for its contrary – say, that Adam be on time. A person who positively desires Adam's timeliness is prone to delight when he is unexpectedly on time, while a person who is simply averse to Adam's lateness will more typically be relieved, not delighted, by such events. Aversion sets one up for anxiety or relief; positive desire makes possible joy or disappointment. (127)

As Schroeder points out, these two kinds of states differ in the emotions that result when they are satisfied or unsatisfied. (I will use his terminology, on which "desire" picks out a class of entities including both positive desire and aversion.) There are similarities between the states that we experience when we get what we desire, and between the states that we experience when we fail to do so – relief and joy are both pleasant feelings, while anxiety and disappointment are both unpleasant. But there are also phenomenological differences, as well as externally visible differences. Anxiety and disappointment feel different from each other, as do joy and relief. The facial expressions and changes in body language that accompany these emotions are different enough for us to be able to tell which of these states someone is in. We can summarize these facts about desire as follows:

> **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.

The distinction between positive desires and aversions may also have some ramifications for direction of attention. While positive desires are more likely to push our attention towards states we try to achieve, aversions push our attention towards states that we try to avoid. This creates an interesting parallel with the relation between direction of attention and object-talk that was brought up in the discussion of Davidson. There I suggested that the reason why ordinary objects like chocolate, necklaces, goldfishes, and dashing young men are sometimes described as the "objects of desire" was because of the way that attention focuses on them, when they are in a desirer's presence. Perhaps there is some deep-seated connection between the way that the "object" locution works and the way we direct our attention. I have described the attention-direction aspect in terms loose enough to allow for a greater focus on the desirable state in the case of positive desire, and a greater focus on the undesirable state in the case of aversion.

Schroeder finds psychological and neurobiological differences underlying the differences between positive desire and aversion. On the psychological level, positive desire is associated with the reward system, while aversion is associated with the punishment system. On the neurobiological level, the reward system is realized in the VTA/SNpc. Punishment appears to be realized somewhere else, as mild aversive stimuli have no effect upon the VTA/SNpc (50). Schroeder says that research on the location of the punishment center is inconclusive, but he argues that the DRN is the place where punishment is realized.

Schroeder makes another useful general point when he replies to the objection that desires should not be taken to continue existing when they are satisfied (137). Since he takes desires to be connections between representational capacities and reward systems, and these connections are stable, he holds that desires continue existing when they are satisfied. Certainly, the motivational effects of desire subside when we believe that we have got what we want. But desires still have psychological effects under these conditions – when we imagine losing something that we have, we often feel displeasure. Schroeder points out that while we do not usually say of even the happiest Harvard students that they desire to have gotten into Harvard, we have plenty of other expressions picking out associated pro-attitudes that are generally used. We can say, for instance, that these students care about having gotten into Harvard, and that they want to be where they are. There are also expressions actually using the term “desire” to pick out states of affairs that we know to obtain – sometimes, things are exactly as we desire them to be.

**The Analysis**

So which of the features of desire are necessarily part of it, and which are only contingently connected with desire? The arguments for the conclusions I will present here are implicit in the many discussions of otherworldly creatures in this chapter.

Two of the actual features of desire cited in this chapter – the two flavors and intensification by vivid images – have not been tested for necessity. This is because they are not plausible candidates for necessary conditions of desire. It is hard to see what kind of necessary truth could possibly be drawn out of the fact that desire has two flavors. It seems that a creature could have only positive desires, or only aversions, and these states would clearly count as desires. Perhaps there are even other possible flavors of desire, where pleasure and displeasure are mixed with other kinds of experiences to create emotions that humans do not have. So the fact that desire has two flavors is only contingent. Intensification by vivid images is also a merely contingent feature of desire – there is no reason to think that all possible desirers would have it.

Two other features of desire were tested for necessity in this chapter – DODI and the attention-direction aspect. DODI was tested in the example involving the Angels, while the attention-direction aspect was tested in the case of the Absent-Minded. Neither feature was an intuitively plausible candidate for being a necessary condition for desire.

Now for the motivational aspect and the hedonic aspect. If we stipulate into the case of the Weather Watchers that giving them the appropriate means-end beliefs would cause them to attempt actions, they seem to have desires. But if the cases are constructed so that even the addition of means-end beliefs does not lead to action, they do not seem to have desires. So it seems that the motivational aspect is necessary. While intuitions about the Neutrals are less clear, there is some reason to think that they lack desires because their motivational states lack the proper counterfactual conditions to experiences of pleasure and displeasure. If this is right, the hedonic aspect is also necessary. The motivational aspect and the hedonic aspect, then, are the two things that are necessary for a mental state's being a desire.

# Chapter 3: Explaining Deliberation

## Introduction

In the first chapter, I presented the two theses that make up the Humean theory of motivation:

> **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.
>
> **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 argued that even if this formulation of the Humean theory not a conceptually necessary truth, its contingent truth would create difficulty for the combination of cognitivism and internalism. If the Humean theory is only true as a matter of nomological necessity – that is, if it is merely true within the range of human psychological possibility – anyone who holds that human beings can make moral judgments will have to choose between cognitivism and internalism.

The second chapter further elaborated the Humean theory by clarifying the nature of desire. I laid out six features of desire – its motivational aspect, its hedonic aspect, its attention-direction aspect, its two flavors, its capacity for intensification by vivid images, and the fact that DODI is true of it. While only the first two of these aspects are conceptually necessary for desire, desire has all six of these aspects throughout the realm of human psychological possibility.

In this chapter, I will defend the version of the Humean theory presented in the first chapter by relying on the features of desire that were discussed in the second chapter. Many opponents of the Humean theory have objected that it is not successful in explaining actual features of our deliberation and action. Many of these objections can be answered if we get clear about the nature of desire, and about the role it would play in shaping our deliberation in various situations. Once we understand what desire is actually like, and consider how desires would interact with each other and with other mental states, it will become clear that Humeans can explain the phenomena that their opponents present as problematic. My aim is to show that explaining what happens in even the most complex cases of deliberation requires nothing beyond the resources of the Humean theory.

First I will go over the methodology that will guide me throughout this chapter, and explain why we need to rely on ordinary folk psychology to evaluate the Humean theory. Then I will address objections to the Humean theory. I will start with perhaps the most famous opponent of the Humean theory, Immanuel Kant. Then I will consider a classic objection to the Humean theory dealing with the phenomenology of obligation. Next I will look at a proposed counterexample from G.F. Schueler, who holds that a Humean theory cannot account for the causal efficacy of deliberation without falling back on a trivially weak notion of desire. Then I will turn to a detailed counterexample from Stephen Darwall, who joins Thomas Nagel and John McDowell in holding that DODI is false. Thomas Scanlon's objections dealing with the phenomenology and structure of deliberation will follow. Then I will consider John Searle's voluntarist account and his criticisms of the Humean theory on the issue of weakness of will. Finally, I will turn to the neo-Kantian account of Christine Korsgaard, which grants agents a surprising amount of power to determine their motivational states through processes of reasoning.

## Methodological Questions

The central thesis of this dissertation is not as modally strong as many philosophical theses are. I argue that the Humean theory of motivation is true of human psychology, not that it is a conceptual or metaphysical necessity. Sometimes I make claims about conceptual or metaphysical necessities – for example, in the previous chapter, where I argued that desire has counterfactual connections to action and pleasure as a matter of conceptual necessity. But my central thesis is that human beings are motivated in the way that the Humean theory describes. If true, this is only a contingent truth.

The usual way of establishing contingent truths is through empirical evidence, and the usual way of establishing a psychological claim about the nature of motivation is through the methods of empirical psychology. At this point, however, there is insufficient evidence from the discipline of psychology to conclusively prove or disprove the Humean theory. As I will explain in discussing Darwall's proposed counterexample, it is a point in favor of the Humean theory that agents faced with some unexpected and unfortunate situation typically feel displeased even before they form any plans about how to change the situation. But it is difficult for psychologists to gather data on when agents typically experience pleasure and displeasure in the course of making decisions.

In addition to the specific difficulties involved in precisely mapping out the mental lives of deliberating agents, there are broader reasons why systematic research of the kind that could prove or disprove the Humean theory has not been done. Major research programs in psychology have not emphasized this particular kind of research. Social psychologists have found many interesting results about motivation, but have not built these into a systematic and general motivational theory of the sort that the Humean theory of motivation is supposed to be. Cognitive scientists do very systematic work on mental processes, but have traditionally focused more on beliefs and how they are formed than on desire and motivation.[12] And while the old behaviorists worked systematically on motivation, their anti-mentalist theoretical commitments caused them not to focus on precisely characterizing the aspects of our mental lives where arguments for and against the Humean theory can best be grounded. Behaviorists' lack of interest in internal, unobservable mental processes prevents them from producing the kind of results that would prove or disprove DODI, which turns on how our processes of reasoning are to be characterized. While there is reason for hope that cognitive scientists focusing on motivation will do the kind of research that would allow us to prove or disprove the Humean theory in the future, such research has not yet been done.

One might also hope that results from neurobiology could help us prove or disprove the Humean theory. Consider the previously mentioned issue of mapping out when agents have experiences of pleasure and displeasure in decision-making. One might think to have agents play games that involve decision-making under various kinds of

[12] I thank Art Markman for describing this to me.

circumstances while undergoing magnetic resonance imaging. Then we might be able to see when the neural seats of pleasure and displeasure are activated, and which other mental processes occur. But research of this kind is not currently possible, because there is still substantial disagreement between neuroscientists about where the neural seats of pleasure and displeasure are located in the brain. Leonard Katz writes that while "We are now beginning to understand how pleasure is organized in the brain… these are very early days yet, so results should be taken only as illustrative of the space of live possibilities rather than as indicative of what pleasure is."[13] Katz's own criticisms of Timothy Schroeder testify to the lack of consensus about the neural seat of pleasure.[14]

Given the lack of scientific data, how should we proceed? Should we withhold belief on the issue entirely, accepting neither the Humean theory nor its negation, on grounds that psychological claims cannot be justified without support from rigorously collected empirical data? That would be a drastic step. To hold that we must await rigorous scientific confirmation or disconfirmation before accepting any beliefs whatsoever about human psychology would be to give up the common-sense folk psychological judgments that have always guided us in our interactions with each other. When we try to determine how someone will react to some news or whether they would enjoy a particular activity, we make predictive judgments about how their mental states will interact to produce emotions and behavior. To hold that our lack of rigorously collected empirical data leaves us unjustified in holding our background conception of human psychology would

[13] Katz, Leonard. "Pleasure" Stanford Encyclopedia of Philosophy.

