1
What is Desire-as-Belief?

1.1 Desire and Belief

Let's start by contrasting our *wants* with our *preferences*. When you have a preference, you prefer one thing over another. But often we talk more simply about what people want, without comparing that thing with some alternative. I will almost always do the same. Many of my claims are much simpler to state when expressed in terms of what people want, and how strongly, rather than in terms of people's preferences between alternatives. By stating my claims more simply, I hope to make them easier to evaluate. What I say could later be extended, in a natural manner, to make sense of preference.

In fact, this emphasis might reflect the underlying reality. Preferences seem to be more complex attitudes than wants, because they involve comparing two things, rather than assessing only one alone. As a result, it is tempting to think of preferences as complex states, composed of wants: your preference for one thing over another is explained by your wanting the former more than you want the latter (for an argument, see Pollock 2006, 22–7). If our wants are more basic than our preferences in this way, it makes good sense to investigate wanting first, and then later use that theory to develop a corresponding theory of preference. This might provide a further reason for focusing on wants, since preferences are plausibly composed of such wants.

Talk of what we 'want' is somewhat ambiguous: sometimes we mean to refer to what someone wants *most*, and sometimes we mean to refer to what they want *to some extent*. The former is implicitly comparative, like preferences are. Unless I say otherwise, I will always have the latter sense in mind, the sense in which it is clearly true that all of us have many conflicting wants. If I mean to talk about what someone *most* wants, I will explicitly say so.

I will sometimes refer to our wants using the word 'desire'. The word 'desire' in English is often associated with particularly strong desires, and sometimes associated more narrowly still with sexual desires. But the notion I have in mind need not have these associations: in my sense, it is true that I desire to get a good night's sleep tonight. To this extent, the word 'want' is probably a better fit than 'desire' as a label for the state of mind I focus on. But the word

Desire as Belief: A Study of Desire, Motivation, and Rationality. Alex Gregory, Oxford University Press.
 DOI: 10.1093/oso/9780198848172.003.0002

'desire' is useful since it can more naturally be used as a noun, and so I will make use of both 'want' and 'desire', meaning the same by each.

Another important mental state is *belief*. To believe p is to represent p as true. Again, in ordinary English the word 'belief' has distracting associations: it is often used to refer solely to our religious commitments. But as is now standard in philosophy, I will use the word more broadly: we have beliefs about the population of China, about the best theory of gravity, about whether it will rain later today, and so on. In everyday English we use phrases like 'he *thinks*' and 'in her *opinion*' to talk about beliefs, but the word 'belief' is less ambiguous than words like 'think' and 'opinion', and is anyway firmly entrenched in the philosophical literature.

States of mind like beliefs and desires are *attitudes*, and those attitudes have *contents*. The content of a belief is the thing you believe (e.g. that the population of China is large), and the content of a desire is the thing you desire (e.g. that you buy a sledgehammer). For clarity, I will often place the contents of our attitudes in square brackets, so that you might believe [that the population of China is large] and desire [that you buy a sledgehammer]. I normally treat the contents of beliefs and desires as propositions: things that are expressed in English by 'that' clauses, and which can be grammatically preceded by 'it is true' or 'it is false'. But I don't think very much hangs on this assumption, other than ease of presentation.[1] (A little more on this shortly.)

Our beliefs can have many different contents, about just any subject matter. But an especially important subset of our beliefs is the set of beliefs with *normative* contents, such as beliefs about what we ought to do, or about what is good. I'll refer to beliefs with normative contents as 'normative beliefs'. You have very many normative beliefs. Perhaps some of those are moral beliefs, as when you believe you ought to keep your promise to Ahmed, believe that no-one should be cruel, believe that it's good to be a vegetarian, or believe that it's bad to be at war. But you also have large numbers of nonmoral normative beliefs, as when you believe that you ought to keep doing exercise, that no-one should wear a bowtie, that pie for dinner would be good, or that it's bad to have no pension. Understood in this extramoral way, we have very many normative beliefs indeed.

[1] For example, I rely on no assumptions about the metaphysical nature of propositions. A distinct worry is that we have some desires for *objects*, not propositions, such as when you desire chocolate. I agree with those who think that there are no such desires, and that sentences that appear to ascribe such desires are really elliptical claims about propositional desires (Sinhababu 2015; for opposition, see Brewer 2006; Thagard 2006). But note that this is consistent with thinking that many *other* attitudes are not propositional, including attitudes such as likings (Grzankowski 2015; Montague 2007). For the contrast between desiring and liking, see §7.3.

1.2 Reasons and Reasons Beliefs

Our normative beliefs fall into various subcategories. For example, there are beliefs about what you ought to do, beliefs about what you have reason to do, beliefs about things being good overall, and beliefs about things being good in certain respects. See Figure 1.1 for a simple taxonomy.

Most of my discussion will focus on beliefs about just one of the nodes in Figure 1.1: beliefs about reasons. For example, if you are debating whether to opt for surgery, you might well form beliefs about the reasons for, or against, surgery. Or you might believe you have reason to do more exercise, to help your sister, or to drink some tea. I will *very* often refer to such beliefs, and I will refer to them as 'reasons beliefs'.

