Chapter 1 draft: What Humeans Say About Desire
Neil Sinhababu

Table of Contents

This paper is longer than the other papers we’ve read in the dissertation working group. Don’t feel compelled to read the entire paper unless you feel like it. The paper breaks down pretty nicely into self-contained sections, most of which deal with one particular philosopher and one set of issues, so feel free to read only the sections that seem interesting to you. I do, however, offer to buy a beer for anyone who reads the whole thing for Wednesday’s discussion.

Table of Contents  (You’re reading this)  1
Introduction  (everyone should read this to get an idea of what’s going on)  2
Hume, Passions, and Vivid Imagination  4
Mill and Pleasure  10
Nietzsche, Direction of Attention, and the Standpoint of Reflection  13
Davidson’s Pro-Attitudes and the Objects of Desire (p.18 is widely seen as amusing)  15
Strawson and the Connection to Pleasure  20
Schroeder’s Reward Theory (a little bit of neuroscience here, it’s not too hard though)  26
Actual Properties of Desire (sort of summarizes everything, will be contentious)  41
The Analysis (probably even more contentious than the previous section)  45
Works Cited  48
Introduction

Among the issues that have concerned philosophers of action are these three: How is the concept of action to be analyzed? What is the nature of a reason for action? What kinds of mental states cause human beings to act? This dissertation will focus on the third of these questions. In particular, I will try to discover what kinds of mental states explain our behavior and our experiences when we deliberate about what to do.

Correctly understanding the way that human action is caused has been of interest to philosophers who work on the other two questions. Some philosophers working on the first question regard it as a conceptual truth that actions are events caused in a certain way by some particular set of mental states. If we are unable to locate those mental states among the causes of an actual event that we intuitively regard as a human action, we will have a counterexample to their theories. Some philosophers working on the second question have argued that having a reason for action requires the presence of particular kinds of motivational elements in an agent’s psychology. This claim, combined with a claim about the motivational elements that actually influence our deliberations and cause our actions, implies a claim about what we have reason to do.

The theory I will defend in this dissertation is a Humean theory. Humean theorists about motivation stress the importance of desire in explaining how actions are caused. They accept the following two claims about human motivation:

**The Desire Theory of Motivation**: Whenever someone performs an action, the action must be caused, at least in part, by one of her desires.

**Skepticism About Noninstrumental Reason**: If a desire is formed as the result of a process of reasoning, other desires must enter into the reasoning.
Philosophers who deny one or both of these claims are rationalists. Skepticism about noninstrumental reason is the more often denied part of the Humean theory, though some rationalists deny both parts. This dissertation will focus mainly on deliberation as opposed to action itself. Vindicating skepticism about noninstrumental reason against objections – particularly, objections from rationalists who hold that Humeans cannot account for the phenomenology of deliberation in some cases where we reason about what to do – will be the central focus of future chapters.

This chapter, which focuses on desire, has four aims. The first is historical. To map out the conceptual territory and position my view in relation to the others that have been put forward, 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. In doing this, I will accomplish a second goal – mining their writings for material that will be useful in overcoming the rationalist objections 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. Their explanations often involve interesting interactions between desires and other mental states, and this brings me towards the third goal of this chapter – to characterize the way in which desire interacts with other mental states and mental phenomena, such as belief, pleasure, imagining, reinforcement learning, 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 fourth goal of this chapter is to build towards an account of the necessary and sufficient conditions for a mental state’s being a desire. 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. At the end of this chapter, I will arrive at a set of individually necessary and jointly sufficient conditions for desire.

One terminological point ought to be made before I begin. Some of the philosophers to be discussed – in particular, Hume and Nietzsche – discuss the role of “passions” and “drives” in motivation, and do not explicitly discuss “desires” in the passages that I will cite. Context does not suggest any reasons to regard these terms as inequivalent to “desire.” (The German word “Trieb,” in fact, is sometimes translated as “desire.”) So I will regard Hume’s “passions” and Nietzsche’s “drives” as the same things that modern philosophers call “desires.”

**Hume, Passions, and Vivid Imagination**

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 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 has laid out two of Hume’s arguments for this conclusion. The 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 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. Rationalists will 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 has
some kind of representative property – perhaps, that it aims at the good, and that reason can affect desires in virtue of their representing states of the world as good\(^1\).

