*Preprint version.*

*As of October 2016, this article has been accepted for publication in* *Synthese*.

Please reference the published version.

This preprint contains several typos and points of unclarity that are fixed in the published version.

**Doxastic Desire and Attitudinal Monism**

Douglas I. Campbell

How many attitudes must be posited at the level of reductive bedrock in order to reductively explain all the rest? Motivational Humeans hold that at least two attitudes are indispensable, belief and desire. Desire-As-Belief theorists beg to differ. They hold that the belief attitude can do the all the work the desire attitude is supposed to do, because desires are in fact nothing but beliefs of a certain kind. If this is correct it has major implications both for the philosophy of mind, with regards the problem of naturalizing the propositional attitudes, and for metaethics, with regards Michael Smith’s ‘moral problem’. This paper defends a version of Desire-As-Belief, and shows that it is immune to several major objections commonly levelled against such theories.

**Keywords:** Propositional Attitudes, Belief, Desire, Motivational Humeanism, Cognitivism, Internalism

**1. Introduction**

There are numerous attitudes. For any given proposition, *p*, one might *believe that p*, or *desire that p*, or *be sorry that p*, or *imagine that p*, or *be excited that p*, or *doubt that p*, or *grimly suspect that p*, and so on. However, some of these attitudes can be reductively explained in terms of others. For example, the following reductive conceptual analyses appear at least approximately correct:

1. One *thinks that p* iff one *believes that p*.
2. One *wants that p* iff one *desires that p*.
3. One *wishes that p* iff one *desires that p but believes that p is unlikely* (Smith 1994, 117).
4. One *fears that p* iff one *desires that p but believes that ¬p is possible, and this belief-desire combination causes a feeling of fear*.

Notice that these analyses are *reductive*, not *eliminative*. For instance, to endorse IV is not to deny that the attitude of *fearing* exists. It is merely to claim fearing that *p* is constituted of having a certain combination of belief, desire and suitably caused fearful feeling.

How many *irreducible* attitudes are there? How many distinct attitudes must be posited at the level of reductive bedrock in order to supply the resources necessary to reductively explain all the others? A widespread view—standardly associated with Hume, and forcefully advocated by Smith (1994)—is that there are *at least two* irreducible attitudes, *belief* and *desire*. Beliefs are about how things *are*, while desires are instead about how things *should be*. Beliefs are *representational* (or *theoretical*, or *cognitive*, or *indicative*, or *thetic*, or *descriptive*), while desires are *motivational* (or *practical*, or *appetitive*, or *imperative*, or *telic*, or *normative*). They have different, and opposing, *directions of fit* with the world (Anscombe 1957; Platts 1979; Smith 1994; Humberstone 1992). When one’s beliefs don’t fit the world, one is to correct the discrepancy *by changing one’s beliefs*. In contrast, when one’s desires don’t fit how things are, the discrepancy is to be corrected by *changing the world*. Because beliefs and desires have these diametrically opposing natures it is commonly held than no attempt to reduce desires to beliefs, or beliefs to desires, or both beliefs and desires to some third attitude, could possibly get off the ground.

My aim is to challenge this idea. My own view is that both the following doctrines are true:

*Desire-As-Belief:* desires are identical to beliefs of a certain kind.

*Attitudinal Monism:* there is only one irreducible attitude, belief.

Here I argue for Desire-As-Belief. In so far as Desire-As-Belief removes the primary obstacle to accepting Attitudinal Monism, I thereby also argue for Attitudinal Monism.

By way of arguing for Desire-As-Belief I describe and defend a theory of desire, the *Doxastic Theory of Desire* (DTD). According to DTD, one desires that *ψ* with strength *x*, iff one believes *that D(ψ,x)*, with *D* being a very special mental predicate about which much will be said below. I argue for DTD by showing that it is a *serviceable* theory, and indeed every bit as serviceable as the orthodox theory according to which beliefs and desires are separable and mutually irreducible mental states. More specifically I show that DTD *simulates* the orthodox theory, in the sense that it provides explanations of rational behaviour that align with explanations provided by the orthodox theory. In other words, it can do everything the orthodox theory can do, but more cheaply, since it posits only one fundamental attitude, belief, instead of two fundamental attitudes, belief and desire.

If DTD is true then why in the larger philosophical scheme of things is this important? The answer is that it helps shed light on at least two major philosophical problems. The first is a problem in the philosophy of mind—that of *naturalizing the propositional attitudes*. To naturalize the propositional attitudes one must explain, in adequately naturalistic terms (e.g., in computational, or neurophysiological, or teleo-functional terms), what is involved in the brain being the seat of any given propositional attitude. This problem decomposes into two sub-problems:

*Propositional sub-problem*. For any given proposition *p*, what are the naturalistic properties of the brain in virtue of which it mentally grasps *p*?