In fact, by 'reasons beliefs', I have something relatively narrow in mind. You might believe [that *Sarah* has a reason to help you], believe [that you *don't* have reason to jump through the window], or believe [that *if* dragon fruit is tasty *then* you have reason to eat it]. Though these are beliefs about reasons, I won't include them as 'reasons beliefs', in my sense. In my sense, 'reasons beliefs' are *atomic* beliefs about single reasons you yourself have: beliefs with the content [I have reason to v].[2] This shouldn't be too confusing: such reasons beliefs are the canonical kind. And don't worry: I'll remind you of this restriction at the most crucial points.

What, exactly, are reasons? Two clarificatory remarks are crucial.

First, by 'reason', I mean *normative* reason. Normative reasons are facts that count in favour of actions, or (equivalently) that contribute towards justifying actions. I will never use the word 'reason' to refer to *motivating* reasons (or any other purely explanatory reasons). Motivating reasons are the reasons

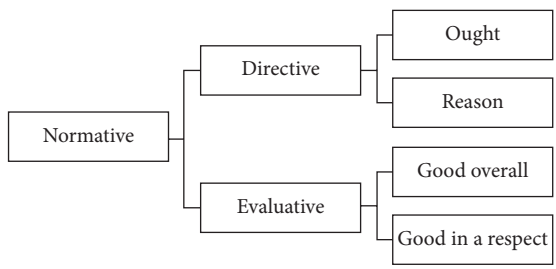

Figure 1.1 Taxonomy of normative beliefs.

[2] To be clear, 'v' here might be logically complex, as when you believe [that you have a reason to (either A or B)]. But any kind of wider logical complexity ensures that the belief in question is not a 'reasons belief', in my sense.

why people act, and merely explain, rather than contribute to justifying, their actions. For example, we might describe Hitler's reasons for invading Poland: such motivating reasons explain what he did rather than justify it. In contrast, normative reasons justify actions, or show that they are appropriate. We will talk a lot about motivation, but since other phrases are available to refer to motivating reasons, I'll keep things clearer by reserving the word 'reason' for normative reasons only. Similarly, by 'reasons belief', I mean beliefs about *normative* reasons, so that your beliefs above amount to beliefs that there is something to be said for doing exercise, helping your sister, or drinking tea.

Second, if I leave it unqualified, by 'reason' I will always mean 'reason to act', where an act (action) is something you intentionally do, such as cutting your own hair, or buying a toupee. Really, normative reasons can favour attitudes as well as actions—you might have some reason to believe that the butler did it, or to feel sad. But since I focus on reasons for action, and say little about reasons for attitudes, I will just use 'reason' to mean 'reason to act'. Similarly, by 'reasons belief', I mean only beliefs about normative reasons for actions, never beliefs about reasons for attitudes.

In short, by 'reason' I mean the things that count in favour of certain actions—things that go in the 'pros' column for an act. And when I talk about 'reasons beliefs', I mean our beliefs about such things.

Can we say anything more about such normative reasons for action? Two broad points are worth mentioning.

First, as I said in the Introduction, desire-as-belief fits most neatly into the objectivist tradition according to which normative reasons are relatively independent of your own state of mind.[3] On this view, you might have various reasons to have surgery, donate money to charity, or keep your promises, and those reasons are independent of what you think or feel on these issues. Since I assume this kind of objectivist view, I assume that there are real facts of the matter about whether you have certain reasons or not, and your reasons beliefs aim to track objective facts about those reasons, just as your beliefs about planets aim to track certain objective facts about those planets.

But though I assume some kind of objectivist view, I will stay silent on exactly which kind of objectivist view we should endorse.[4] There are many

[3] I say 'relatively' independent because of course *some* reasons can depend on your own state of mind: if you are in pain, that might be a reason to take a painkiller. But objectivists think that these are special cases and certainly that not all reasons are like this.

[4] In my (2016), I defended some claims about reasons that are in some ways awkward partners for my claims about desire in this book. I think the two views could be reconciled, but I won't undertake that task here, and I shall proceed to ignore my claims from that paper.

kinds of objectivism, and desire-as-belief should be compatible with the vast majority of those. For example, it is compatible with pluralist views like Ross', as well as more monistic views like utilitarianism. It is also consistent with metaethically naturalist views as well as metaethically non-naturalist views. Desire-as-belief tells us that desires are beliefs about reasons, and these different objectivist theories will disagree about which such beliefs are true, and what makes them true. These seem like relatively independent issues. Just about any view that makes the truth of these beliefs independent of our desires is likely to fit perfectly well with desire-as-belief. For that reason, I won't express my own views about the respective merits of different objectivist theories, since almost all are compatible with desire-as-belief. Whatever your preferred objectivist theory, you can combine it with desire-as-belief. (One notable exception are objectivist theories that incorporate desire-based theories of wellbeing—I discuss those in §10.3.)

But second, I will make one small assumption: I shall assume that reasons connect in a systematic way with what you ought to do. More specifically, I assume that you ought to do something just when you have *most* reason to do it. The idea is that our reasons weigh against one another, and the balance of reasons determines what you ought to do overall. As a result, at some points I shall make claims about what you ought to do, trusting you to understand that these relate in an obvious and systematic way to what you have reason to do. I hope this assumption seems as natural to you as it does to me.

To summarize: I will use the label 'reason' to refer to normative reasons for action. In turn, I will use the label 'reasons beliefs' to refer to our beliefs about such normative reasons for action—beliefs about justifications for various acts. Most crucially, you should never get misled by the alternative use of 'reason' where it means a *motivating* reason: a mere explanation of what moved someone to act. I will talk about such things, but never using the word 'reason'.