A second response, perhaps more effective, is to argue that reason is not limited to evaluating mental states in terms of truth and falsity. This response is implicit in the position of Aristotle, who holds that reason can evaluate potential practical syllogisms and determine the rationality of performing the actions that are their conclusions; and taken by instrumentalists, who hold that reason can take our desires and beliefs into view and determine whether we have a desire-belief pair that would make an action rational. One virtue of the Humean account of reason, however, is that it gives a unifying account of why reason can affect some mental states – where truth is at stake, reason can play a role, and where truth is not at stake, reason has no role to play. Showing which unifying feature of beliefs and motivational states explains their susceptibility to rational guidance is a challenge for Aristotelians and instrumentalists\(^2\).

At this point, I would like to address a way in which empirical disputes about the nature of our motivational mental states interact with normative disputes about practical rationality. If a theory is better for giving a unified account of what makes a mental state subject to rational criticism – that is, assessment as rational or irrational – theorists interested in arguing that a particular mental state is apt for rational criticism should be interested in the descriptive properties of that mental state. If it turns out that the mental state in question shares a particular property with other mental states agreed to be apt for rational criticism, it may be possible to cite this property as the one in virtue of which the mental state is apt for rational criticism. The effect of this move will be to show that the aptness of the mental state in question for rational criticism

---

\(^1\) David Velleman deals with objections of this kind in “The Guise of the Good.”

\(^2\) Certainly, there are plenty of responses to this challenge, for instance Allan Gibbard’s view that normative judgments of the rationality of a mental state are appropriate when it is possible to agree or disagree with the mental state.
is a consequence of a suitably unified theory of which mental states can be seen as rational or irrational.

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, 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.

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. Rationalists 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 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.

Mill and Pleasure
In *Utilitarianism*, John Stuart Mill advances a few different theses about the relationship between desire and pleasure. Mill avoids psychological hedonism, the view that people are only motivated to pursue increases in pleasure. Shortly after saying similar things about virtue, health, and music, Mill accepts that “money is, in many cases, desired in and for itself” – “not for the sake of an end, but as part of the end” (46). What he actually says about the relationship between desire and pleasure is more complicated than psychological hedonism, and much closer to being correct. He claims that “practiced self-consciousness and self-observation, assisted by observation of others” will show that desiring a thing and finding it pleasant, aversion to it and thinking of it as painful, are phenomena entirely inseparable or, rather, two parts of the same phenomenon – in strictness of language, two different modes of naming the same psychological fact; that to think of an object as desirable (unless for the sake of its consequences) and to think of it as pleasant are one and the same thing; and that to desire anything except in proportion as the idea of it is pleasant is a physical and metaphysical impossibility. (49)

First I will focus on Mill’s claim that we feel some pleasure in thinking of something that we desire. Then I will consider Mill’s intriguing remarks about the epistemic and modal status of the link between being pleased at the thought of some state of affairs and desiring that it come about.

Psychological hedonism is clearly incorrect – people occasionally desire things that are not themselves pleasurable, and the intensities of their desires often fail to correspond with their beliefs about how pleasurable the desired states of affairs will be. However, as Mill points out, thoughts about desired states of affairs are accompanied by experiences of pleasure. Thoughts of states of affairs to which we are averse are accompanied by displeasure. The amount of pleasure experienced in imagining a desired state of affairs, or in coming to believe that such a state of affairs will come about, tracks the intensity of desire far better than the pleasure that one believes one will have in the desired state of affairs. This is obvious, as Mill says, from self-
consciousness and self-observation – imagining states of affairs that we desire, as we often do in
daydreams, is a pleasant activity, while vividly imagining states of affairs that we are averse to is
an unpleasant activity. This even applies to unconscious desires and aversions, which are no less
able than their conscious counterparts to trigger feelings of anxiety and pleasure when we
imagine their objects, or things that we associate with their objects. Inexplicable feelings of
anxiety or disappointment may lead me to start wondering whether I am subject to some
unconscious aversion, or whether one of my unconscious desires is being frustrated. What is
unusual about unconscious desires is that we are unable to directly introspect their contents, as
we are able to introspect the contents of conscious desires.