[14]" That pleasure is *uniquely* situated in the PGAC, rather than instead or equally in other regions showing up similarly in brain imaging studies, is not well supported by Schroeder's citation of studies, selected from a large literature, showing results for varying cingulate loci and many other correlations besides."

be to deny that we are justified in our most basic predictive judgments about how people will respond to situations.

Views differ about how we generate these predictive judgments. According to some theorists, we generate the predictions from an explicit theory of mind. According to other theories, we generate the predictions by internally simulating other people's mental processes. There is also debate about how much of our capacity to make correct predictions about human psychology is learned through observation of ourselves and others, and how much is innate. All parties to these debates, however, agree that we are able to make fairly reliable predictive judgments about how human beings behave and feel. Setting aside eliminativism about the mental – a radical position that I will not be dealing with here – all parties to the debate agree that human beings, given adequate information from their ordinary dealings with each other, are capable of making reliable (though fallible) judgments about the mental states of others.

One might think that the less controversial question of whether we are able to make accurate predictions about the mental states of ourselves and others should be treated separately from the more controversial question of whether we are equipped to defend a well-articulated psychological theory of the sort that I defend here. But if we have the capacity to make a rich and accurate set of predictions about the psychological states that people will go into under various conditions, we can use this capacity to develop the psychological theory. Usually, one builds a theory from observations of the phenomena that the theory is supposed to describe, but if one has an independent source of reliable claims about the phenomena, those can play a role in developing and defending the theory. Physics researchers operate on this principle when they run computer simulations of difficult-to-observe phenomena, and use the outputs of these simulations to guide theory-construction. If our predictive judgments about human psychology are rich and accurate, we ought to fit our psychological theories to them. Of course, actual observational data will trump folk-psychological predictive judgments, and if enough observational data contradicts these judgments, we will be forced to conclude that these judgments are generally unreliable. But in the absence of such data, and in view of the general reliability of our predictions, using our capacity to make accurate folk-psychological judgments is a fine way to build a psychological theory.

This capacity for judging which mental states people will be in under various sorts of conditions is presupposed by those who propose counterexamples to the Humean theory of motivation. In many cases, they describe situations and elicit our judgments about how agents would feel or act in those situations. They then argue that the Humean theory commits one to conclusions incompatible with our judgments, and that this gives us reason to reject it.

In responding to the opponents of the Humean theory, I will rely on the same general methodological principles that they accept. When they argue that the Humean theory cannot explain how human agents are able to engage in familiar kinds of deliberation and action, I will respond by showing that the Humean theory is compatible with – and indeed, capable of explaining – the phenomena of deliberation and action described in the proposed counterexamples. In many cases, the very phenomena offered as counterexamples to the Humean theory are better explained by the Humean theory than by its competitors.  This gives us reason not only to set aside the anti-Humean objections, but to accept the Humean theory.

A few of the objections I deal with will present particular anti-Humean views as having appealing consequences, or show that Humean views lead to some bad consequence.  In these cases I will usually attempt to show that the consequences described do not obtain.  In the one case where neurobiological data is brought out in defense of an anti-Humean view, I will respond with neurobiological evidence showing that the conclusions being drawn from this evidence do not follow.

Folk psychology is by no means infallible, and in the surprising event that it is radically mistaken, the judgments about human psychology underlying my arguments (and, for that matter, the arguments of my anti-Humean opponents) may be in error. Perhaps in decades to come, new psychological research will render all the considerations advanced here obsolete, and give us a more reliable way of testing the Humean theory. But at this early point in the history of psychology, the only way to determine whether the Humean theory is contingently true is with reference to our folk-psychological judgments.  And so I will proceed by relying on those judgments.

## Kant, Freedom, and Deliberation

The attempt to show that the Humean theory cannot explain everything about how we decide and what we decide in cases of practical deliberation has a long history.  Consider this passage from Immanuel Kant's *Critique of Practical Reason*, in which Kant argues that our experience forces us to believe in freedom, by showing us reason's power to guide our actions, independently of what we might happen to desire, through the moral law:

> Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him. (5:30).

Kant takes this case to be an empirical example, grounded in our experience of the actual world, that demonstrates our capacity for freedom. As he says immediately before introducing the example, "one would never have ventured to introduce freedom into science had not the moral law, and with it practical reason, come in and forced this concept upon us. But experience also confirms this order of concepts in us" (5:30). Here Kant is not only arguing for the possible falsity of DBTA – he is arguing on *a posteriori* grounds for its actual falsity. This makes him an opponent of even a contingent version of the Humean theory. While Kant offers many *a priori* arguments against the Humean theory involving what would be required for rational action, this argument is distinguished by its dependence on empirical facts that are supposed to show the contingent truth of an anti-Humean theory of motivation.

If we assign the most plausible strengths to the agent's desires for the various options, a Humean view can easily account for the fact that the agent would quickly decide not to satisfy his lust, but would regard sacrificing his life to prevent the execution of an innocent man as “possible for him.” For most of us, the aversion to being killed today is far stronger than the desire to have sex once. While the aversion to participating in the execution of a good man may be weaker than the aversion to being killed today, it is much stronger in most of us than the desire for one episode of sex. We can see this in the fact that most people would not participate in the unjust execution of a good man, just to have sex once. So while we easily make the decision to forgo sex when our lives are at stake, we are torn between our powerful aversions in the case of the other decision. Even if it is not the desire that eventually will determine our action, our aversion to participating in the unjust execution of a good man is strong enough to direct our attention towards this possibility for a while before choosing. It is because this possibility is strong enough to hold our attention and prevent us from choosing quickly, causing us to deliberate even in when the other choice is our own deaths, that we say that averting the man’s execution would be possible for us.

Presented with the choice between this Humean explanation where deliberation is shaped by the different strengths of our desires, and the Kantian explanation where the agent’s freedom plays a role in guiding deliberation, which explanation should we prefer? The Humean explanation should be preferred, because of its greater simplicity. It successfully accounts for all of the data without committing us to the sort of freedom that the Kantian explanation does. The agent’s positive desires and aversions are brought in by both explanations (the emotions associated with the motivation to not be killed track the profile of aversion far better than positive desire, and the Kantian will have to posit aversion in order to deal with them) but the Kantian goes an additional step by bringing in freedom. The Humean theory makes better use of its explanatory resources by getting positive desire and aversion to do all the work.

Here, and later in this chapter, the ability of desire to direct attention towards the object of desire or towards something we associate with the object of desire will do a lot of work in dealing with putative counterexamples. A significant part of the experience of deliberating is the experience of having our attention focus on salient aspects of our options. This can manifest itself in a sensory focus on some nearby object, or on the contemplation of some possible or actual state of affairs. In the case of the choice between sex and life, the difference in desire strengths causes the decision to go quickly enough that we do not spend much time looking at each option. But in the case of the decision between dying and participating in the judicial execution of an innocent man, we would spend more time considering the options, which involves focusing our attention on aspects of the choices that are salient with respect to our desires. In this kind of case, aversions will probably play a bigger role than positive desires. Among other things, an aversion to unjustly killing people will cause us to reflect with horror on the possibility of his execution, and an aversion to acting dishonestly will make us reflect unhappily on the prospect of telling a lie.

## 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 does so by invoking explanatory resources that both Humeans and their opponents will have to admit, while not invoking any mental states other than desire that are capable of motivating action. Thus it offers a simpler and better explanation of the feeling of obligation than anti-Humean theories do, giving us reason to accept it.

Some opponents of the Humean theory have claimed that this 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 heteronomy of the will when it is driven by desire, is the most famous representative of this strand of the anti-Humean tradition.

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 she is under an 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 feeling different from many of our ordinary experiences of desire, and which 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 20th 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" (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 motivation:

> 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 which 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 are not present in many clear cases where we are motivated by desire. Imagine that 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 announce a spur-of-the-moment plan to go 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 than 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 Davidson’s inclusion of regarding something as “dutiful” or “obligatory” in the category of pro-attitudes (169). 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” (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 which is stronger than the desire to go to the party. While this would successfully explain the agent’s behavior in the case where she decides to keep her 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 her to go to the party. Simply positing another desire can explain her 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.

In the first case that I will consider, the grader seriously considers each of the choices before her, and weighs whether to go to the party or keep her promise. As she does this, she will feel different emotions in considering the options before her. 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 she has to do, and does not seriously weigh the possibility of going to the party and leaving her promise unfulfilled. She will feel some disappointment at not being able to go to the party, but she will not seriously consider leaving the papers ungraded. While the desire to go to the party pulls at her, grading the papers will seem to have a kind of necessity, and she will not have the experience of weighing one desire against the other. The feeling generated by the motivational force that causes her to stay in her office and grade the papers will be less intense than the feeling generated by the motivational force that pulls her towards the party. But despite its lesser emotional vehemence, the former force will determine the course of her reflection and her decision. This is the case of the Unwavering Grader.

Now I will respond to this objection by showing how a 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.

And this is how the case of the Tempted Grader can be best explained. The different feelings associated with each of the options before her – the excitement of thinking about the party versus the duller satisfaction when she thinks about being able to hand back graded papers, the disappointment when she thinks of missing the party versus the anxiety when she 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 of the phenomena in the case of the Tempted Grader provides as good an explanation as my version of the Humean view does. 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, 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 what reason is there to prefer the Humean explanation?

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 simple 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 realizes on the way home that he does not have his wallet, and decides to walk back to the bar where he thinks he left it.