1.3 Desire-as-Belief

In this book, I will argue for *Desire-As-Belief*.[5] This view identifies our desires with some of our normative beliefs, and more specifically with our reasons beliefs:

[5] Similar views can be found in Campbell (2018), McNaughton (1988), and Little (1997).

> *Desire-As-Belief*: To desire [to v] just is to believe [that you have reason to v].

Desire-as-belief says 'desire' is really just a label that refers to a certain subset of our beliefs. It says that when a person desires to do something, that is just the very same thing as their believing that they have reason to do it. For example, desire-as-belief says that to desire to start work early is to believe that you have reason to do so (perhaps you work best then). Vice versa, desire-as-belief says that every reasons belief qualifies as a desire, so that if you believe you have reason to buy some wellies, you thereby desire to buy some.

As stated above, desire-as-belief made a claim only about desires [to __]. But it seems that some of our desires are desires [that __], rather than desires [to __], as when I desire [that Spurs win]. What should defenders of desire-as-belief say about such propositional desires? One option here is to insist that desire-as-belief is a theory only about desires [to __], and to treat propositional 'desires' as distinct states—perhaps as hopes, or wishes, rather than desires (cf. §6.5).

But a second more ecumenical option is to instead extend desire-as-belief and make a further related claim about desires [that __]:

> *Desire-As-Belief*: To desire [that p] just is to believe [that you have reason to bring about p].

We could separately accept both this analysis of desiring [that __] and the above analysis of desiring [to __]. Or, more neatly, we could simply treat the original view as a mere consequence of this second one.[6] We could treat the desire [to v] as the desire [that you v], and could treat the belief [that you have reason to v] as the belief [that you have reason to bring it about that you v]. If that were right, the initial analysis of desiring [to __] is really just a special case of the above analysis of desiring [that __].

In what follows, I proceed on those assumptions. That is, I treat desire-as-belief as a claim about desiring [that __], and assume that claims about desiring [to __] are just shorthand for related claims about desiring [that __]. This way of proceeding is certainly cleanest, since it allows me to largely stay with the orthodoxy and treat desires as propositional attitudes, and because it allows me to deploy examples without much care as to whether they involve desires [to __] or instead desires [that __]: again, I assume that desire-as-belief

[6] For potentially relevant discussion, see Schroeder (2011).

covers both and in fact that the former are just a special case of the latter. But I don't believe I rely on this assumption at any point, and it really serves only to keep my presentation of the issues nice and clean.

A related thing I should emphasize is that we might distinguish a more general formulation of desire-as-belief which merely identifies desires with normative beliefs of some kind, and my more specific version of desire-as-belief which identifies desires with reasons beliefs in particular.[7] The choices here are interesting—I discuss related matters in Chapters 3 and 6, where I argue for my more specific formulation of desire-as-belief. Much of what I say would remain relevant if we instead pursued some other nearby view. Still, for ease, by 'desire-as-belief' I will always mean my specific version of that view.

To help us understand desire-as-belief, let's turn our attention to another state of mind: disbelief (cf. Price 1989, 120–1). As I said above, we believe many things. Some of these things involve negation (¬). For example, you might believe [that it is *not* Christmas today]. Indeed, it seems that we have many beliefs in negations: you presumably believe [that it's not the case that grass is tasty], [that it's not the case that sheep wear top hats], [that it's not the case that 2 is larger than 10], and so on. Given how common beliefs with negated contents are, it might in some contexts be useful to use the word 'disbelief' to refer to such states of mind. Disbelief is not some new state of mind over and above belief: it is just a belief with a negated content. With this word in place, we can talk about the things you disbelieve: [that grass is tasty], [that sheep wear top hats], [that 2 is larger than 10], and so on. When we talk about these disbeliefs of yours, that is just another way of talking about the above beliefs of yours. In this way, we might put the word 'disbelieve' to good use, where 'disbelieving [that p]' is just shorthand for 'believing [that ¬p]'. The introduction of the word 'disbelief' allows us to describe beliefs with negated contents in a more concise manner, where we move the negation out of the content of a belief and into our description of the attitude itself. Of course, we are not actually moving anything around: this is just a convenient way of talking. When you use the label 'disbelief', you aren't committing yourself to the existence of a new state of mind beyond belief. Rather, you are just using a convenient label that lets you talk about some of our beliefs slightly more concisely.

We can now get a better understanding of desire-as-belief. Remember, it says:

[7] A slightly different option is to identify desires not with beliefs, but with perceptual states, or similar. I discuss and reject such views in §4.4.

Desire-As-Belief: To desire [that p] just is to believe [that you have reason to bring about p].

Desire-as-belief says that 'desire' functions exactly like 'disbelief'. Desire-as-belief says that to 'desire [that p]' is just to believe [that you have reason to bring about p]. Just as the word 'disbelief' allows us to move negation out of the content of a belief and into our description of the attitude, the word 'desire' allows us to move reason-to-bring-about out of the content of a belief and into our description of the attitude. And just as the word 'disbelief' might allow us to talk more concisely about beliefs with negated contents, by using this label 'desire', we might talk more concisely about our reasons beliefs. Desire-as-belief literally identifies desires and beliefs with a particular normative content, in just the same way that we identified disbeliefs and beliefs with a particular negated content. It says that when we talk about desires, this is just a particular way of talking about a particular subset of our beliefs. That is the view I will defend.[8] For a simple illustration, see Figure 1.2.