*Attitudinal sub-problem.* For any given attitude, *A*, what are the naturalistic properties of the brain in virtue of which it grasps propositions under the aegis of *A* instead of some other attitude?

Let’s focus on the attitudinal sub-problem. Suppose there were, at the level of reductive bedrock, say, *five* mutually irreducible attitudes—*A1*, *A2*, *A3*, *A4* and *A5*. Then, obviously, a solution to the attitudinal sub-problem would need to include a *five-way decision procedure* for classifying any given grasping of *p* as being a token of one of the following five propositional attitudes: *A1p*, *A2p*, *A3p*, *A4p* or *A5p*.

Suppose instead that there were only *two* fundamental attitudes—belief and desire. Then a solution to the attitudinal sub-problem could be considerably simpler. It would only need to include a *binary* decision procedure for classifying a grasping of *p* as being either a *belief that p* or a *desire that p*.

Finally, imagine Attitudinal Monism was true—i.e., that there was only *one* fundamental attitude, belief—and that all the other attitudes, including desire, could be reductively analysed in terms of belief (using analyses like I—IV and DTD). In this case there would no longer be *any need* to decide which attitude a proposition was being grasped under, there being, at the level of reductive bedrock, only a *single* attitude under which a proposition *could* be grasped, belief. A decision procedure that must always yield the *same* answer—‘belief’—is no real decision procedure at all, there being no decision to make. Thus in so far as DTD, by reducing desires to *D*-beliefs, suggests Attitudinal Monism is true, it *dissolves* the attitude sub-problem (by making it trivial), and thereby dissolves *one half* (the *attitudinal half*) of the problem of naturalizing the propositional attitudes. Of course this still leaves the *other half* of the problem—the propositional half—intact, but progress can be claimed even on this front since a satisfactory solution to the propositional sub-problem is liable to be much easier to find when it needn’t be compatible with a separate solution to the attitudinal sub-problem.

The second major philosophical problem DTD bears on concerns the following three famous meta-ethical claims (the first of which has two parts):

*Motivational Humeanism (MH)*

*MH1:* An agent can be motivated to act only by some combination of a desire and a means-end belief; and

*MH2:* Desires and beliefs are ‘distinct existences’ in the sense that they are *modally separable*. That is, for any belief, *b*, and desire, *d*, it is logically possible for an agent to have *b* without *d*, or *d* without *b*.

*Cognitivism:* moral judgments are judgments about matters of fact—i.e., they are beliefs.

*Internalism:* moral judgments, when combined with suitable means-end beliefs, have motivational force.

Together these claims form an ‘inconsistent triad’.[[1]](#footnote-1) They are inconsistent because: (i) cognitivism states that moral judgments are beliefs; (ii) internalism and MH1 jointly entail that moral judgments are desires; and (iii) MH2 entails that if a moral judgment is a belief then it is not a desire, and *vice versa*. The challenge of resolving this inconsistency—Michael Smith’s ‘moral problem’—is ‘the central organizing problem in contemporary meta-ethics’ (Smith 1994, 11). Non-cognitivists deal with it by abandoning cognitivism, and externalists by abandoning internalism. Both these approaches are costly since cognitivism and internalism possess considerable *prima facie* plausibility. A third approach—that of the cognitive internalist—is to reject MH (either MH1 or MH2).

Now if DTD is true then MH2 is false, for if desires are identical to D-beliefs then they are not modally separable from these beliefs. Thus, in addition to helping to dissolve the attitudinal half of the problem of naturalizing propositional attitudes, DTD also supports a cognitive internalist response to the moral problem.

DTD is important because it has these larger philosophical applications. However my focus in the rest of this paper is not on the applications but on DTD itself—on what it says and why it is plausible. §2 describes DTD. §3 explains why DTD is superior to other extant Desire-As-Belief theories, and in so doing explains how it sidesteps objections commonly levelled against such theories. §4 wraps things up.

**2. DTD**

Imagine an agent, Homer, who strongly desires that [1].[[2]](#footnote-2) He also has the means-end belief that [2]. He has no other relevant beliefs or desires.

(1) I drink beer.

(2) If I get off the couch and fetch a beer from the fridge, then it will become the case that [I drink beer].

Lo: Homer’s brain issues appropriate signals to his muscles. He gets off the couch and fetches a beer from the fridge. How are such triumphs of practical rationality achieved? According to the orthodox belief-desire theory of practical rationality the following pair of rules are involved:

*Means-end rule, version 1 (MER1):* if an agent desires, with strength *x*, that *ψ*, and if she believes that by *φ*-ing she will make it the case that *ψ*, then she has a motivating reason, of strength *x*, to *φ*.[[3]](#footnote-3)

*The maximising rule (MAX):* if an agent has motivating reasons to *φ* that are collectively stronger than her motivating reasons for performing other actions incompatible with *φ*, then the action of *φ*-ing is to be initiated.