The experience of having this realization is not typically one of disappointment, but one of anxiety and worry, so a positive desire to regain the wallet will not explain his emotions.  This is not a case where believing a moral principle is explaining his emotions or his decision.  And while one might think of explaining the motivation and phenomenology by appealing to beliefs in principles of prudence, one would still have to explain why prudence creates this particular phenomenology in this situation.  In other cases (for example, the case of an investor who thinks she 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 merely appealing to prudential motivation will not explain the particular emotions in this case.

An aversion to losing his possessions, 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 lost wallet.  Rather than invoking a new primitive motivational force, the Humean builds a simpler theory by using a motivational force that both sides must allow.

This methodological point will be relevant to the explanations that I offer throughout this chapter.  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 for that particular purpose, 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 cannot, in the end, do without. 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 which we see before us are more violent than desires for more distant things. The beauty of the Humean theory is that it explains complex cases using explanatory resources that both sides accept. Using only these resources, 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 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 obligation, and generate 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 do not have many close associations between concepts at this level of abstraction and things that we can have vivid imaginative or sensory representations of. Second, people who reliably fulfill their obligations often have confidence in their abilities to do so, and 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.

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 which he closely associates with morally wrong action in the same way that we associate food very closely 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 others 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 the aversion to breaking promises may be when we are away from the person to whom we have made a particular promise, it will usually become more violent when we are looking into the promisee's eyes. In these cases, the passion moving us to action gains violence because of the vivid sensory representations that we experience.

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.

Contrast the case where I decide against jumping out the window, without seriously considering it, with a case in which I have to seriously consider jumping out the window. Perhaps some billionaire with strange and gruesome preferences appeals to my utilitarian 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 she misses her party because she has to finish her grading.

I will now return to her case. In making her decision, the Unwavering Grader was faced with two choices which felt different to her. 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 her towards 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 her mind as she 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 she knows herself to be reliable in fulfilling her obligations, she will not have particularly vivid mental representations of failure to fulfill her obligations. And since she has a prior commitment to grading, backed up by powerful desires, she 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.

### Schueler and Deliberation

In *Desire: Its Role in Practical Reason and the Explanation of Action*, G.F. Schueler distinguishes between two senses of desire. He takes the ordinary use of the term "desire" to be ambiguous between them. Scheuler argues that the first of these senses, which resembles my account of desire, is unable to explain what happens in cases where agents deliberate about what to do. I will argue that desire in this sense is able to provide good explanations of deliberative actions.

The first sense of desire that Scheuler considers is that of "desires proper," which include "such things as cravings, urges, wishes, hopes, yens, and the like, as well as some motivated desires, but not such things as moral or political beliefs that could appear in practical deliberation as arguing against the dictates of one's urgings, cravings, or wishes" (35). "Motivated desires" here is a term from Thomas Nagel that picks out desires arrived at after deliberation. The second sense of "desire" is that of "pro attitudes," which includes motivational forces of all kinds.[15] If a moral belief could directly motivate action, it would be a pro attitude but not a desire proper.

While Schueler does not attempt a precise characterization of desires proper, it seems that desires proper are closer to desires as I defined them in the previous chapter than pro-attitudes are. The notion of desire that I have been working with does not merely involve connections to motivation, as is the case with Schueler's pro attitudes. Desire, on my view, also involves necessary connections to pleasure, and connections to lots of other psychological states which hold throughout the space of human psychological possibility. So when Schueler claims that "if we stick with desires proper, then the desire/belief model of agents' reasoning… is not true," I take him to be rejecting the Humean theory as I have laid it out (48).

In the final chapter of *Desire*, Schueler considers two models of how actions might be motivated. He calls the first the "blind-forces model." On this model, what the agent does "is simply the outcome of the causal interaction of his or her desires with the 'information states' in his or her brain" (171). Schueler contrasts the blind-forces model with the "deliberative model", on which "the agent weighs up the various considerations

[15] Schueler does not claim that Davidson was using the term "pro attitude" in the same way he does. He seems to regard the matter as an open question (48).

that seem to him or her to count for and against performing the various actions available and then performs the action that seems to have the most to recommend it, at least in the optimum case where there is no failure of rationality" (173). On this model, the considerations favoring or opposing an action are not limited to ones that figure in the contents of an agent's desires proper. If desire is understood as desire proper, the deliberative model is inconsistent with DBTA.

Schueler holds that the blind-forces model will succeed in explaining actions that we perform without deliberation, while the deliberative model will (unsurprisingly) succeed in explaining actions where we deliberate. However, he does not see how either model can explain all cases of action. The deliberative model will fail in cases of impulsive action where the agent does not deliberate. The blind-forces model, if it incorporates only desires proper[16], will fail in cases of deliberative action. This causes him to wonder, at the very end of his book, "what if, as I am suggesting, neither of these two models turns out to be plausibly eliminable in favor of the other?" (196).

Schueler would probably characterize my Humean view as a version of the blind forces model. Since I include psychological states like imagination (which might be characterized as non-informational) among those that can causally interact with desire to determine how an agent eventually acts, there might be some questions about whether my view falls under his exact definition of the model. But I am happy enough to be regarded

[16] The blind forces model, it seems, could account for all actions if paired with the pro attitude view of desire. I am happy enough to follow Schueler in considering his preferred pairing of the blind forces model with the desires proper view, however, since that is the general outline of the theory I have proposed.

as a blind-forces theorist, and to argue that my view will succeed in dealing with cases of deliberation.

Schueler's concern about the inability of the blind-forces model to explain deliberation arises from difficulties that previous blind-forces models have faced in doing so. The blind-forces model that he focuses on the most is that of Fred Dretske. While Dretske's account is like mine in making desire-belief pairs causally necessary for action, it also tries to build in a story about the causal genesis of desire, and in doing so it brings in several complexities that my account avoids. On Dretske's picture, desires produce particular movements when the agent believes that she is in the same circumstances under which those movements produced a reward for her in the past. Thus, past rewards play an important role in the development of desire.

Dretske acknowledges that his view does not give a satisfactory explanation of deliberation. He considers a case in which a hungry jackal sees a tiger eating an antelope. The jackal desires to eat the antelope, but also fears getting close to the tiger. So the jackal may decide to wait at a safe distance until the tiger eats its fill and wanders away. This is not the course of action demanded by either of its desires, though it may be the path to the greatest aggregate desire-satisfaction. Dretske says that "How this novel third result is synthesized out of control structures already available is, at the biological level, a complete mystery" (Dretske 141). As Schueler notes, the problem here is not really at the "biological level" (Schueler 141). Rather, "The mystery is how we are supposed to extend the theory he gives for, and explains in terms of, cases of a single motive (the practical-syllogism case) to cases of multiple motives." If desires are tied to the performance of movements when the agent believes that conditions obtain under which those movements have previously led to reward, how can we explain an agent's decision, formed through deliberation, to engage in novel movements?

Different explanations may be appropriate for jackals and humans, and some people may not want to attribute any sort of deliberative processes to the jackal in the above case. But humans certainly deliberate in this way, and a theory of motivation must explain how this is possible. The Humean theory I have offered is able to use the attention-direction feature to explain how an agent's deliberations here might start. An agent's desires have the ability to direct her attention towards a possible state of affairs that she associates with their object, and in a situation like this both of her desires might combine to direct her attention towards the possible state of affairs in which she avoids the tiger by waiting nearby and gets the food by seizing it once the tiger leaves. The formation of an occurrent means-end belief about how to jointly satisfy her desires thus naturally accompanies the attention-direction. The agent becomes motivated to wait nearby until the tiger leaves, and then feast on the antelope. Since my account only brings in the agent's movements insofar as they figure in the antecedent of the means-end belief (unlike Dretske's, which attaches a fixed set of movements to each desire) I am able to account for novel movements while Dretske cannot.

Schueler also considers the view of Alvin Goldman, which he regards as an "unhappy hybrid that tries unsuccessfully to combine elements of both the blind-forces model and the deliberative model" (174). On Goldman's view, actions are caused by the agent's beliefs about which actions are likely to satisfy the most of her desires (Goldman 74). It appears similar to the blind forces model in that the concept of desire is essential for the explanation of action. However, it is not the case that actual desires are necessary. An agent could, for example, falsely believe that she desired something when she had no desires at all, and thus act without desire. Such an agent not actually be motivated by desire. Since the mental state that actually motivates action is a belief and not a desire, Goldman's model is not a blind forces model, and also not a Humean model.

To argue that neither of the two models is sufficient to explain all action, Schueler presents two examples, both of which involve him eating chocolate from a candy jar. In the first case, he eats it impulsively. In the second case, he deliberates about whether he should eat it or not, decides that there are no good reasons not to, and eats it. He argues that in the two cases, the actions are explained by different psychological states. In the first, a desire for chocolate and a means-end belief that he can eat the chocolate by taking it out of the jar, unwrapping it, and putting it in his mouth explains the action. In the second, the role of the desire and the means-end belief is occupied by a judgment. As one possible form that the judgment may take, Schueler offers this:

**[J]** I judge that I want to eat some chocolates. (178)

Since impulsive action involves desires while deliberative action involves judgments, Schueler argues, neither the blind-forces model nor the deliberative model will be able to deal with all cases of action.

Schueler first considers the response, from the blind-forces theorist, that the causation of action in the deliberative case is just the same as the causation of action in the nondeliberative case. He rejects this response because it leaves J out, denying the judgment any genuine explanatory power. If the desire and the belief explained the deliberative action in just the same way as they explained the impulsive action, "it would simply be false that I acted *on the basis of* my deliberation" (180). Deliberation, on this view, "would be simply a kind of epiphenomenon" (180). While Schueler accepts that some cases of rationalization may be like this, he does not want to say that all cases of deliberation go this way.