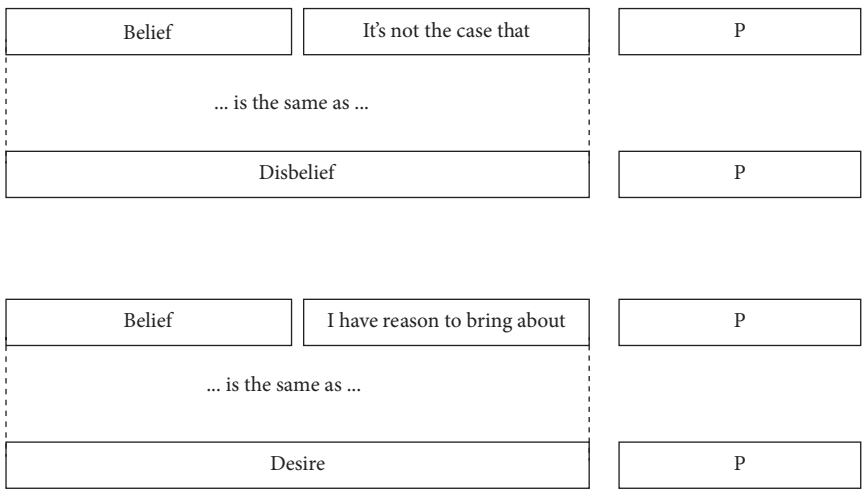

Figure 1.2 Desire-as-belief illustrated by analogy with disbelief.

[8] Are there other words like 'disbelief' and 'desire'? If so, they are probably words for attitudes with contents governed by monadic propositional operators, where attitudes with such contents are sufficiently common that greater parsimony of expression is helpful. There are some candidates: perhaps someone doubts [p] just when they believe [probably not p], or someone expects [p] just when they believe [p will happen in the future] (Searle 1983, 31). Another possibility is that someone suspends judgement whether p when they believe [I have insufficient reason to believe p or its negation] (Raleigh Forthcoming; see also Friedman 2013). On this broad topic, see also Campbell (2018).

1.4 Direction of Fit

With desire-as-belief clarified in this way, this is a good place to address the *direction-of-fit* metaphor (Humberstone 1992; Gregory 2012; Smith 1994, 111–16; Platts 1997, 256–7). The idea is that desires and beliefs have different 'directions of fit'. Whilst in each case something goes right if the content of the attitude matches up to the world, the thought is that the direction of this fit travels in opposing directions: whereas beliefs aim to fit the world, desires aim to have the world fit them. That is, we try to revise our beliefs to bring them into line with the way the world is, whereas we try to bring the world into line with how we want it to be. (If you like, think of beliefs as soft wax that press against the world and have the world imprint the truth upon them, and desires as stamps that press against the world and imprint their contents upon it.) This metaphor seems to suggest that when we think about beliefs and desires, we are thinking about states of mind that are in some fundamental sense the opposite of one another. That might seem to count against desire-as-belief, which identifies desires with a particular subset of our beliefs (Smith 1994, 116–25).[9]

But with desire-as-belief clarified via the comparison with disbelief, we can see why this line of thought is mistaken. According to desire-as-belief, we can describe desires in two different ways, and depending on how we describe them, we describe their content in a different way.[10] In turn, when we talk about the direction of fit of a mental state like desire that can be described in more than one way, we need to be careful about which content the relevant direction of fit is being ascribed to. Desire-as-belief says that to desire [that p] is to believe [that you have reason to bring about p]. The former content—[that p]—is plausibly one that you are supposed to impose on the world. The latter content—[that you have reason to bring about p]—is plausibly one that

[9] Smith presents the problem as a problem for the 'besire' theory, rather than desire-as-belief, though he surely thinks it would undermine both. I discuss the besire theory in §6.1 and reject it for reasons that are independent of the present issue.

[10] This claim also has some bearing on views according to which desires are not states with normative contents, but instead have non-normative contents that they represent with normative *force* (Schafer 2013; Tenenbaum 2008). These views are sometimes pitched as alternatives to a view like desire-as-belief. But I agree with these views insofar as reasons rarely feature in the contents of our desires: I rarely desire [that I have reason to v]. This fact is consistent with desire-as-belief: according to desire-as-belief, that desire would be an unusual belief—one about a higher-order reason: a reason to bring it about that I have reason to v. Those defending these views also tend to say that desires relate to the good in the same way that beliefs relate to the true. Let's set aside the fact that I identify desires with beliefs about reasons rather than beliefs about goodness. Even so, this clam isn't right. Since claims about goodness are themselves truth-apt, it would be more accurate to say that desires relate to the *truly* good as beliefs relate to the true. But so understood, this claim fits perfectly well with the claim that desires are a subset of our beliefs. In these ways, I think these views are either consistent with desire-as-belief or else implausible.