If Homer’s practical reasoning system instantiates both these rules then MER1 will cause his belief that [1] and his strong desire that [2] to produce a strong motivating reason to get off the couch and fetch a beer from the fridge. Absent conflicting motivating reasons, MAX will then issue in Homer getting off the couch and fetching a beer from the fridge. Homer’s rational behaviour is thereby neatly explained.

While accepting this story, I maintain that *another* story about Homer’s trip to the fridge can be told *at a lower level*. The two stories are consistent, each being translatable into the other. But the lower-level story explains Homer’s trip to the fridge in *purely doxastic* terms, with all mention of desire being stripped from the explanation and replaced with talk of belief. The explanation it offers is similar to the above explanation in respect of involving both the means-end rule and MAX. However, the version of the means-end rule it uses differs from MER1 in being framed in *purely doxastic* terms. It does not, as MER1 does, make any mention of an agent’s desires. The lower-level, purely doxastic story has six parts—now described—which together comprise the theory I call ‘DTD’.

*Part 1.* The first issue that needs tackling concerns how a mere belief could be motivationally forceful. How, for example, could any mere belief play the motivational role of Homer’s strong desire that he drink beer? By way of answering this question, I posit a two-place mental predicate, *D* (the ‘D-concept’), which accepts a *proposition* and a *scalar value* as its arguments.[[4]](#footnote-4) I also posit the following, purely doxastic version of the means-end rule:

*Means-end rule, version 2 (MER2):* if an agent believes that *D(ψ,x)* and if she believes that by *φ*-ing she will make it the case that *ψ*, then she has a motivating reason, of strength *x*, to *φ*.[[5]](#footnote-5)

To see how this works, let’s suppose that Homer’s practical reasoning system instantiates MER2, and that in addition to having the means-end belief that [2], Homer also believes that [3]:

(3) *D([I drink beer],100)*

Any such belief, having a content of the form ‘*D(ψ,x)*’, is what I will call a ‘D-belief’. MER2 will cause Homer’s believing that [2] and that [3] to issue in him having a motivating reason, of strength 100, to get off the couch and fetch a beer from the fridge.

Voila! We have just succeeded in getting a motivating reason out of a pair of beliefs, without desires (explicitly) entering the picture. How did we do this? Simply by positing a version of the means-end rule, MER2, which treats certain beliefs—namely, D-beliefs—exactly as MER1 treats desires.

It is definitive of the concept I am calling ‘the D-concept’ *that it produces motivating reasons by engaging MER2*. If, upon analysing a creature’s mind, you find a concept, *X*, such that X-beliefs pair up with means-end beliefs to produce motivating reasons in the sort of way MER2 describes, *then you have thereby found D itself*: i.e., *‘X’* and *‘D’* are two names for one concept. In other words, the D-concept, as I conceive it, is *functionally* characterized, so that any concept sharing the functional role of the D-concept vis-à-vis motivation *is the D-concept*, and any concept not having this functional role *is not the D-concept*. It is thus built into the notion of the D-concept that it is motivationally ‘pushy’.

*Part 2.* Now that we have helped ourselves to MER2, and thereby explained how some beliefs—namely, D-beliefs—can acquire motivational ‘oomph’, the next task is to say how such motivationally forceful beliefs are related to desires. I propose the following analysis:

*IDENTITY: An agent desires that ψ with strength x iff she believes that D(ψ,x).*

In other words, the idea here is that every desire is *identical to* a corresponding D-belief, in the sense that having the desire is just a matter of having the D-belief, and *vice versa*. For example, Homer’s having a desire, of strength 100, that [1], simply consists of him believing that [3].[[6]](#footnote-6)

Notice that if IDENTITY is true then an agent will automatically conform to MER1 by virtue of conforming to MER2. For instance, if Homer is motivated to get off his couch because he conforms to MER2 and believes both that [2] and that [3], then IDENTITY implies that it is equally true that he is motivated to get off his couch because he conforms to MER1, believes that [2], and desires that [1]. Where MER1 references a desire, MER2 instead references the underlying D-belief, but other than that the two rules align precisely. And so by identifying desires with D-beliefs, IDENTITY reduces MER1 to MER2. We needn’t choose between the two explanations of Homer’s behaviour—one citing his belief that [2], his desire that [1], and MER1, and the other citing his beliefs that [2] and that [3], and MER2—because IDENTITY implies they amount to two ways of saying the same thing.