Schueler then considers the response that J plays a genuine explanatory role in some kind of desire-belief causation. But he claims that this would require widening the spectrum of desires to include all pro-attitudes. After all, J (and the other claims that Schueler considers to replace the desire-belief pair in the deliberative case) can be made by an agent who does not have any desire proper for chocolate, or even by an agent who has no desires proper at all. Such claims would be false if made by such an agent, but they still could be made – for example, if the agent introspected badly about which desires he had. J is not a plausible candidate for being a desire proper. So if J can be considered a genuine cause of action, Schueler says, desires proper will not explain all actions.

I will now take a closer look at Schueler's example, and try to present a Humean account that employs only desires proper and gives J genuine explanatory power. In Schueler's example, he notices the candy jar, his craving for chocolate awakens, and then the following happens:

> It occurs to me that there are some reasons for me not to eat any chocolates, reasons to weigh against my desire to eat some. For one thing, I have already had several chocolates today, and I believe that chocolates are both fattening and bad

> for the teeth. Then too other members of my family have been complaining that I eat a disproportionate share of the candy in our household and that I should leave more for them.
>
> I decide, however, that none of these reasons is really all that telling when weighed against my craving for more chocolates. My teeth are fine, and I get enough exercise that gaining weight is not a problem. And since there are plenty of chocolates in the jar and I plan to take only a few, I will be leaving a lot for others. So I stop, open the jar, take out a couple of chocolates and start to eat them, replace the lid on the jar, and proceed on my way (175).

I will now offer a Humean story that explains what happened here, and makes it fit with a desire-belief model of action that involves only desires proper.

The example begins with the agent (we can call him George) having his desire to eat chocolate brought to the forefront of his mind by the sight of something directly relevant to its object – chocolate. This combines with his simple means-end belief about how to eat the chocolate, and he finds himself motivated to eat it. But before George can act on this motivation, the idea of eating the chocolate makes another of his previously latent desires occurrent – his prudential desire for his own desire-satisfaction. He associates eating chocolate with things like being fat and suffering tooth decay, which he desires not to happen. His aversion to disharmony in the home comes to the fore as well, as the thought of decreasing the chocolate supply makes him worry about irking his chocolate-loving family members. With these desires occurrent, he casts his attention on every side of their objects (as Hume would say). As he considers things related to the possibility of a strife-causing chocolate shortage, he notes that there are actually enough leftover chocolates to avert disharmony. Once he sees this, he sees nothing relevant to his aversion to family strife around him, so that aversion fades back to being latent. Similarly, as he attends to the possibility of tooth decay and weight gain, he notes that because of his present condition, eating more chocolate is unlikely to bring these negative consequences upon him.

But perhaps, before his desire for his own desire-satisfaction fades away into latency, it causes George to direct his attention toward something else related to its object. He notes that he wants to eat some chocolates. In doing this, he makes judgment J, judging that he wants to eat some chocolates.[17] J, together with the facts of his situation, implies the truth of a means-end belief – the belief that by reaching into a jar and eating the chocolate, he will be able to satisfy a desire. His prudential desire for desire-satisfaction and his belief that he can satisfy a desire by reaching into the jar and eating the desired chocolate come together, causing him to reach into the jar and eat the chocolate.

In my explanation of the example, the only motivational states involved were desires proper. J played an important explanatory role, as it led to the formation of a crucial means-end belief. So it is possible for a blind-forces theorist to explain deliberation using only desires proper, and without depriving deliberation of its explanatory power.

## 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

[17] Personally, I do not think I would make the judgment J if I were in George's shoes. My desire for my own desire-satisfaction would fade quickly along with my aversion to family strife, and I would eat the chocolate simply out of a desire for the taste of chocolate (much as in Schueler's nondeliberative case) rather than out of a desire for my own desire-satisfaction. If most people would not have J in such a case, this is not to the discredit of the Humean theory, but rather to the discredit of Schueler's commentary on his example, since it is his commentary that forced J upon us. Other cases that do not involve chocolate may, however, involve a judgment like J.

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 states. Rather than being a counterexample to the Humean theory, Darwall's case demonstrates the superiority of the Humean theory over competing explanations.

Early in *Impartial Reason*, Darwall attacks 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" (Darwall 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. I will consider an objection that he presents to a form of the DBR thesis that I accept, and examine this objection to see how my Humean view would fare against it.

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" can have a motivational reading, on which a reason to act is something that explains action (as opposed to something that justifies action). Then 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" (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.

Many other prominent rationalists have expressed views like the one that Darwall expresses here – that pre-existing desires are not necessary for the formation of a new desire through reasoning. In *The Possibility of Altruism*, Thomas Nagel attacks the view that "all motivation has desire at its source" (27). While Nagel accepts that "all motivation implies the presence of desire" (32), he thinks that we often come to have desires through processes of reasoning that do not include desire as a premise. In these cases, our actions are not motivated by any desire, but by the reasons that drove our deliberation and caused us to act. While attributions of a desire to an agent follow trivially from the fact that she was motivated, this desire does not play any substantial role in explaining the action or the decision to act, and no other desire need be invoked to explain the action or the decision. In Nagel's words, "That I have the appropriate desire simply follows from the fact that these considerations motivate me… But nothing follows about the role of the desire as a condition contributing to the motivational efficacy of those considerations" (15). New desires, then, can come into existence through processes of reasoning that do not have desire as a premise. All that has to happen is that the processes of reasoning must motivate an agent.

John McDowell, who cites Nagel in “Are Moral Requirements Hypothetical Imperatives?” shares this view. According to McDowell, if an agent acts to promote his future happiness, we credit him with a desire for his future happiness. However,

> the commitment to ascribe such a desire is simply consequential on our taking him to act as he does for the reason we cite; the desire does not function as an independent extra component in a full specification of his reason, hitherto omitted by an understandable ellipsis of the obvious, but strictly necessary in order to show how it is that reason can motivate him. Properly understood, the belief does that on its own. (15)

Like Nagel, McDowell denies DODI. On his view, all that is necessary to explain motivation is a belief. All action necessarily involves desire, but sometimes an agent has a desire only in virtue of the fact that she reasoned her way from having a belief to being motivated to act.

As a defender of a merely contingent version of the Humean theory of motivation, I am willing to concede to all these philosophers that the psychology of motivation which they describe is within the realm of conceptual possibility. We can imagine creatures that are motivated in the way that they describe. I presented such an example in the second chapter – the example of the Angels. But human beings are not Angels. In order to make it plausible that the Humean theory misdescribes the psychology of human motivation, these rationalists would do well to offer examples in which psychologically normal human agents form new desires in a way inconsistent with the Humean theory. Their examples would have to involve nothing beyond the kinds of processes that we recognize from ordinary human psychology.

I have chosen to focus on Darwall because he goes the farthest in trying to present such an example. His example runs as follows:

> 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 the film there is a discussion of what the students might to do 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.

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 pre-existing 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" (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 will be claiming that new desires can be generated by processes of reasoning that do not have desires as premises. Then he will – like Nagel and McDowell – 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 watched the film. Desire, as I have argued in the previous chapter, 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. And 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. It is hard to see how he could do so. 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 pre-existing desire combined with a sudden realization that a deeply undesirable situation obtains, can Darwall invoke? The hedonic aspect of desire neatly accounts for the fact that we are subject to these emotions.

It might be argued here, on Darwall's behalf, that a belief that others are being mistreated could generate the sentiments and motivation in this case, and that such an explanation is as simple as the Humean explanation. And if we were to look at the case of Roberta without thinking about the broader ontology of motivational states that Humeans and anti-Humeans are committed to, this view might be convincing. But when we consider the motivational states that Humeans and anti-Humeans are committed to, we can see how the Humean explanation of Roberta's case allows us to develop a simpler overall picture. While Humeans can plausibly claim to be extending a simple model of action that covers all the other cases, no similar claim can plausibly be made on the anti-Humeans' behalf.

Consider the cases of hunger, thirst, and sexual lust. Anti-Humeans generally concede that the motivational states arising in these cases are best understood as desires. It is hard to see how they could do otherwise, considering things like the ways in which these states are produced and the inability of theoretical reasoning to affect them. So both the Humean and the anti-Humean ontologies of motivational states will include desire. Against this background, an explanation of Roberta's deliberation and action in which desire is the only motivational state will be simpler than one that includes motivational states not attested elsewhere. As we come to Roberta's case (and the other cases discussed in this chapter), we have no similarly uncontroversial cases of the motivational forces that anti-Humeans invoke. Cases like Darwall's and the others discussed in this chapter are supposed to do the work for the anti-Humean that hunger, thirst, and lust do for the Humean. So if the Humean can explain these cases in terms of desire while capturing all the phenomena that the anti-Humean does, the Humean theory will have demonstrated its greater theoretical economy, and thus its superiority.

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 what he regards as the way that she would 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" (40). As Darwall says, "this need not be what happened."

Upon reading Darwall's example, one certainly does not imagine Roberta seeing an "opportunity" to satisfy a previously held desire in the plight of the textile workers. If the Humean theory claimed that she regarded the workers' misfortunes as an opportunity, 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. Part of the trouble here may involve a version of Davidson's mistake in "Actions, Reasons, Causes" – regarding the action, and not the state of affairs it produces, as the object of desire. Typically, people acting to relieve 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.[18] 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 desires that others not suffer, 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

[18] 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.

inquire into whether there are conditions that enable its realization" (40). Darwall is right to say that desire can make agents think about its object. This is actually a feature that I have built into my account of desire – desire causes us to direct our attention towards things we associate with the object, starting with the object itself. The "inquiring" that Darwall talks about can be reduced to a combination of attention-direction towards 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 Darwall intends to develop an objection to Humean views including mine out of this. The objection is simple, and goes as follows: if this sort of thought and inquiry is a necessary condition for desire, why is thought and inquiry about how to relieve 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" (39). Stimuli that would activate a latent pre-existing desire that others not suffer, then, are largely absent from 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" (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, she is healthy, and she does not take unnecessary risks, so I believe that the possibility 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 had 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" (39). I have argued above that Roberta's sensitivity to suffering is best understood as being at least partially constituted by a pre-existing 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.