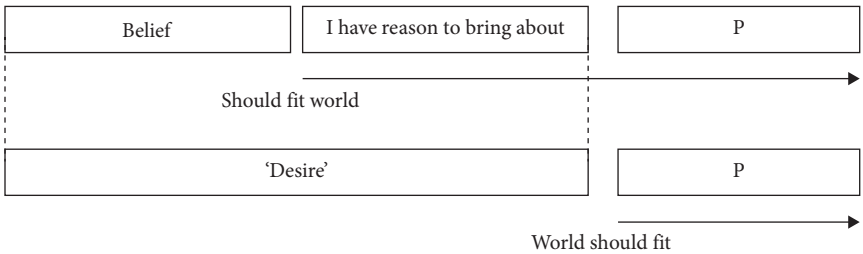

Figure 1.3 The two directions of fit of desire.

you are supposed to make fit the world as it already is. This state of mind has two directions of fit, each with respect to a different content: it is both supposed to bring about p, and also supposed to be responsive to whether it's true that you have reason to bring about p (Little 1997, 63–4; Price 1989, 120–1).[11] See Figure 1.3.

Understanding this may be easier if we step away from the metaphor. Really, the idea behind the direction-of-fit metaphor is that whereas beliefs are, or should be, responsive to evidence, desires do, or should, influence us to act in ways that are productive of their content. Desire-as-belief says that *both* of these things are true of desires, which is to say, both of these things are true of reasons beliefs. Clearly, your beliefs about what you have reason to do should be sensitive to evidence about whether you really have reason to do those things. But such beliefs should also rationally influence your actions in appropriate ways. And—I here repeat the previous claims in different words—your desires should be sensitive to what you have reason to do, and ought to rationally influence your actions in appropriate ways. Desire-as-belief is perfectly consistent with the distinction between directions of fit: it merely says that this state of mind has both directions of fit at once, each with respect to a different content.

It might also help if we return briefly to our comparison with disbelief. Imagine someone reasoning as follows:

> 'Disbelief cannot be a belief of any kind, because those states of mind have opposing directions of commitment. Whereas the belief [that p] commits

[11] Some might deny that that one state of mind could have two contents. Perhaps it would be better to say that this state of mind has just one content—that [I have reason to bring about p]—but also has [p] as a part of its content, and the desire-like direction of fit applies to a content-part rather than a content. I am not so clear about this—too much hinges on how we think of mental contents—and at any rate the analogy with disbelief ensures that there must be *some* appropriate way to understand states of mind which can be described in different ways and which seem to get ascribed different contents depending on how they are described.

> one to [p], the disbelief [that p] commits one to [¬p]. Beliefs commit one in *favour* of their contents, whereas disbeliefs commit one *against* their contents. So disbelieving cannot be a kind of believing'.

This is bad reasoning. The very idea is that disbelieving [p] just *is* believing [¬p], and as such disbelieving [p] *both* commits you in favour of [¬p] and commits you against [p]. Similarly, according to desire-as-belief, desiring [that p] just *is* believing [that you have reason to bring about p], and as such the desire [that p] *both* aims to fit whether [you have reason to bring about p] and aims to have the world fit [p].

A final point: desire-as-belief is consistent with the idea that *describing* a state of mind as a belief, or as a desire, makes one of its directions of fit more salient. For example, when we describe a state of mind as a belief [that you have reason to bring about p], by referring to it as a 'belief' we might thereby emphasize the way in which this state should be responsive to the truth about whether you in fact have reason to bring about p. Vice versa, when we describe this very same state of mind as a desire [that p], by referring to it as a 'desire' we might thereby emphasize the motivational role it plays in affecting your actions. So it is possible that by describing states of mind as 'beliefs' or as 'desires', we thereby emphasize different features they have, and in turn we might emphasize one direction of fit by describing the state in a certain way. To this extent, one direction of fit is more closely associated with beliefs, and another is more closely associated with desires. But this is all consistent with desire-as-belief, which says that ultimately, there is just one state of mind here with both directions of fit. Our ability to draw attention to one direction of fit at the expense of the other is consistent with this state of mind ultimately always having both.

I conclude that the direction-of-fit metaphor is perfectly fine, but does absolutely nothing to undermine desire-as-belief.

Before we move on, I can briefly respond to another simple objection to desire-as-belief. Whereas we refer to beliefs as 'true' or 'false', we never attach these labels to desires. Isn't that a simple but effective argument against desire-as-belief? It's again helpful to start by thinking about disbelief. To the extent that we might successfully make use of the word 'disbelief', we would be unlikely to describe disbeliefs as 'true' or 'false'. Why? Because it would be confusing: if someone described the disbelief [that pigs fly] as 'true', it would be unclear whether they meant that the content *of the disbelief* was true or that the content *of the belief* was true: whether they meant that pigs *do* fly or that they *don't*. To this extent, it would be far clearer to communicate

the relevant facts by describing the state of disbelief as 'appropriate' or 'inappropriate', or else by describing the relevant belief as 'true', 'false', 'appropriate', or 'inappropriate'. Given that we have these helpful ways of communicating the relevant facts, it would be bizarre to try to communicate those facts with the highly unclear assertion that the relevant disbelief is 'true'.[12]

Similar reasoning applies to desires, given desire-as-belief. You *could* accurately describe desires as 'true' or 'false'. But it would be extremely confusing. If someone called the desire [that pigs fly] 'true', it would be unclear whether they meant that the content of the desire was true, or that the content of the belief was true: whether they meant that pigs *do* fly, or instead that there is a *reason to make* them fly (i.e. whether they meant that the desire is *satisfied*, or that it is *appropriate*). As a result, it's clearest to make the intended claim by saying that the relevant desire is 'appropriate' or 'inappropriate', or else by saying that the relevant belief is 'true', 'false', 'appropriate', or 'inappropriate'. To this extent, desire-as-belief positively predicts that we don't refer to desires as 'true' or 'false'. Doing so would be unnecessarily confusing, and that is a good explanation of why we do not talk in that way.