Notice, also, that if all desires are identical to D-beliefs, as IDENTITY claims, then the D-concept is a *very special* concept—a concept one must possess as a precondition for possessing desires and for possessing instrumental, means-end rationality. Australopithecines presumably had very few of the concepts we have, but if they had desires—as they presumably did—then IDENTITY implies that they had D-beliefs, and thus the D-concept too. Similarly, IDENTITY implies that if human babies have desires then the D-concept is already wired into their conceptual lexicons. In other words, IDENTITY implies that the D-concept is *phylogenetically and ontogenetically primal*—as primal as is desire itself. It implies that the D-concept is to instrumental and practical rationality much as basic logical concepts like *negation* and *entailment* are to theoretical rationality—utterly fundamental.

*Part 3.* Because DTD implies that desires are identical to D-beliefs, it also implies that any causal power of a desire *must also be a causal power of the D-belief that the desire is identical to*. Chief among the causal powers of desires are, of course, their ability to motivate rational action. We have already seen how DTD explains this particular causal power of desires—namely, in terms of the operation of MER2. We have also seen that D-beliefs are *characterized* in terms of their having this particular causal power. (A belief that lacked the ability to motivate in accordance with MER2 would, for this very reason, not qualify as a D-belief, and thus would not qualify as a desire, according to what has been said.)

But desires have other causal powers, in addition to their motivational role. Neil Sinhababu (forthcoming) argues, very plausibly, that the causal powers of human desires include (among other things): (i) a role in directing attentional focus; and (ii) causing hedonically charged (pleasant or unpleasant) feelings as the subjective expectation of their being satisfied rises or falls. For example, Homer’s strong desire for beer might cause his attention to be caught by a TV advert for beer. It might also cause him to experience an unpleasant feeling of disappointment upon his discovering there is no beer in the fridge.

How can DTD explain these causal powers of human desire? The third ingredient of DTD is simply the idea that, in addition to the motivational role which D-beliefs possess *essentially*, D-beliefs can also possess other causal powers, such as those Sinhababu mentions, *as a contingent matter of fact*.[[7]](#footnote-7) Why does Homer’s strong desire for beer cause his attention to focus on the TV advert? Because it is identical to his belief that *D([I drink beer],100)*, and because *this D-belief* *causes Homer’s attention to focus on the advert*. Why does the desire cause him to feel disappointment when he learns there is no beer in the fridge? Again, because the desire is identical to the D-belief, and because *the D-belief has this causal power*. Provided the D-concept is phylogenetically ancient, as IDENTITY suggests it must be, then it surely likely that natural selection will, over the aeons, have hardwired numerous such dispositional properties into the human D-concept (and thus into human D-beliefs, and thus into human desires). DTD acknowledges and embraces this implication.

*Part 4.* The next order of business is to say what the D-concept *means*; i.e., to say what its truth-conditions are.

Why do *D*’s truth-conditions matter? To help answer this question let’s return to Homer. Suppose Homer doesn’t initially believe *D([I drink beer],100)*: he doesn’t initially have a strong desire to drink beer and is not strongly motivated to drink it. But he now acquires strong empirical evidencefor the *truth* of *D([I drink beer],100)*. If he is theoretically rational—i.e., disposed to believe what the evidence suggests to be true—then this evidence will prompt him to believe that *D([I drink beer],100)*. IDENTITY implies that he will, in acquiring this D-belief, thereby acquire a strong desire, and motivation, to drink beer. Next, let’s suppose that, after acquiring this D-belief and the associated desire, he now acquires yet further empirical evidence which contradicts and overturns the previous evidence by indicating that *D([I drink beer],100)* is in fact *false*. If Homer is theoretically rational then this new evidence will result in his belief that *D([I drink beer],100)* being *extinguished*. IDENTITY implies that as this D-belief is extinguished, so too will be his strong desire to drink beer.

In short, IDENTITY implies that if Homer is theoretically rational then his desires and his motivations to act will wax and wane *in ways controlled by evidence and against the truth of D-propositions*. This imposes strong constraints on what a theory of D’s meaning must say if it is to be plausible. For example, consider the following absurd hypothesis: *D([I drink beer],100)* is true iff *snow is white*. This hypothesis implies Homer’s desire to drink beer *should be modulated by evidence for and against snow being white*. Show Homer lots of snow which appears white and he should desire to drink beer! Show Homer some snow which instead appears green and his desire to drink beer should be extinguished! The hypothesis is absurd because facts about the colour of snow obviously have no inherent, semantic connection to the desirability of drinking beer.

The moral of the example is that a theory about the meaning of the D-concept can be plausible only if it identifies *the truth-conditions of D(ψ,x)* with *conditions under which it would be appropriate for an agent to desire that ψ with a strength of x*. The task of analysing *D*’s truth-conditional content therefore devolves directly into the task of developing a general, normative theory of desire—a theory capable of telling us, for any *ψ* and *x*, the conditions under which an agent should desire that *ψ* with a strength of *x*.