While imagination, like belief, has a role to play in motivation, it cannot generate new motivations without the assistance of pre-existing desires. In fact, if someone had had a completely different set of desires, the same set of sensory and imaginative experiences might affect him in exactly the opposite way. 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, he might very well despise the textile workers 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 would 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.[19]

After spending so long on Darwall's example, it is time to make some general remarks about how a counterexample to the Humean theory of motivation would have to go. It is entirely compatible with the Humean theory that an agent who never previously felt the effects of some desire can be moved to action by forming a new belief. To develop a counterexample to the Humean theory, one must offer evidence that desires were not present in advance. Darwall attempts to do this, but does not succeed because the situation he describes is set up so as not to trigger any of the activating conditions under which the desire would have its characteristic effects. So the fact that Roberta's desire

[19] 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 agree with me 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.

does not display any effects is not a clear sign of its absence. To develop a counterexample to the Humean theory, one would have to show the desire's antecedent absence by producing a case in which the activating conditions were achieved, but no effects were seen.

Such cases, I maintain, are not conceptually impossible. I presented such a case in the previous chapter – the case of the Angels. They did not display the characteristic effects of desire before they engaged in reasoning that brought them to new desires. When, prior to reasoning, they imagined counterfactual states of affairs in which they performed a particular right action, they felt no pleasure. It was only after reasoning their way to a new desire to perform some right action that imagining counterfactual states of affairs in which they do such an action would give them pleasure. If humans are like this, then the Humean theory of motivation will not accurately describe their behavior. My contention, as a defender of the contingent Humean theory, is that examples like those of the Angels are located outside the realm of human psychological possibility.

### Scanlon, Desire, and 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 response in the first case will be familiar from the discussion of obligation – agents who act despite reporting "no desire" to do something are merely reporting a lack of positive desire, and are motivated by aversion. In cases of bracketing, deliberation is actually 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" (39). This account has several things in common with mine. Most obviously, Scanlon regards desire as capable of directing an agent's attention, while my view posits an attention-direction aspect of desire. One might also regard my hedonic aspect of desire as a way of cashing out how things an agent desires appear to her in a "favorable light." Scanlon, of course, 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." 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. Much like Schueler when he talks about "desires proper", Scanlon can be regarded as rejecting 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'" (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" (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 positive aspects of 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 towards things associated with what she wants, aversions direct an agent's desire toward things associated with the object of aversion – in other words, what she 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 her friend being in the dark about the bad news (though depending on the case, it may be 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 she is going tell her friend, she focuses on the unpleasant duty before her, and the negative aspects of what she has to do loom large in her 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 aversive passion. To overcome the unpleasantness that they feel in these cases, people sometimes choose to focus their attention on their freedom from the even worse consequences that they have chosen to avoid, and draw some relief from that. 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 only to refer 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 which I discussed in the previous chapter, 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 towards the states of affairs that we act to obtain, while aversions direct our attention more towards 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? Cases like this, then, are more fully 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" (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" (52). This adds up to an objection to the Humean view. Scanlon claims that Humeans are committed to regarding the weighing of competing desires against one another as the process by which rational decisions are made, and that this is not always the process by which we make rational decisions about action.

Scanlon gives an example of a kind of decision in which more complex structures than the weighing of competing desires are involved. He points out that many decisions "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" (52). Scanlon does not see how a Humean theory can explain our ability to do this. A decision in which we bracket some of our reasons is a decision in which we do not weigh all of them. If the Humean theory is committed to presenting every rational decision as one in which we weigh all our desires for the options against each other and choose the option that is preferred by a preponderance of our desires, it will not be able to explain the fact that we sometimes engage in this sort of bracketing.

I agree with Scanlon that rational decision-making is not always just a matter of balancing the weights of all of our desires. Certainly, this is not how our experiences present the experience of decision-making – we sometimes feel that we are able to intentionally exclude some interest of ours from the weighing. What I want to show is that the Humean theory is not actually committed to regarding the weighing of all our desires as being the whole story about rational decision-making, and that it can explain why we often make decisions in a way where we do not weigh all the available options. So I will deal with an example of bracketing that Scanlon raises and show how the Humean theory can deal with it.

This brings me to Scanlon's example.[20] It involves the chair of a philosophy department who has strong personal interests at stake in some decision he is making. He may put those interests aside in his deliberation and make his decision based on what is good for his department. This chair 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 explain his behavior reasonably well, but it would not be accurate to the phenomenology. He does not have the experience of weighing, but rather the experience of working towards a goal while he passes by attractive considerations towards 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 would the Humean theory explain how the chair makes his decision? The story should begin with the chair's antecedent desires about how his decision is to be conducted. Perhaps he forms a desire to act only in the best interests of his department.

[20] Scanlon actually spends more time developing a different example of bracketing. In this example, he decides whether or not he should play to win in a tennis match. Playing to win would involve bracketing considerations about his opponent's unhappiness at losing. I do not deal with that example, however, because it seems to me that when the agent decides "that it might be all right either to play to win or not to do so" nothing actually gets bracketed. I do not understand why Scanlon chose to run the example that way.

There are several ways in which this desire could be formed, and many of them follow the familiar Humean pattern in which belief-desire combinations generate instrumental desires. His desire may be formed by a process of considering how he generally ought to make decisions, feeling the weights of his initial desires, and being motivated to act in the department's interests by the preponderance of these. If an aversion to selfish behavior by those in power and his desire to be a good chair are the strongest, these desires will drive the formation of the desire that structures his decision. (In an example of how imaginative representations of a state of affairs can increase the violence of a desire, the fact that he is thinking of decision procedures and not the personal benefits of selfish action will help keep his selfish desires from becoming violent, while his desire for above-board decision-making may become more violent.) Or, when he first considers the question of how he should make his decision, his aversion to selfish behavior by people in official capacities may be strong enough that he goes right to forming his desire without seriously weighing the possibility of making his decision in some other way. Either way, it is not necessary for his antecedent desire to specifically include the decision before him in its content, or to be formed with this decision in mind. It could be that he long ago formed a desire to make official decisions, or generally to act, in some way that suggests that he should consider only his department's interests when making this decision.

Having formed this desire and having the capacity for introspection, the chair will be aware that he wants to act only in the best interests of the department. And if his desire is sufficiently firm, it will be rational for him to believe that he is going to act in the best interests of his department, at least when he is acting in an official capacity. We usually form beliefs like these about our future actions when we form sufficiently strong desires to act in particular ways, on matters that are within our control.

If the chair comes to his decision with a firm antecedent desire to act only in the best interests of the department, and with the belief that he will act in the best interests of his department, he will be able to bracket some of the considerations when he makes his decision. Desiring to act only in his 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 rather than weighing it with everything else, he will regard its object as being unavailable to him. This is why 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 feel 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 have some desire for 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.

Scanlon anticipates a Humean response that is like mine. According to this response, it is the agent's second-order desires that are responsible for the way that bracketing goes. On my explanation, bracketing is driven by the agent's second-order volitions – his desires that a particular desire or set of desires be effective in moving him to action. (Acting in the best interests of his department, as I have been understanding it, is acting in a way that is motivated by the desire for his department to do well, so the desire to act in the best interests of his department is a second-order volition.) 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.

Scanlon's point here is familiar from Gary Watson's response to Harry Frankfurt's account of free will. 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 the issue of bracketing needs to be cast as a matter of authority. As I have presented the issue, higher-order volitions generate the phenomenon of bracketing because of their content, their power, and our introspective awareness of their power. They have no greater authority than lower-order volitions, and they do not need any greater authority to structure deliberation. 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 an authority is involved to grant the permission. And the displeasure we feel with ourselves when we think about 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. When we consider acting on a bracketed-off consideration, we are thinking about acting so that the antecedent desire that structured our bracketing is not satisfied. These thoughts are often unpleasant, as the previous chapter's discussion of the hedonic aspect of desire would suggest. And if the antecedent desire is a second-order volition whose object involves our own actions, our unhappiness will be unhappiness with ourselves.

The way that the Humean theory deals with the case of bracketing supports the point that Nietzsche makes in *Daybreak* 109 – strong higher-order desires are often capable of structuring deliberation, and authority has nothing to do with it. More generally, it is a case that shows us the explanatory power of the Humean theory. The way bracketed-off options look to us – both in their seeming inaccessibility, and in the displeasure we feel when we imagine being the unsavory characters who could pursue them – are continuous with the way that our desires shape our thoughts in simpler cases. Bracketed-off options look like things we desire that we know we cannot get. Thoughts of being bad enough to pursue them cause displeasure, as thoughts of situations we are averse to generally do. It is deep continuities like these between the simple and complex cases of deliberation that make clear the driving role of desire in the complex cases, and give us reason to accept the Humean theory of motivation.

## Searle, Freedom of the Will, and "The Gap"

One of John Searle's major contentions in *Rationality in Action* is that belief-desire pairs are not causally sufficient for a large number of our actions – in particular, our rational actions. According to Searle, "the cases of actions for which the antecedent beliefs and desires really are causally sufficient, far from being models of rationality, are in fact bizarre and typically irrational cases" (12). He tries to support this claim with neuroscientific evidence and by claiming that we cannot regard ourselves as genuine agents if we are only moved by belief-desire pairs. I will defend the Humean theory by showing that his neuroscientific evidence does not support the conclusion that he thinks it does, and a proper understanding of the self will reveal belief-desire pairs to be sufficient for genuine agency.

As an example of action caused by the agent's belief-desire pairs, Searle offers the case of the heroin addict whose overpowering urge to take his drug compels him to do so even as he judges that he should not. By contrast,

> In the normal case of rational action, we have to presuppose that the antecedent set of beliefs and desires is not causally sufficient to determine the action. This is a presupposition of the process of deliberation and is absolutely indispensable for the application of rationality. We presuppose that there is a gap between the "causes" of the action in the form of beliefs and desires and the "effect" in the form of the action. This gap has a traditional name. It is called "the freedom of the will" (13).