1.5 Some Broad Attractions of Desire-as-Belief

In the chapters ahead I'll present some detailed arguments for desire-as-belief. But before we get to those details, we can here note two more simple attractions of the view.

The simplest attraction of desire-as-belief is that the view is highly parsimonious, reducing desires to beliefs. According to other views, we have two distinct and important states of mind: beliefs and desires. But according to desire-as-belief, we have just one important state of mind: belief. On this view, talk of 'desire' is just another way to talk about some of these beliefs. I take it that theories are more plausible if they can explain the data while positing the existence of fewer kinds of entity. So if desire-as-belief *can* explain the data, it is a more plausible theory than many rivals. This is clearly an attraction of the view. True, a lot hinges on whether desire-as-belief *can* explain the data, but as we might put it, desire-as-belief is a good starting hypothesis that we should abandon only if it can't explain the data; other more positive arguments for the view might be superfluous.

[12] Similar reasoning applies to 'deny' and 'reject': denials and rejections are truth-apt, and yet we don't call them 'true' or 'false'. This is surely just because it would be deeply confusing to talk that way.

A second simple attraction of desire-as-belief comes from noting how much there is in common between desiring something and believing there's a reason to bring it about (we will investigate many of these in greater depth in the following chapters). Plausibly, reasons beliefs and desires are both capable of motivating you to do things. Reasons beliefs and desires both play important roles in practical deliberation. Reasons beliefs and desires both seem capable of rendering your actions (ir)rational. Reasons beliefs and desires both come in degrees: you can think you have a weak reason, or a strong reason, and you can want things a little or a lot. You can believe that you have conflicting reasons, and so too you can have conflicting desires. Reasons can be believed to be instrumental, when they favour a means to an end, or ultimate, when they favour an end for its own sake, and this same distinction holds for our desires. Ultimate reasons beliefs and ultimate desires seem relatively stable over time, whereas instrumental reasons beliefs and instrumental desires do not. Reasons beliefs and desires come in both non-comparative and comparative forms: just as you can believe [that you have reason to bring about p] and desire [that p], you can believe [that you have more reason to bring about p than q], and you can prefer [p to q]. Reasons beliefs and desires can both have their demands met: you can comply with the reason you believe you have, or satisfy the desire. Reasons beliefs and desires are both evaluated in a manner that is agent-relative: some reasons beliefs and desires might be more appropriate for me than you, or vice versa.

This list could go on, but the basic and simple point should be clear: it looks as though desires and reasons beliefs are made for each other, in that they have many properties in common. This isomorphism needs explaining, and one obvious explanation of it is that the two are in fact one and the same. To this extent, desire-as-belief seems in a good position to explain some obvious facts about reasons beliefs and desires.

1.6 Some Broad Defences of Desire-as-Belief

Some might think that desire-as-belief is sufficiently implausible that we should just reject it out of hand. Again, in the following chapters, I defend desire-as-belief in detail. But again, before we get to those details, we can note some general reasons for optimism about desire-as-belief. David Lewis claimed that incredulous stares can't be answered (1986a, 133), but I hope they might nonetheless be prevented by casting a view in a different light. We can then move onto more articulate arguments.

Sometimes, when we talk about our desires, what we say seems inconsistent. For example, as you reluctantly drag yourself out of the house, filled with dread, and head towards the dentist, it would be odd to insist that you *want* to go to the dentist. But at the same time, when the bus driver asks where you're heading, you might quite truthfully say that you want to go to the dentist. So which is it: do you want to go, or not? Our thoughts here might seem inconsistent.

In a case like this, I think we should try to find resources that explain the apparent variation in the kinds of claims we make about desire. Rather than treating such variation as demonstrating inconsistency in our thoughts, I shall try to find theories that predict the relevant variation. More specifically, I will make sense of what we say about desire in part by showing how some of what we say about our desires is misleading. We might have thought that a theory of desire is most plausible if it vindicates the truth of everything we say about our desires. But, in fact, we need only vindicate the truth of those claims that ought to be read at face value. We might then argue that the other claims that we tend to make are misleading and ought not be read at face value.

By aiming to vindicate only some claims about desire, we make a difference to the range of phenomena that a theory of desire ought to explain. For example, in the case above I think you really do desire to go to the dentist, and that we can explain away our tendency to *say* that you do not want to go (§2.4.1). If this is right, we can maintain that although our theory of desire must make sense of your desire to go to the dentist, it need not vindicate the appearance that you don't desire to go. Instead, we explain that latter appearance away by appeal to independent linguistic theories that explain why we make such misleading claims.

With the above point in hand, when I say that desires are reasons beliefs, the plausibility of that claim might depend on our prior decisions about which ordinary language claims about desire we take seriously and try to vindicate, and which we instead treat as misleading. Part of my job in what follows is to show that we should understand ordinary desire-talk in a way that makes desire-as-belief more plausible. For example, hunger is not a belief of any sort, but this doesn't threaten desire-as-belief so long as I can defend the claim that hunger is not really a desire, whatever some ordinary language might suggest (§7.2). Or, for another example, I agree that we are sometimes weak-willed, and fail to be motivated to do things that we think we ought to do. But again, I don't think this threatens desire-as-belief, so long as I can defend the claim that motivation is distinct from desire, whatever some ordinary language might suggest (§2.3, §5.5). As Austin said, in philosophy

there's the bit where you say it, and the bit where you take it back (1965, 2): though desire-as-belief is substantial and interesting, it nonetheless survives some kinds of criticism in virtue of having more modest implications than its objectors assume it must have. So we should not reject desire-as-belief until we have gotten clear about exactly what it *does* commit us to.