Here there is good news and bad news. The good news is that well-developed normative theories of desire already exist, so we needn’t construct one from scratch. The bad news is there are *several* such theories, with the question as to *which is* *best* being among the most vexed and controversial in philosophy. Two such theories are now considered by way of illustration—the British empiricists’, and Kant’s.

Here, in the words of Bentham, Locke and Hume, is the view of the British empiricists:

Nature has placed mankind under the governance of two sovereign masters, *pain*, and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do (Bentham 1996, I.1)

For whatsoever good is proposed, if its absence carries no displeasure or pain with it, if a man be easy and content without it, there is no desire of it, nor endeavor after it. (Locke 1995, XX.6)

’Tis from the prospect of pain or pleasure that the aversion or propensity arises towards any object… (Hume 2000, 2.3.3.3)

If this idea is right—if all motivation and all desire is rooted in attraction to pleasure and aversion to pain—then *D*’s truth conditional content is to be analysed roughly as follows:

*Hedonistic Theory of Desirability (HTD):* *D(ψ,x)* is true, relative to an agent, *a*, iff *a* would experience a net total of *x* units of pleasure (positive) and pain (negative) were *ψ* to be the case.[[8]](#footnote-8)

To illustrate the theory of rational agency that emerges when IDENTITY is conjoined with HTD, let’s return to Homer. Homer has—let’s imagine—drunk sufficient beer in the past to have accrued strong inductive evidence that when he drinks beer in the future he will be rewarded with 100 units of pleasure. HTD implies that he has thereby acquired strong evidence for the truth of *D([I drink beer],100)*. Let’s imagine he believes that *D([I drink beer],100)* on this basis. IDENTITY implies that in acquiring this belief he also acquires a desire, of strength 100, to drink beer. Inductive evidence is, of course, sometimes misleading. Let’s imagine it is misleading in the present case, for, although Homer has no way of knowing it yet, he has suddenly become alcohol intolerant: he will experience acute agony next time he drinks beer, not the 100 units of pleasure he expects. His situation is therefore such that although he believes *D([I drink beer],100)*, and thus desires to drink beer, his desire is *misplaced* and identical to *a false belief*—as he will quickly learn to his cost if he succeeds in satisfying it. And indeed (let’s imagine) he *does* succeed in satisfying it. His desire that he drink beer issues, via operation of MER1/MER2, in him being motivated to get off his couch and fetch a beer. He drinks beer, experiences agony, and thereby acquires new evidence which powerfully disconfirms his belief that *D([I drink beer],100)*. In light of this disconfirming evidence, he ceases to believe that *D([I drink beer],100)*, and instead comes to believes some different proposition, such as *D([I drink beer],-500).* IDENTITY implies that as his original D-belief is extinguished, so too, *ipso facto*, is his desire to drink beer.

It is worth pausing briefly at this point to notice two important virtues of DTD. The first concerns *how naturally and effortlessly it explains the acquisition and extinction of desires*. Those who hold that beliefs and desires are independently existing mental states must posit *one* psychological processes (‘theoretical rationality’) to govern what we believe, and another process (‘practical rationality’, or a component thereof) to govern what we desire. But since the DTD identifies desires with beliefs, it implies only the former, theoretical process is needed: for in governing what we believe, this process will automatically govern what we D-believe, and thereby desire, into the bargain. It might have been suspected that by thus bashing the ‘square peg’ of desire into the ‘round hole’ of theoretical rationality we would be left with a confused and incoherent account of the rational governance of desire. But such is not the case. In the above example Homer’s desire to drink beer comes and goes in a way controlled, not by practical rationality, but by theoretical rationality, and yet the resulting story is simple, straightforward, and apparently perfectly unobjectionable. Why believe in two psychological processes when one does the trick all by itself?