Now I will consider the modal and epistemic status that Mill ascribes to his claims about the
relation between desire and pleasure. It seems that he conceives the relation between desiring
that B and feeling pleasure at the thought of B the same way that Putnam regards the relation
between being water and being H2O – as a necessary \textit{a posteriori} identity. He does not seem to
regard it merely as a contingent identity, as he claims that it is metaphysically impossible to
desire something without finding the idea of it pleasant. The path he suggests for discovering
this truth about desire is not one of conceptual analysis or reflection on the meaning of “desire”,
but the \textit{a posteriori} method of “practiced self-consciousness and self-observation, assisted by
observation of others.” Mill’s status as an early direct reference theorist about proper names and
his comment that desiring something and feeling pleasure at the thought of it are “two different
modes of naming the same psychological fact” suggest that ascribing this view to him is not
excessively anachronistic.

However, it does not seem that Mill has successfully discovered an \textit{a posteriori} identity here.
If desiring that B is necessarily identical to being pleased by thoughts of B, nothing more will
need to be the case for someone to have a desire. And there is something more that is widely regarded as necessary. Perhaps the most immediately obvious necessary condition for some mental state’s being a desire is that it play some role in causing action. In the words of Anscombe, “the primitive sign of wanting is trying to get” (68). Of course, Mill might want to claim here that behavior is the primitive sign of desire in much the same way wetness might be offered as the primitive sign of water – merely a major part of the stereotype of the thing, and not its essence. If the motivational power of desire was somehow explained by pleasure at the thought of the desired thing, just as the wetness of water is explained by the polarity and other properties of water’s chemical structure, an H2O-water-style reduction might be possible. But Mill does not offer a psychological account on which this is the case, and there is no obvious reason to think that such an account is correct.

**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 that is tormenting us. We can avoid opportunities for its gratification, gratify it only on a certain schedule, overindulge it until one becomes 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. Our 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 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 it 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 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 B – whether B 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 B and towards things 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 proposes 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.

One effect of Nietzsche’s move in the dialectic between Humeans and rationalists is much like the effect of Hume’s discussion of calm passions. Rationalists 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, rationalists 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.

Davidson’s Pro-Attitudes and the Objects of Desire

According to Donald Davidson, the “primary reason” for an action consists in a pair of mental states – a “pro attitude towards actions of a certain kind”, and a belief that the 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 the claim that all actions by humans 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 rationalists might want to say that convictions whose content ties them to action in this way can cause actions
without the action 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 Hume and all desire theorists about motivation, who say that desires are necessary if humans are to be motivated to act.

Davidson claims that pro-attitudes are directed towards actions of some 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 an action of that kind. 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 seems to necessarily involve an action for the agent to perform, the content of a desire does not. 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 – 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. Much as Mill suggests, 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.

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 evades 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. 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 necklace, make love to a goldfish, or keep Orlando Bloom in her aquarium. 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.

Much of “Actions, Reasons, Causes” consists in arguments against the position that mental states cannot rationalize action while being causes of action. Since these arguments do not make any essential reference to the kinds of mental states involved, they do not deal in any great depth with the issue that I am addressing.

**Strawson and the Connection to Pleasure**

In *Mental Reality*, Galen Strawson has argued against “neobehaviorism,” claiming that one can understand mental states without understanding how they are related to behavior. In arguing for this position, he offers a view of desire similar to that of Mill, discussed above. 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.

One example that Strawson offers to support his view 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.

I have presented many people with an example involving creatures (the “Neutrals”) with M states but no capacity for pleasure or displeasure. This example covers much of the same conceptual territory as Strawson’s, since pleasure and displeasure play essential roles in making our affective states what they are. Intuitions about whether the Neutrals’ M states were desires were as far from consensus as could be imagined, with respondents equally divided and sometimes strongly disagreeing with each other. What should we make of these differing intuitions? It does not seem implausible to say that we have multiple concepts of desire, some of which require counterfactual connections to pleasant and unpleasant experiences, and some of which do not. If this is the case, the concept of desire that will be analyzed here and brought into the discussions in future chapters is one that has ties to pleasure and displeasure. It is also
possible that there is some vagueness in the necessary conditions for desire. While this may present a problem for my analysis, it will not be a problem for the explanatory work that follows in subsequent chapters, since the mental state to be used in my explanations will be a clear case of desire.

I use a notion of desire on which connections to pleasure are necessary so that I can respond to concerns that the desire-belief theory of motivation is trivially true. If any mental state with motivational power turns out to be a desire, it will be trivially true that all action is caused by desire, granting the plausible assumption that actions must be caused by mental states with motivational power. There is a version of the desire-belief theory of motivation that is more interestingly true – one that builds into the concept of desire some of the key aspects of how desires feel. This version is not trivially true, since it is conceivable for creatures to be motivated without having mental states that counterfactually connect to experiences of pleasure. In view of the unclarity of our intuitions about the concept of desire, I am choosing the version of the concept that leads to the more interesting theory, and the one that is more challenging to defend.