According to Searle, the vast majority of actions are caused not only by the agent's pre-existing beliefs and desires, but also by free will. That the agent wills one action or another is not causally determined by facts about the agent's antecedent psychological states. On Searle's view, antecedent facts about the agent's psychology are not sufficient to determine the way in which her free will is exercised.

According to Searle, we have experiences of freedom of the will in the phenomenon of the "gap".

> The gap can be given two equivalent descriptions, one forward-looking, one backward. Forward: the gap is that feature of our conscious decision-making and acting where we sense alternative future decisions and actions as causally open to us. Backward: the gap is that feature of conscious decision-making and acting whereby the reasons preceding the decisions and the actions are not experienced by the agent as setting causally sufficient conditions for the decisions and actions. (62)

I will use Searle's terminology for the experience that he refers to in using the phrase "the gap." In doing this, I should not be regarded as agreeing that the representational content of this experience is exactly as Searle characterizes it. While the "Forward" description seems like an apt description of some part of our experience of action, there is reason to doubt whether the "Backward" description correctly characterizes the content of the experience.

It is important to note that the "Forward" and "Backward" descriptions of the representational content of our experiences of the gap are not equivalent to each other. On its most natural reading, the Forward description is consistent with the possibility that the agent's future decisions and actions are causally determined by the agent's beliefs and desires, and that the agent is reducible at least in part to these psychological states. Then, even though our beliefs and desires causally determine our actions, our actions might be "causally open to us" in a way that a compatibilist account of free will would support. If we desired otherwise, we could do otherwise, and the causal determination of our actions by our pre-existing beliefs and desires would not prevent our futures from being "causally open to us," at least if the last two words of the phrase "causally open to us" are taken seriously. The Backward description, however, excludes this compatibilist possibility. It contains the stipulation that the agent's antecedent psychological states cannot constitute causally sufficient conditions.

To help us see what the gap is, Searle describes an experiment conducted by Wilder Penfield, in which Penfield stimulated the motor cortexes of his patients with a microelectrode and caused them to make various bodily movements. When asked, the patients denied that the bodily movements counted as their actions, saying "I did not do that, you did it" (64). Searle says that "the patient's experience, for example, of having his arm raised by Penfield's stimulation of the brain is quite different from his experience of voluntarily raising his arm" (64). In the cases where Penfield stimulates the patient, the patient observes his arm moving, while in the normal case, the patient makes the movement happen. It is at least some part of the difference between the two cases that in the normal cases, the patient does not regard the motions as causally open to him. No matter what he wishes, they will occur anyway.

If this example is merely meant to help us get clear on the component of our phenomenology that Searle is calling "the gap", there is no reason to quarrel with it. But it seems that Searle wants the example to do another kind of work as well. He concludes the paragraph where he describes Penfield's experiment with a description of what happens in the normal case when someone raises his arm. "First, I cause the bodily movement by trying to raise my arm. The trying is sufficient to cause the arm to move; but second, the reasons for the action are not sufficient causes to force the trying" (64). Searle is trying to use the causal sufficiency of motor cortex stimulation for motion that is not action in the Penfield case to argue for the causal insufficiency of our antecedent psychological states for action in the normal case.

Penfield's experiment, however, does not show or in any way suggest that this is how the normal case goes. Motor cortex stimulation is not the sort of thing that gives agents new reasons to act. This is true even on a Humean theory where motivating reasons are constituted by the agent's desires and beliefs. Stimulating an agent's motor cortex would not cause her to desire or believe anything new. So if motions caused by motor cortex stimulation are not regarded as actions by the people who make them, that has no bearing on the truth of the Humean theory of motivation.

Tim Schroeder's work on the neuroscience of desire bears this out.[21] The VTA/SNPC, where Schroeder says that desire is realized, is neurobiologically far upstream from the motor cortex. While the route by which desire causes action goes through the motor cortex, one cannot give the agent a new desire by doing things to the motor cortex, just as one cannot give the agent a new desire by doing things to the spinal cord. Stimulation of the motor cortex will not have many of the effects that desire necessarily or actually does

[21] *Three Faces of Desire* (2004) and "Moral Responsibility and Tourette Syndrome," *Philosophy and Phenomenological Research*, July 2005.

– for example, it will not generate the counterfactual connections to pleasure that are necessary for desire, or the direction of attention that is contingently part of it. All it will do is cause motion. So Searle should not claim that the subjects in Penfield's experiment acted because of their beliefs and desires, and that only volition was missing. (From what Searle writes, it is not clear whether he intends to make this claim or not. While he does not explicitly endorse this claim, his argument from the Penfield experiment to the insufficiency of desires and beliefs for action depends on it.) Penfield's experiment does not involve any of the agent's beliefs or desires, and thus is no more interesting in the debate between Searle and the Humeans than a case where the experimenter grabs the patient's arm and lifts it. Neither party to the debate would consider that a case of action, and neither party is committed to saying that Penfield's experiments involve action either.

Searle considers the position that our experiences of the gap are illusory, and that our antecedent psychological states are sufficient causes of action despite the fact that it seems to us, when we act, as if multiple alternative possibilities are open to us. Against this position, he claims that "we have to presuppose that there really is a gap, that the phenomenology corresponds to a reality, whenever we engage in choosing and deciding, and we cannot give up choosing and deciding" (71). According to Searle, there is a "practical inconsistency" in holding both of the following theses:

1. I am now trying to make up my mind whom to vote for in the next election.
2. I take the existing psychological causes operating on me now to be causally sufficient to determine whom I am going to vote for.

The idea is that sincere belief in (2) would make the effort in (1) unnecessary. Searle thinks that someone who genuinely believed (2) would, rather than acting, sit back and let the existing psychological causes move him.

One only sees a “practical inconsistency” between (1) and (2) if one has an antecedent commitment to Searle’s view that the agent is irreducible to his psychological states. If one assumes that the agent is separate from the psychological causes of his actions, it looks curious that the agent is trying to accomplish something that is already going to be accomplished no matter what by his psychological states. There may be something irrational about acting to bring about some state of affairs that you know will occur whether you act or not. If the agent is distinguished from his psychological states, and only the operations of the former are regarded as actions, that is how actively deciding how to vote in this situation will look.

But a Humean picture on which the agent is reducible to his psychological states makes this problem fall away. Consider what would actually happen with an agent who decided to just sit back and let his psychological states move him. At some point, when it came time to act, his beliefs and desires would take over, and he would make his decision. If the Humeans are right about how beliefs and desires contribute to action, this would be an action like any other, caused by beliefs and desires as actions always are. Perhaps an hour before the polls closed, he would be reminded of the election by the television news and realize that time was running out. His desire to vote for the better candidate would become occurrent, and the possibility of missing the election entirely might make him feel anxious about the fact that he had not yet voted. His desire to vote for the better candidate would cause him to direct his attention to things relevant to its object, like the candidates' policy positions. Scrutinizing these policy positions, he would come to realizations about which candidate he agreed with on more issues. Seeing this, he would know who to vote for, and form an intention to vote for that person.

At some point in this process, his belief in (2) would cease to be occurrent, because he would be busy thinking about how to vote in the election. But there is no reason to think he would cease to have the belief in dispositional form at any point during the process, accept anything that contradicted it, or find himself acting to bring about a situation that will come about whether he acts or not. For him to choose and act is simply for his psychological states to operate in their usual choice-driving and action-causing way.

Some of Searle's arguments against Humean accounts of the nature of the self are independent of his view that our experiences of the gap correctly represent our free will. These arguments deal in large part with the semantics of action-explanations. Since these arguments do not have anything to do with the nature of desire, the particular psychological states that cause action, or the phenomenology of deliberation, I will not consider them here.

Searle does, however, seem to think that our experiences of the gap provide us with sufficient reason to accept non-Humean accounts of the self. As he says in a fully italicized sentence, "*The intelligibility of our operation in the gap requires an irreducible notion of the self*" (74). It is this claim that I have argued against. While we do have experiences of the gap according to the Forward description – experiences where our future decisions seem causally open to us – explaining these experiences and the decisions that we eventually make does not require an irreducible notion of the self, or even require the agent to accept one. Neither do our experiences support any claims about the causal insufficiency of our antecedent psychological states for action.

### Searle on *Akrasia*

Searle argues that Humean theories, 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 she 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 be provide a better explanation.

In cases of *akrasia*, an agent's judgment about what to do differs from how she acts, even at the moment of action. 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" (220). This is the same class of views he was criticizing in his discussion of the gap, and Humean theories fall within the scope of his criticism.

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" (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. (233-234)

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 she 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.

But if there is a problem that this account solves, it is not the problem of *akrasia*. The problem 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 so often 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 her mind at the last moment and decides to do something she did not plan to do before.

The description of the heroin addict who compulsively takes his drug, which Searle presents early in his book and which I referenced in the previous section, 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. In doing this, I will rely mainly on facts about desire that I presented in the previous chapter. I hope to show that the issue of *akrasia* is not a weakness for the Humean theory, but a strength. Once we understand how the Humean theory describes desire's interactions with other mental states, 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 AM. 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 exciting sports highlights on the screen and the amusing commentary of the SportsCenter announcers have me in 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 her environment is such that she has vivid sensory or imaginative representations that relate to the object of one desire. At the same time, she 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, as Hume would say, 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 her attention to various possibilities as she makes her judgment about what she ought to do. Though her calm passion is too weak to overpower the violent passion and determine her behavior, the calm passion is still able to determine the course of her reflection. As I discussed in the previous chapter, vivid images are especially powerful in increasing the motivational and emotional force of a passion, but they do not give an equal boost on all of the passion's effects. The agent's reflection and judgment are dominated by one desire, while her behavior is dominated 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 itself (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 towards it. This will make that desire more capable of driving akratic behavior.

Now I will consider an objection to my proposal, which threatens to drive a wedge between Humeans about motivation and Humeans about practical rationality.[22] It is an objection that Christine Korsgaard advances against this combination of Humean views in "The Normativity of Instrumental Rationality":

[22] By calling this view the "Humean theory of practical rationality", I do not mean to assert that it is the view that Hume himself actually held. Elijah Millgram's "Was Hume a Humean?" provides interesting arguments that Hume might have thought otherwise.