The second virtue of DTD is its invulnerability to a well-known ‘directions of fit’ objection levelled against Desire-As-Belief theories by Michael Smith (1994, 118). Smith has us a consider a mental state, *B*, that has both the representational role of a belief and motivational role of a desire. If there were a mismatch between *B* and the world then, *qua* representational state, *B* should change to fit the world. Yet, *qua* motivational state, *B* should remain unchanged and instead cause the agent to change the world. Thus *B* is—so it seems—required both to change and remain unchanged—a contradiction. This argument against Desire-As-Belief theories has an oft-noted loophole (McDowell 1978; Pettit 1987; McNaughton 1988; Price 1989; Little 1997; Tenenbaum 2006; Bromwich 2010; Kriegel 2012; Park 2013) which Smith himself acknowledges (Smith 1994, 118). While *B* would indeed be ‘just plain incoherent’ (ibid.) if both its representational and motivational elements had the same content, *ψ*, the contradiction disappears if the agent is motivated to make *ψ* come true in virtue of representing the world as being a place where *some different* proposition, *f(ψ)*, is true. DTD is immune to Smith’s ‘direction of fit’ argument for precisely this reason. Homer is motivated to drink beer, not by believing that [I drink beer], but by believing that *D([I drink beer], 100)*. In having this belief Homer will be both: (i) motivated (via the operation of MER2) to act in ways that will make [I drink beer] come true (as required by the desire’s world-to-mental-state direction of fit); and (ii) disposed to retract his belief that *D([I drink beer],100)*, and thereby retract his desire, if he acquires evidence that his drinking beer is not, after all, pleasurable (as required by the belief’s mental-state-to-world direction of fit). There is no incoherency in supposing Homer has these two dispositions simultaneously.

Kant’s theory about the conditions under which desires should be acquired and extinguished is more obscure than that of British empiricists. However, it is at least clear that for Kant all rational desire is subservient to the ‘categorical imperative’—the requirement to ‘act only according to that maxim whereby you can, at the same time, will that it should become a universal law’. This suggests the following analysis of *D(ψ,x)*’s truth-conditional content:

*Kantian Theory of Desirability (KTD):* *D(ψ,x)* is true iff a rationally perfect system of universal laws would prescribe, rather than proscribe, agents being motivated, with a strength of *x*, to make it the case that *ψ* is true.

I hesitate to claim that KTD perfectly captures Kant’s views about the rational governance of desire. But even if it captures Kant’s views only approximately it is obvious that British empiricism and Kantianism support radically different analyses of *D*’s truth-conditional content.

How then are we to decide *which theory to use*? Should we follow the British empiricists, and choose HTD? Or should we follow Kant, and choose KTD (or some improved variant thereof)? Or should we opt for some different theory altogether—e.g., a virtue theory along the lines of (McDowell 1978) or (Little 1997)? I am strongly sympathetic to HTD. But where my purposes in the present paper are concerned I don’t want to be tied to any particular normative theory of desire. I instead leave the reader free to choose whichever such theory she prefers. In order to see why, having posited the D-concept, I can nevertheless be agnostic as to its precise truth-conditions, we may distinguish three questions:

1. How, if at all, are facts about what an agent desires related to facts about what she believes?
2. Under what conditions *will* a desire of strength *x* that *ψ* be created, or extinguished?
3. Under what conditions *should* a desire of strength *x* that *ψ* be created, or extinguished?

My answer to (i) is that every desire is identical to a corresponding D-belief (as per IDENTITY). My answer to (ii) is that a desire, of strength *x,* that *ψ*, will be created if and when the agent in question comes to believe that *D(ψ,x)*, and extinguished if and when this D-belief is extinguished. Notice that in answering both these questions I haven’t had to take *any position* as regards *D*’s truth conditions. Therefore in so far as it is my intention *only* to answer (i) and (ii) I can afford to remain neutral between theories like HTD and KTD. It is only when we reach (iii) that *D*’s truth conditions matter. My answer to (iii) is that an agentshould come to desire with strength *x* that *ψ* if and when she acquires adequate confirming evidence for *D(ψ,x)*’sbeing true, and that this desire should be extinguished if and when the confirming evidence is offset by disconfirming evidence. This answer is incomplete since to know *what evidence* would count as confirming or disconfirming *D(ψ,x)* we first need to know what is involved in *D(ψ,x)* *being* true, or false. That is, we need a theory of *D*’s truth-conditions. A reader who wants a complete answer to question (iii) is invited to ‘plug in’ her favourite normative theory of desire at this point—HTD, or KTD, or some other such theory. That is, she can take *D* to mean whatever she thinks it should mean given its role in determining desire and motivation.

While I remain officially neutral as regards *D*’s meaning, I take it that any plausible normative theory of desire will be consistent with the following, deliberately loose and equivocal analysis:

*Generic Theory of Desirability (GTD): D(ψ,x)* is true iff *ψ*’s being the case would be *valuable* (i.e., *good*, or *morally right,* or *attractive*, or *pleasurable*, or *utility-inducing*, or *well-being-inducing*, or *eudemonia-inducing*, or *happiness-inducing,* or *rewarding as opposed to punishing*) to degree *x*.

Philosophers who subscribe to different normative theories of desire will presumably be able to agree in accepting GTD, if only by dint of diverging radically in how they unpack the term, ‘valuable’. In what follows I will speak of *D(ψ,x)* as having the content that *ψ* is ‘valuable’ (or ‘dis-valuable’) to degree *x*, in line with GTD. The scare quotes signal that the meaning of ‘valuable’ is up for grabs.