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
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. Perhaps even the necessary connection that Strawson accepts – that desire is necessarily able to combine with belief to bring about 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 support pleasant daydreams and moments of horror at imagined future calamities, or 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 H₂O. 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 should 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 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 hedonic experience 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.

**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 this could happen 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. A similar thing
happens when I learn to properly input a password to enter my email account. When my fingers
slip and I incorrectly input the password, I experience the mild frustration of not being able to
see my mail. But when I properly input the password, I am rewarded by being able to check my
mail. After numerous opportunities for punishment or reward, my fingers tap out the password
in a swift and automatic fashion – I have acquired firm mental dispositions to move my fingers
in one way and not another. That I represent checking my email in a way that allows for this
kind of reinforcement learning makes it true that I have a desire to check my email.

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 feeling 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 – 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 euphorogenic 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 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\textsubscript{2}O 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 H₂O 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 H₂O, 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 a case of this kind. 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 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?

Another counterexample can be deployed to attack the claim that a reward system is sufficient for desiring. Imagine the Creatures of Habit, whose mental lives are largely unlike ours, but who 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. 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. Do the Creatures of Habit have desires, in virtue of the fact that they represent certain states of affairs as rewards?
Intuitively, the Unconditionable have desires, while the Creatures of Habit do not. 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 not be aware 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 candidates for being 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.”³

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 \( \pi \) 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. 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 \( \pi \) change, and it seems
right to attribute a desire for \( \pi \) 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: “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 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.

³ 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.
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: 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 purportedly 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. As Schroeder claims, we want to say that depressed people still desire the same things that they desired when they were not depressed. 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 fully occurrent, 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. Once, desiring to help him, I told Al Martinich that I would administer an exam for him several days later when he was out of town. But when the time for the exam came, my desire to help Dr. Martinich was latent, and to my great shame I completely forgot! Had something reminded me (for example, had I passed my eyes across the books by him on my shelf), perhaps my desire would have been activated by the association, and then I certainly would have administered the exam.

Schroeder’s second objection arises from his theory of what pleasure is. 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 even 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 are presented with 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 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 somewhat 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, and 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 this difference. 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.

Before I finish with Schroeder, I would like to go over 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.

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 is inconclusive, but he argues that the DRN is the place where punishment is realized.

Schroeder makes another useful general point when he addresses 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 have got what we want (or actually, when we believe that we have got what we want). But desires still have psychological effects under these conditions – when we imagine losing the thing we desire and have, we 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. There are also expressions actually using the term “desire” that are used for states of affairs that we know to obtain – sometimes, things are exactly as we desire them to be.

Actual Properties of Desire

Now the time has come to enumerate the properties of desire. I will list them, and then attempt to separate the ones that are necessary for desire from the ones that are only contingently associated with desiring.

Before beginning, it may be useful 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 B and knows that doing A will bring about B, 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 can seem unsatisfactory – situations where an agent has conflicting desires ought to be dealt with as standard 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 blown up by a bomb 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 and about many other things whose causal powers we wish to understand. So the claims that follow should all be understood as restricted to some hard-to-characterize set of situations in which psychological explanations are the ones we seek.

First I will discuss the exact nature of the connection between desire and motivation. Hume and Davidson, among others, gave voice to the commonsense view that desire-belief pairs are necessary for action, and this view is implicit in the following characterization of desire’s motivational role. The following formulation brings in the changes that were suggested in the previous discussion of Davidson’s view.

**The Motivational Component:** If an agent occurrently desires D, and she occurrently believes that she can bring about D by engaging in some behavior B, she will have some motivation to engage in B. Her degree of motivation to engage in B will increase with the strengths of the desire and the belief. If at any time there is some behavior that she is most motivated to engage in, she will initiate that behavior.

The point of using “behavior” rather than “action” in stating the motivational component is to make possible a noncircular characterization of action in terms of belief and desire. While I will not try to defend any view about the concept of action here, it may be regarded as a virtue of this account of desire that it makes such an account of action possible. The term “motivation” 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.
Mill and Strawson have pointed to another component of desire – the hedonic component. Mill regards it as a necessary truth that one feels pleasure at the thought of what one desires, while Strawson is at least sympathetic to the view that pleasure-involving affective experiences are necessary for desiring. To more precisely spell out Mill’s idea of pleasure at the thought of some desired state, we should perhaps consider two kinds of mental states that can cause pleasure in coordination with desire – believing and imagining.