> Hume identifies a person’s end as what he wants most, and the criterion of what the person wants most appears to be what he actually does. The person’s ends are taken to be revealed in his conduct. If we don’t make a distinction between what a person’s end is and what he actually pursues, it will be impossible to find a case in which he violates the instrumental principle. (230)

Korsgaard offers a general criticism of theories – like the Humean theory of practical rationality when combined with the Humean theory of motivation – that present us as guided by norms that we cannot violate: “how can you be guided by a principle when anything you do counts as following it?” (229)

*Akrasia* provides a case where the force of Korsgaard’s argument becomes clear. If *akrasia* occurs when vivid representations make one desire strong enough to drive behavior, then it seems that akratic action is just another case where the agent acts on her strongest desire. And if the Humean theory of practical rationality is correct, acting on one’s strongest desire (at least in cases where only two desires are relevant to the situation, and we are certain of getting what we choose) will be rational. So this combination of Humean theories implies that akratic actions are rational. But this is deeply counterintuitive – akratic actions are supposed to be among the paradigm cases of irrationality. So it seems that my account of *akrasia*, when paired with a theory of practical rationality preferred by many people who are also Humeans about motivation, will make paradigmatically irrational behavior look rational.

The key to answering this objection is to see that a Humean theory of practical rationality need not – and should not – run along exactly the same lines as the Humean theory of motivation. There are ways to build a Humean theory of practical rationality that sometimes endorses different actions than those predicted by the Humean theory of motivation. For example, in looking at desire strengths to determine which action is rational, one might only take into account the strengths of desires under conditions where the agent is presented with equally vivid representations of the various things she desires. Then the fact that my desire to stay in bed is temporarily stronger than my desire to teach well, merely because its object is more vividly represented at the time, would not make it rational to stay in bed. To determine the rationality of staying in bed, however, we should not look at how strong the desires were at that particular moment. Instead, we should consider how things looked when the images associated with both desires were equally distant from me – perhaps, the way they looked the previous afternoon as I made my plans for the next day. Different actions would be endorsed by a Humean theory of practical rationality that worked with desires under conditions of equally vivid information than those predicted by the Humean theory of motivation as I have spelled it out. This would leave room for agents to act irrationally. A theory of practical rationality that calculated the rationality of actions by looking at desires under conditions of equally vivid imagination would still be distinctively Humean, as its judgments of actions would be grounded in the agent's desires and means-end beliefs.

This combination of Humean theories accords with our pretheoretical judgments about how common rational action is. It accounts for the truth of our pretheoretical belief that people act irrationally only in a minority of cases, since the general kinds of mental states that explain action – desires and means-end beliefs – are the same ones that justify it. The process by which these psychological states cause actions is geared towards producing rational actions, most of the time. But there is room for irrationality, in cases where one of the desires has become violent because of a vivid sensory or imaginative representation of a desired object (and perhaps in other cases as well). Thus the theory can also explain the truth of our pretheoretical belief that people sometimes act irrationally.

This way of setting up the Humean theory of practical rationality should accord fairly well with our intuitions about rationality. While it often happens that agents sacrifice their strongly desired but distant goals for smaller satisfactions that are right before them, it is part of our idea of a rational agent that she will not act in such shortsighted or impulsive ways. A Humean picture of motivation can account for the behaviors of rational agents as well. Presented with a nearby satisfaction that would require her to sacrifice a strongly desired goal in the future, the rational agent may – either as an automatic process driven by her desire for the strongly desired goal, or as an act of volition instrumentally generated by this desire – turn her attention to that goal. This is something that an irrational agent would not do. As she imagines that goal, her desire for it will become more violent, giving her sufficient motivation to set the nearer satisfaction aside. One thing that contributes to an agent's rationality, then, is having a psychological tendency to keep the big picture in mind, thinking of one's distant but strongly desired goals when one might have otherwise sacrificed them for a lesser satisfaction. The Humean theory offers us a simple explanation of how this tendency operates.

### Korsgaard and the choosing of aims

So far in this chapter, I have concentrated on explaining phenomena that opponents of the Humean theory cite and which I also accept. Rather than denying their descriptions of the surface phenomena, I have tried to show that Humean explanations offer a better account of these phenomena. For example, I took Darwall's story about Roberta, as he presented it, and showed that the Humean theory had a good explanation of her experiences and history. Now I will turn to an area where the anti-Humean accounts of motivational psychology conflict with the surface phenomena, and where this should be decisive in favor of Humean accounts of motivation. The target of my criticism will be Christine Korsgaard's view that we can choose to enter into new motivational states when we choose the aims for which we act. I will argue that this view is not correct, and the way of generating new motivational states that she describes is not psychologically possible for human beings.

In her *Locke Lectures*, Korsgaard argues for an anti-Humean position according to which we can choose not only what we do, but also the motivational force that causes us to do it. Her way of expressing this view involves a distinction between "acts" and "actions." An action involves both an act and an aim. Making a false promise and committing suicide are examples of acts, while making a false promise to get money and committing suicide to avoid misery are examples of actions, since they include the aims for which the acts are performed – getting money and avoiding misery. According to Korsgaard, "it is the action that is strictly speaking the object of choice" (1.2.5). She cites Aristotle as a predecessor of hers, saying it is his view that "that the *aim* is included in the description of the action, and that it is the action as a whole, *including the aim*, that the agent chooses" (1.2.4). Kant is also cited as holding this view, with the combination of act and aim constituting the maxim that the agent accepts in deciding to act. Korsgaard says that the form of a Kantian maxim is "I will do Act-A in order to promote end-E" (1.2.5). With "Act-A in order to promote end-E" being a description of what Korsgaard calls an "action", accepting a maxim is choosing an action.

Korsgaard is not alone in holding the view that agents can choose their aims. John Searle agrees. According to Searle, "when one has several reasons for performing an action, one may act on only one of them; one may select which reason one acts on" (65). He offers a case in which someone has several reasons for voting for a particular candidate, and claims that one can vote for the candidate for one of these reasons, but not for the other reasons that he has.

While Korsgaard accepts an anti-Humean view of human action, she has what appears to be a Humean view of animal action. According to her, "an animal does not choose the principles of its own causality – it does not choose the content of its own instincts. We human beings do choose the principles of our own causality – we choose our maxims. And the categorical and hypothetical imperatives are rules for doing this – rules for the construction of maxims. It is because we, unlike the other animals, must choose the laws of our own causality that we are subject to imperatives." (3.6.1). Korsgaard's use of "instinct" seems fairly close to the ordinary use of "desire" – she sets aside the major difference between the two terms, saying that she does not intend any contrast between "instincts" and motivational states that can be learned. (4.2.2). She allows that animals are capable of basic forms of intelligence that allow them to attain new desires through processes of conditioning, and also through instrumental reasoning. These are all psychological processes that a Humean theory will happily admit.

Aims, on Korsgaard's view, seem to be something like the contents of motivational states. Choosing our maxims is choosing not only what act to perform, but choosing the "principles of our own causality." The principles of an animal's causality, she says, are its instincts. So the principles of a creature's causality seem to be its motivational states. When we choose our maxims, then, we are choosing which motivational states to enter into. And since choosing a maxim involves choosing an act and an aim, it makes sense to identify the choice of an aim with the generation, by choice, of a new motivational state.

As Korsgaard regards the categorical imperative as a rule that humans can follow in choosing maxims, it is clear that she is some sort of anti-Humean about human action. Following the categorical imperative to choose how one will be motivated must be regarded as a process of reasoning. And since the categorical imperative applies to an agent regardless of what desires she has, the generation of new motivational states by following the categorical imperative is barred by the Humean theory. The character of her anti-Humean view depends on her theory of motivational psychology. If the choice of motivational states that occurs when one chooses a maxim involves a choice of desires, Korsgaard denies DODI in claiming that new desires can be generated by processes of reasoning that do not involve pre-existing desires. But if the choice of motivational states in this case does not involve a choice of desires, she admits that humans can be motivated in the absence of desires, and thus denies DBTA. Since she does not use the term "desire" in spelling out her theory, or any clearly equivalent expression, it is unclear which part of the Humean theory she denies. It may be that she denies both.

Now I will begin my criticism of Korsgaard's claim that agents choose not only their acts but the complex of their acts and their aims. Often when agents would like to act for one aim rather than another, they find themselves unable to act for the preferred aim. Suppose that Bob has received his draft notice from the army. He is very worried about the consequences to his future political career if he dodges the draft. But he would like to be the sort of person who obeys his nation's call out of a proud desire to do his duty, and not for craven political reasons. If he could choose his aims, he would gladly choose to serve in the military out of duty. But since he lacks a sufficiently strong desire to do his duty, he will join the military out of a desire to protect his political career, and there is nothing he can do to make it the case that he joins the military with a different aim. What makes it the case that we have certain aims is that the relevant desires drive our action. If we lack these desires, if they are not strong enough to drive our actions, or if they fail to operate for some reason, there is nothing we can do about it.

That may put the point too strongly – sometimes there are things we can do to make some aim our own, but they are not the kinds of things that would be useful to Korsgaard's account. For instance, we can try to engage with stimuli that strengthen some of our desires. As he walks into the Army recruiting station, Bob may look at a photo of the Iwo Jima memorial that he carries in his wallet or think of an older relative's military heroism in an effort to stimulate his desire to serve his country. In doing this, Bob would be trying to intensify his desire by vivid images, as was discussed in the previous chapter. But this does not count as choosing an action – choosing an act and the aim for that act. Whether he employs an external stimulus or directs his attention by an act of will, he is not choosing the aim for the act that he is choosing at the moment – he is trying to affect his desires so as to determine the aim of some future act. The object of choice is still the act, even if it will have causal consequences for some future aim.