*Part 5.* A concept, *X*, is said to be ‘intensional’ when there exists another concept, *Y*, that shares *X*’s extension (i.e., its referent, or denotation, or truth-conditions) but has a different intension (i.e., a different *sense*, or *connotation*, or *cognitive significance*). When *X* and *Y* are so-related let’s call them ‘intensional siblings’. Examples of intensional siblings include the ‘Superman’ and ‘Clark Kent’ concepts, and the ‘square root of four’ and ‘least prime number’ concepts. Arguably the vast majority of concepts—perhaps *all* concepts—are intensional, since it is generally easy, for any concept *X*, to gerrymander a new concept, *Y*, that shares *X*’s extension but has a different intension.

The fourth ingredient of the DTD is simply the idea that the D-concept is like most or all other concepts in respect of being intensional. If this is correct then *D* can have intensional siblings that share with it the same extension but have different intensions. Let *E* be some such intensional sibling of *D*. Since *D* and *E* are extensionally equivalent, *D(ψ,x)* and *E(ψ,x)* will be related to each other much as, say, [Superman can fly] and [Clark Kent can fly] are related: they will be truth-conditionally equivalent propositions that nonetheless differ in cognitive significance. Since *D(ψ,x)* has (roughly) the truth conditional content *that ψ is ‘valuable’ to degree x*, so too will *E(ψ,x)*. But because they have different intensions, an agent might believe one of these propositions without believing, or being able to infer, the other. Just as Lois Lane believes that Superman can fly but not that Clark Kent can fly, so an agent might believe that *E(ψ,x)* but not that *D(ψ,x)*, or *vice versa*.

How will the intensions of the D-concept and E-concept differ? It is, as I have said, essential to the D-concept that *D-beliefs produce motivating reasons by engaging MER2*. This motivational ‘pushiness’ will be a vital and distinctive component of *D*’s intension, with the ‘felt violence or intensity’ (Charlton 1988, 127) of a desire grounded in the belief that D(*ψ*,*x*) being proportionate to *x*. The E-concept, in contrast, *will not* have motivational pushiness as part of its intension. Why not? *Because MER2 interfaces with an agent’s D-beliefs, not with her E-beliefs.* (When MER2 is ‘scanning’ an agent’s beliefs, it will, so to speak, ‘see’ only her D-beliefs and her means-end beliefs, not her E-beliefs.) Thus D-beliefs will have motivational ‘oomph’ or ‘kick’, while E-beliefs won’t. Because E-beliefs don’t interface with MER2 and therefore lack motivational force, *they are not identical to desires*. For this reason an agent who believes that *E(ψ,x)* but not that *D(ψ,x)* will thereby believe that *ψ* is ‘valuable’ to degree *x* *in such a way that she will not, in so doing, desire with strength x that ψ*.

This idea—that the D-concept is intensional and that its intensional siblings share its truth-conditions but not its motivational role or relation to desire—is important because it helps explain cases like the following that could otherwise be construed as counterexamples to IDENTITY:

*The Addict.* An unwilling addict strongly desires to take heroin, but believes taking heroin to be *very bad*, not *good* (Frankfurt 1971). What is going on here? Well, there is (we may suppose) a mismatch between what she *D*-believes and what she *E*-believes. Specifically, she believes that *D([I take heroin],1000)*, thus having an overwhelmingly powerful desire to take heroin, but at the same time believes that *E([I take heroin],-500)*, thus believing taking heroin to be of very great ‘disvalue’. The *D*-belief has motivational force (because of MER2), while the *E*-belief doesn’t (since MER2 engages only *D*-beliefs, not *E*-beliefs), and so she is powerfully motivated to take the drug despite knowing full well she oughtn’t.

*The Depressive.* A depressive believes that her trying to get a job would be very ‘valuable’. Yet she is listless and lacks any corresponding motivation to try to get a job. What is going on? Again, we can posit a mismatch between what is *D*-believed and what is *E*-believed. She believes that *E([I try to get a job], 100)*, and so believes going job hunting to be ‘valuable’. However she doesn’t have the corresponding motivationally forceful belief that *D([I try to get a job], 100)*, and so doesn’t desire to go job hunting.

*The Amoralist.* An amoralist happens to be an expert where moral facts are concerned, so she knows as well as anyone that her being kind to others is ‘valuable’. But because she is amoral, she has no corresponding desire or motivation to be kind to others. Again, this can be explained in terms of a mismatch between *D*-beliefs and *E*-beliefs. She believes that *E([I am kind to others], 100)*, but not that *D([I am kind to others], 100)*, and so she lacks the desire and motivation to be kind.