I used to think that philosophers should be interested only in the world itself, and not interested in contingent facts about the arbitrary language we use to describe it. But Williamson (2008, 284–5) rightly points out that this is like an astronomer insisting that astronomers ought to be interested in the stars and not their telescopes: though it is true, their understanding of the stars will improve if they come to understand how their telescopes might distort their vision. In this book I defend desire-as-belief in part by applying the parallel lesson to our theorizing about desires and desire-talk.

1.7 Reason and Passion

Many of the details of desire-as-belief are new. But in many ways, desire-as-belief is an old view. The underlying picture on which our desires are entwined with our normative or ethical beliefs is hinted at, or outright endorsed, by many of those who form the canon of at least Western philosophy. So the view is not an aberrant flight of fancy, but instead a development of a long-standing tradition of thinking of our desires and ethical views as deeply connected.

The oldest and most well-known figures in this tradition are probably the Stoics, who apparently claimed that 'impulses are acts of assent' (Long and Sedley 1987, sec. 33I1; see also 53R, 53S, 60F, 65A4, 65G1, 65I4). But many other figures also seem to endorse some view in this vicinity. For example, Aristotle wrote that 'The object of desire always moves, but this is either the good or the apparent good' (DA 433a27–9; see also EN III.3 1113a23–4, EE VII.2 1235b25–7), and Aquinas wrote that 'there is appetite only for a good which is proposed to it by a cognitive power' (DV 24.2; see also ST I–II.1.1). Some of the rationalists may have also held views of this sort: Spinoza wrote that 'The will and the intellect are one and the same' (Ethics II P49 Cor.; for some discussion see Youpa 2007), and Leibniz claimed that 'volition is the effort or endeavour (conatus) to move towards what one finds good and away from what one finds bad, the endeavour arising immediately out of one's awareness of these things' (1982, 172; see also 1989b, sec. 13 and 1989a, 279). More recent influential authors endorse something in the vicinity of

desire-as-belief as well. For example, Davidson writes, 'If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y' (2001b, 23), Murdoch writes that 'Will and reason then are not entirely separate faculties in the moral agent' (2013, 39), and Scanlon writes that 'desire, in order to play the explanatory and justificatory roles commonly assigned to it, needs to be understood in terms of the idea of taking something to be a reason' (1998, 7–8).

Still, not everyone has been happy to subscribe to a view of this kind. Even the Stoics were not united in their views: most notably, Posidonius appears to have explicitly rejected this kind of view.[13] Posidonius seems to have been concerned with cases where we act irrationally (see e.g. Long and Sedley 1987, sec. 65K3), and this concern is shared by others. For example, Locke resists the 'established' and 'settled' view that we desire the good on the same basis (Essay, II.XXI.31–8). I address such objections in Chapter 5.

Another famous objector to desire-as-belief is Plato. In *The Republic*, Plato presents an argument that Reason and desire form different parts of the soul (436b–439d). ('Reason' here has a capital R, and thereby refers to the faculty of the mind, rather than a normative reason.) If correct, this would presumably entail that desire-as-belief is false.

Plato's basic idea is that we can be conflicted about what to do, and this requires that our conflicting impulses come from different parts of the soul. That is, one and the same part of the soul could not simultaneously be F and not-F: that would be a contradiction. And so, one and the same part of the soul could not simultaneously be in favour of, and not in favour of, drinking. Now imagine a man who is thirsty, but who is unwilling to drink. Plato argues that his thirst—a desire—must originate in a different part of his soul from whatever ethical principles lead him to avoid drink.[14] Otherwise, we would have to say that he both wants to drink and doesn't want to drink, and that would be a contradiction. So desire and belief must originate in different parts of the soul.

But Plato's argument can't be right (Annas 1981, 137–8). We should all allow that conflicts can occur *within* individual parts of the soul: your desires might conflict with one another, and your reasons beliefs might conflict with one another, as when you think that there is something to be said for going to

[13] See e.g. Galen's remarks in Long and Sedley (1987, sec. 65K2); for detailed discussion, see Cooper (1999). Some of the other Stoics were apparently equally unsure—see Long and Sedley (1987, sec. 65K1).

[14] Later (§7.2), I argue that thirst is not even a desire, but the general points here could easily be made with other examples.

the pub and something to be said for staying in. To this extent, everyone needs to make sense of conflicting attitudes in a manner that doesn't require us to locate each conflicting attitude in a different part of the soul.

So we should reject Plato's argument. And that is easily done, since it involves a scope fallacy: Plato conflates desiring [not p] (D¬p) and failing to desire [p] (¬Dp).[15] Plato seeks to avoid the conclusion that the thirsty man both wants [to drink] and fails to want [to drink], since that would be a contradiction. But even if the soul is unitary, we can avoid that conclusion. We should say that the thirsty man who is unwilling to drink is someone who wants [to drink] and also wants [not to drink]. That combination of attitudes is perfectly consistent with the law of non-contradiction, and is enough to explain his conflicted psychology. So we should not accept Plato's argument that Reason and desire form different parts of the soul.