There is, then, an asymmetry between acts and aims. Often we choose between two acts to accomplish the same aim. When I play tic-tac-toe, I may choose between the act of putting my X in the corner and the act of putting it in the center. Either way, I have settled on the aim of winning the game. But we do not settle on an act, and then choose which aim we should do it for. So it is hard to see why we should regard "actions", in the aim-including sense in which Korsgaard uses the term, as the objects of choice, with new motivational states being generated by choosing. All deciding what to do involves choosing between acts, and making "actions" the objects of choice by attaching aims to acts is philosophically unmotivated.

Korsgaard (and anyone who follows her in building an anti-Humean theory on which agents can affect their motivational states through a simple act of choice) will have to find some way to explain why we are so often dissatisfied with our stock of motivational states, and why our dissatisfaction is so often impotent. I wish that my desire to work hard were stronger. Then I would work harder, be happy doing it, and be more successful. The overarching goals that drive my life would stand a much greater chance of being achieved. If there were some way to strengthen these desires through the processes of reasoning that Korsgaard says that we differ from animals in having, I am fairly sure that I would have met with some success in strengthening my desires through those processes. I do not see what resources Korsgaard's view has to explain why I am never able to successfully carry out this activity. On her theory, I am distinguished from the animals by my ability to perform it. But as it stands, the only methods I have to strengthen my motivation are methods similar to the ones I described Bob as using. I remind myself of the need to work by changing my desktop background to a picture of my dissertation co-chair, whose image brings up associations to my academic goals (and to the possibility of being rebuked if I am lazy). And when I find my mind wandering, I force myself to imagine the rewards of work and the costs of indolence. The short and simple internal route of strengthening my desire by choosing my aims is closed to me.

The Humean theory, on the other hand, accounts for my inability to change my motivational states perfectly well. According to the Humean theory, desires can only be generated through reasoning by instrumental processes. They can be strengthened this way too – if I previously had a purely gustatory desire for coffee, I may come to want it more upon realizing that its caffeine will allow me to be more alert during class. A second-order desire like the desire to desire to work hard, however, has no way of transmitting its force directly to the first-order desire that it focuses on. The only means-end beliefs by which I can transmit the motivational force to the first-order desire involve external processes like the ones I described above. (For example, the belief that by making my advisor's face my desktop background, I will strengthen my desire to work harder.)

Korsgaard presents a few arguments to support her claim that we choose actions rather than acts. Perhaps the most substantial one, which runs through several of the lectures, is that a view on which we choose actions rather than acts will help explain how we can be morally responsible. "Because we are self-conscious, and choose our actions deliberately," she argues in Lecture 1, "we are each faced with the task of constructing a peculiar, individual kind of identity – personal or practical identity – that the other animals lack. It is this sort of identity that makes sense of our practice of holding people responsible" (1.3.3). In the fourth lecture, she says that "it is because our actions are expressive of principles we ourselves have chosen, principles we have adopted as laws of our own causality, that it makes sense for us to hold one another answerable in this way" (4.5.4).

It is not clear, however, how allowing an agent to generate her own motivational states by an act of choice opens the door to any sort of distinctive theory of how we can be morally responsible. Korsgaard evidently means to offer a compatibilist account, and her view bears some similarities to that of Harry Frankfurt in "Freedom of the Will and the Concept of a Person." According to Frankfurt, free action requires not only a first-order desire that motivates a particular action, but also a second-order desire for the first-order desire to be effective. With Korsgaard, free action requires more than the first-order desire as well – it requires some kind of higher-order mental state or process by which the first-order desire (or whatever psychological state embodies the agent's aim on Korsgaard's preferred theory of motivation) will be chosen. So it seems that Korsgaard's view may be subject to the same objection that is often regarded as decisive against

Frankfurt: if first-order mental states are unable to make action free, why should we think that higher-order mental states are able to do so? As Tim O'Connor summarizes the point, referencing Gary Watson, "Why suppose that they inevitably reflect my true self, as against first-order desires?"[23] Korsgaard's account differs from Frankfurt's in that the higher-order mental states are of a different kind than the lower-order ones, but it is not clear why this would be of any real benefit in developing an account of free action.

## Conclusion

Near the end of *Desire*, G.F. Schueler says,

> A very plausible research project, represented in this book by Dretske's views, hopes to extend the blind-forces model to cover all cases of reasoning and intentional action, including the ones for which we use reflective explanations and that, I have argued, are amenable only to explanation only by means of the deliberative model. (195)

This project – the project of covering all cases of deliberative reasoning with a simple Humean model – is the one that I have tried to advance in this chapter. Even on a strong notion of desire where an agent's being motivated does not make it a conceptual truth that a desire is present, it does not seem that the actual world contains any genuine counterexamples to the Humean theory of motivation. The theory offers elegant and powerful explanations of our deliberation and action in terms of desires and other familiar mental states coming together.

As I argued in the first chapter, the truth of the Humean theory throughout the space of human psychological possibility forces the difficult choice between cognitivism and

[23] O'Connor's Stanford Encyclopedia article on Free Will

internalism upon us.  Theorists who desired to hold both of these views may be disappointed (or worried) by this conclusion.  But it is a conclusion to which our Humean nature forces us.

## Bibliography

Darwall, Stephen, *Impartial Reason*. Ithaca: Cornell University Press, 1983.

Darwall, Stephen, "Reasons, Motives, and the Demands of Morality". In *Moral Discourse and Practice,* Darwall, Gibbard, and Railton, eds., 1997.

Darwall, Stephen, with Allan Gibbard and Peter Railton, "Towards Fin De Siecle Ethics". In their *Moral Discourse and Practice*, Oxford: Oxford University Press, 1997.

Davidson, Donald, "Actions, Reasons, and Causes". *Journal of Philosophy* 60: 685-99, 1963.

Dretske, Fred, *Explaining Behavior*. Cambridge: MIT Press, 1988.

Frankena, William, "Obligation and Motivation in Recent Moral Philosophy". In *Perspectives on Morality: Essays of William Frankena*, Kenneth Goodpaster, ed., Notre Dame: Notre Dame University Press, 1976.

Frankfurt, Harry, "Freedom of the Will and the Concept of a Person". *Journal of Philosophy* 68: 5-20, 1971.

Hume, David, *Treatise of Human Nature*. L. A. Selby-Bigge, ed., Oxford: Clarendon Press, 1888.

Kant, Immanuel, *Critique of Practical Reason*. In *Practical Philosophy*, Mary Gregor, trans. and ed., Cambridge: Cambridge University Press, 1996.

Katz, Leonard, "Pleasure". In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta ed. URL = http://plato.stanford.edu/entries/pleasure/

Korsgaard, Christine, *Locke Lectures*, 2002. URL = http://www.people.fas.harvard.edu/~korsgaar/#Publications

Korsgaard, Christine, "The Normativity of Instrumental Reason". In *Ethics and Practical Reason*, Garret Cullity and Berys Gaut, eds., Oxford: Clarendon Press, 1997.

McDowell, John, "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society*, 52:supplement, 13-29, 1978.

Millgram, Elijah. "Was Hume a Humean?" *Hume Studies* 21:1, 75-93, 1995.

Montague, Michelle, “Against Propositionalism”. *Nous* 41:3, 503-518, 2007.

Nagel, Thomas, *The Possibility of Altruism*. Princeton: Princeton University Press, 1970.

Nietzsche, Friedrich, *Daybreak.* R.J. Hollingdale, trans., Maudmarie Clark and Brian Leiter, ed., Cambridge: Cambridge University Press, 1997.

O’Connor, Tim, “Free Will”. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, ed. URL = http://plato.stanford.edu/entries/freewill/

Pettit, Philip, and Michael Smith, “Backgrounding Desire”. *Philosophical Review* 99, 565-592, 1990.

Scanlon, T. M., *What We Owe To Each Other*. Cambridge: Harvard University Press, 1998.

Schroeder, Timothy. “Moral Responsibility and Tourette Syndrome,” *Philosophy and Phenomenological Research* 71:1, 106-123, 2005.

Schroeder, Timothy. *Three Faces of Desire*. Oxford: Oxford University Press, 2004.

Schueler, G. F., *Desire: Its Role in Practical Reason and the Explanation of Action*. Cambridge: MIT Press, 1995.

Schurman, J. G., The Consciousness of Moral Obligation”. *Philosophical Review* 3:6, 641-654, 1894.

Searle, John, *Rationality in Action.* Cambridge: MIT Press, 2001.

Smith, Michael. *The Moral Problem*. Oxford: Blackwell, 1994.

Sorley, W. R., *Moral Values and The Idea of God*. Cambridge: Cambridge University Press, 1919.

Strawson, Galen, *Mental Reality*. Cambridge: MIT Press, 1994.

Thorndike, Edward, *Animal Intelligence*. Darien, CT: Hafner, 1911.

Van Roojen, Mark, “Moral Cognitivism vs. Non-Cognitivism”. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, ed., URL = http://plato.stanford.edu/entries/moral-cognitivism/

Watson, Gary, “Free Action and Free Will”. *Mind* 96, 145-172, 1987.

Wedgewood, Ralph, “Practical Reason and Desire”. *Australasian Journal of Philosophy* 80:3, 345-358, 2002.

# VITA

Neiladri Sinhababu was born in Lawrence, Kansas on March 14, 1980, the son of Achintya Kumar Sinhababu and Pranati Sinhababu. He received his B.A. in Philosophy from Harvard University in 2001 and began work on his Ph.D at the University of Texas that fall. He was a visiting graduate student at the University of Michigan from 2004 to 2005, and received a Newcombe Fellowship in 2006. He is (with Brian Leiter) the editor of *Nietzsche and Morality*, published by Oxford University Press in 2007, and the author of “Vengeful Thinking and Moral Epistemology” in that volume. He is also the author of "Possible Girls", forthcoming in the *Pacific Philosophical Quarterly*. In July 2008 he will take a position as Assistant Professor at the National University of Singapore.

Permanent Address: 1329 Vancouver Avenue, Burlingame, CA 94010

This dissertation was typed by the author.