**The Hedonic Component**: If an agent desires that D, the following four things are true of her:

a. Increases in the subjective probability of D will cause her pleasure. The stronger the desire and the greater the increase in subjective probability, the greater the pleasure will be.

b. Decreases in the subjective probability of D will cause her displeasure. The stronger the desire and the greater the increase in subjective probability, the greater the displeasure will be.

c. Vivid imaginative or sensory representations of D will cause her pleasure. The stronger the desire and the more vivid the representations, the greater the pleasure will be.

d. Vivid imaginative or sensory representations of a state of affairs incompatible with D will cause her displeasure. The stronger the desire and the more vivid the representations, the greater the displeasure will be.

I have given reasons for rejecting Schroeder’s view that there is nothing more to pleasure than the representation of desire satisfaction relative to expectation. There are experiences of pleasure that do not represent desire satisfaction. But Schroeder is right to claim that pleasure results when one experiences an increase in expected desire satisfaction. We also experience pleasure
in imagining states of affairs that we desire, as we do when we daydream. Being forced to imagine undesirable states of affairs is unpleasant.

The claim that desire has a role in directing our attention is present, implicitly or explicitly, in the work of Hume, Nietzsche, and Damasio. As Hume points out, desire does not cause us to focus our attention solely on the object of desire, but also on other things we associate with it. This allows it to play a role in bringing together the desire-belief pairs necessary for motivation. One who desires that D focuses her attention on possible ways of getting D – for example, by doing B. As one does this, the belief that one can get D by doing B becomes occurrent, and the agent can be motivated to act. The direction-of-attention component of desire can be expressed as follows.

**The Direction of Attention Component:** All else being equal, an agent who occurrently desires that D will be more likely to focus her attention on things she associates with D than things she does not associate with D.

The ceteris paribus clause is here because this claim needs to be fairly weak. 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 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 – desire can cause an agent to focus one’s attention on non-actual states of affairs as easily as physical objects in her environment.

As Schroeder claims, desiring that D has an important role to play in reinforcement learning. I will draw from Schroeder’s own formulation in characterizing this role.
The Reinforcement Learning Component: When an agent desires D, representations of D will contribute to reinforcement signals in that agent, in the sense made clear by contemporary theories of reinforcement learning.

While reinforcement learning is important to the development of successful organisms, it is not going to play an especially large role in the explanations of the phenomenology of decision-making that will come in subsequent chapters. Reinforcement learning has its characteristic effects across several episodes of decision-making and action, not within any one episode.

The Analysis

So which of the components 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. Assuming that desire can cause action and motivation if the agents are appropriately modified, both the Aldebaranians and the Weather Watchers have desires. But if the cases are constructed so that these counterfactuals do not hold, neither the Aldebaranians nor the Weather Watchers have desires. So it seems that both the Hedonic Component and the Motivational Component are necessary. The Unconditionable have desires, while the Creatures of Habit do not. So the Reinforcement Learning Component seems to be only contingently tied to desire.

How about the Direction of Attention Component? Clearly, directing one’s attention towards certain things is not sufficient for having desires. 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 the hedonic component of
desire? 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 interact 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. 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 the Direction of Attention Component were removed.

Insofar as I can imagine this, though, it does not seem that the Direction of Attention Component is necessary for desire. Imagine one more group of otherworldly creatures – the Absent-Minded. While they are sometimes capable of directing their attention towards things they desire, the systematic connection between desire and attention-direction that exists in humans is absent in them. (I name them the Absent-Minded because imagining an 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.) They are just as motivated by their desires as we are, and when something causes them to fully focus on states of affairs that they desire, their experiences are the same as ours. 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.

Now that these four components have been considered for necessity and sufficiency, an analysis of desire can be offered. Of all the actual properties that our desires have, two are individually necessary and jointly sufficient – the Motivational Component and the Hedonic Component. Intuitions are more firm in favor of the necessity of the Motivational Component, but it seems that there is, at least, some ordinary sense of the term “desire” for which the
Hedonic Component is necessary. The Direction of Attention Component and the Reinforcement Learning Component are neither necessary or sufficient – while they are part of desire in the actual world, a creature need not have the automatic tie between desire and attention-direction or a capacity for reinforcement learning in order to have desires.
Works Cited