There is a further—better—objection to desire-as-belief, related to Plato's. The idea is that it is perfectly rational to have conflicting desires, but definitely irrational to have conflicting beliefs. It seems to follow that desires could not be beliefs: they do not stand in the right consistency relations with one another (Archer 2016, 3–4; Tenenbaum 2007, 38–9). But this objection also fails. Desire-as-belief says desiring [that p] is believing [that you have reason to bring about p]. So now imagine that you desire [that p], but that you also have a conflicting desire [that ¬p]. If desiring [that p] is believing [that you have reason to bring about p], your desire [that ¬p] should be understood as the belief [that you have reason to bring about ¬p]. It should be clear that these beliefs are perfectly consistent: they can both be true, since you might well have reasons to bring about p and competing reasons to bring about ¬p. Here, your conflicting desires might accurately represent a genuine normative conflict. As I have formulated it, desire-as-belief entails that conflicting desires are beliefs about conflicting reasons, and not conflicting beliefs about reasons. So desire-as-belief permits that conflicting desires are rational, even though conflicting beliefs are not. That is, desire-as-belief does not analyse conflicts between desires as conflicts between beliefs (i.e. B[Rp] & B[¬Rp]), but rather as conflicts between reasons you believe you have (i.e. B[Rp] & B[R¬p]). It thereby entails that conflicts between desires are totally unlike conflicts between beliefs: conflicts between desires can accurately represent conflicting normative pressures, whereas conflicts between beliefs can never accurately represent anything (cf. De Sousa 1974; Williams 1976). So this further objection to desire-as-belief also fails.

[15] This crucial distinction features again in §2.4.1 and §5.4.

Plato's objection to desire-as-belief, and its descendant, both fail. But might we still want to maintain *some* distinction between Reason and passion? And if so, might that distinction undermine desire-as-belief? No. Though we should draw some useful distinctions in this vicinity, none of them conflict with desire-as-belief.

For example, we should definitely distinguish between beliefs and emotions (see also §2.4.4, §7.1). But this is consistent with desire-as-belief, which is a theory of desire, not emotion. To the extent that the distinction between Reason and passion tells us something only about emotion, it is consistent with desire-as-belief.

Or for another example, we should definitely distinguish between those states of mind that are produced by (conscious?) reasoning, and those that are not. But that distinction is consistent with desire-as-belief, since it cuts across the belief/desire distinction: many beliefs are not produced by reasoning (but instead, say, by socialization), and many desires are produced by reasoning (such as our political preferences).[16]

Or for a final example, we should definitely distinguish between states of mind that are rational and those that are not. But again, that distinction is consistent with desire-as-belief, since it also cuts across the belief/desire distinction: many desires may be irrational, but so too are many of our beliefs. We have always known that our beliefs can go awry in various ways, but the catalogue of our failures is constantly growing (e.g. Kahneman 2011). Moreover, our normative beliefs are especially prone to irrationality: more hangs on them, and so they are more liable to distortion from incentives such as self-interest. To this extent, desire-as-belief is not only *consistent* with the claim that our desires are often irrational, but in fact positively *explains* such irrationality (for more, see §5.3). So again, the distinction between rational and irrational states of mind does nothing to undermine desire-as-belief.

In these ways, we should be wary of rejecting desire-as-belief out of hand because it conflicts with the distinction between Reason and passion. That distinction is ambiguous, and on obvious disambiguations it is perfectly consistent with desire-as-belief.

After Plato, Hume is the main canonical figure who cemented the division between Reason and passion.[17] He writes:

[16] I discuss our ability to change our desires by reasoning more thoroughly in §8.2.

[17] That said, Hume's views are not perfectly straightforward. For example, Hume seems to suggest some sympathy for something like desire-as-belief when he says things like 'Desire arises from good consider'd simply' (T2.3.9.7; see also T2.1.1.4, T2.3.9.2), though interpreting such claims is difficult given that Hume seems to use the words 'good' and 'pleasant' interchangeably (T2.3.9.8).

> A passion is an original existence, or if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high. 'Tis impossible, therefore, that this passion can be oppos'd by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represent.
>
> (T2.3.3.5)

Hume here suggests that desire-as-belief is false. One concern is that Hume gains unfair rhetorical force by switching between distinct states of mind: he moves back and forth between desires and emotions, even though these are clearly distinct states of mind. Defenders of desire-as-belief should not be too concerned if emotions like anger do not represent the world, given that they have a theory of desire and not emotion. But even if we set this issue aside, Hume doesn't actually give any argument in this passage against desire-as-belief. Rather, he simply *asserts* that passions do not represent the world: he asserts that our passions make no reference to any other object (see also T2.3.3.6). But this is just the very claim that is in question when we assess desire-as-belief. And so Hume's only real argument for that division is the argument that Reason alone cannot motivate us: I address that across the next two chapters.

1.8 Summary

In this chapter I outlined some basic terminology and assumptions that I make, and described desire-as-belief, according to which desiring something just is believing that you have reason to bring it about. I showed that the view is consistent with the direction-of-fit metaphor, has some simple attractive features, and can avoid some initial objections, including those relating to inconsistency in desire. Our next chapter discusses the role that desires play in motivating people to act, and defends the view that everything anyone has ever done can be explained in a systematic manner by appealing to their desires. That view will be important for Chapter 3, since it is a premise in an argument there for desire-as-belief.