In each of these cases the apparent misalignment between an agent’s desires and her beliefs about ‘value’ is explained in terms of her *E*-beliefs being misaligned with her *D*-beliefs.

*Part 6.* Beliefs are sometimes *unruly*, which is to say they sometimes refuse to pop into existence or pop out of existence at the behest and command of theoretical rationality. Perceptual illusions are one case in point. For example, I believe on an intellectual level that the two central circles of the Ebbinghaus illusion (Figure 1) are of identical sizes, having carefully measured them with a ruler.

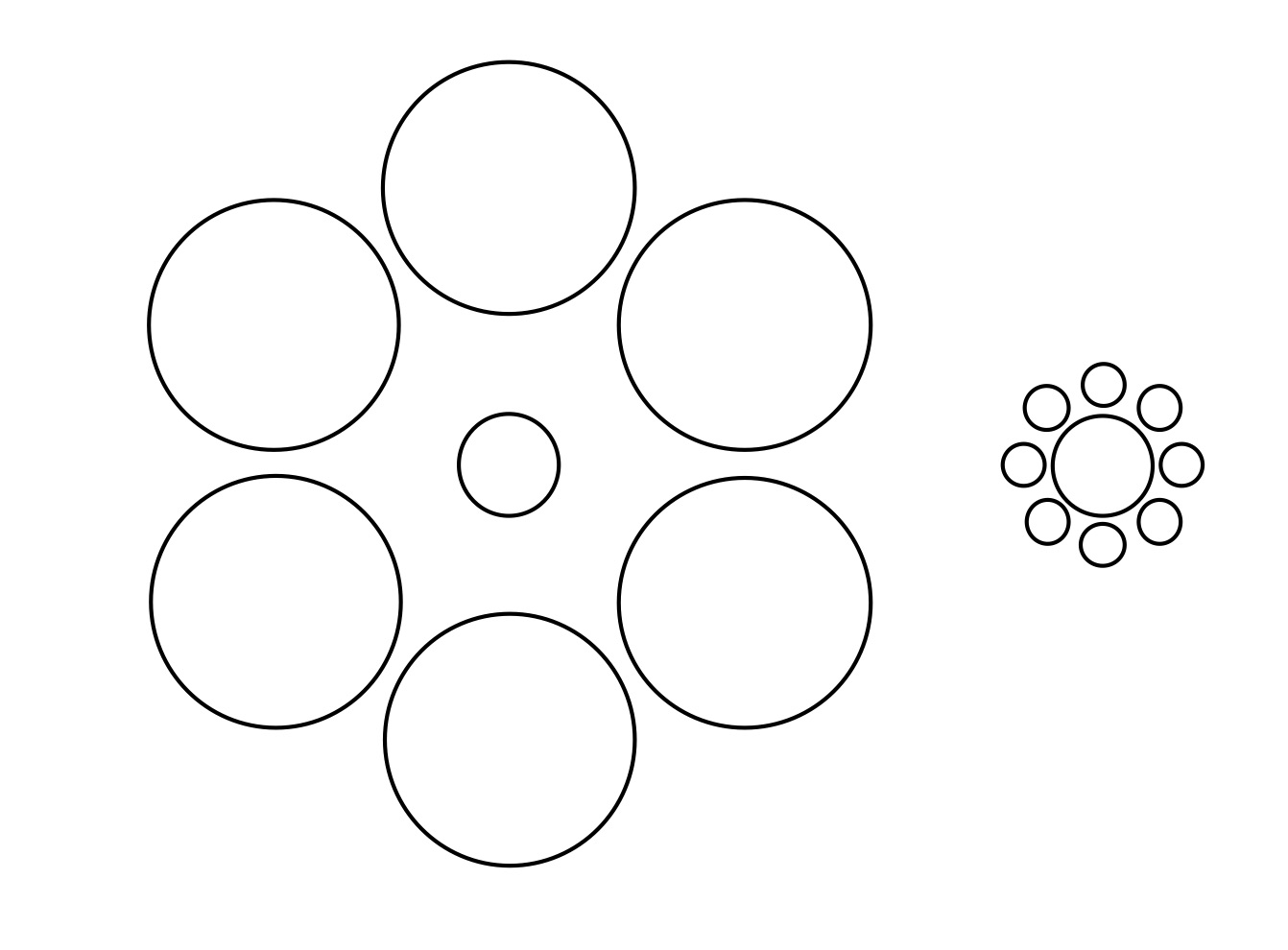


Figure 1. The Ebbinghaus Illusion

Yet when I look at the circles I can’t help having the perceptual belief that the one on the left is smaller than the one on right. This perceptual belief is an unruly belief, that persists no matter how firmly my rational faculties repudiate it. Moreover, if I reach a hand to grasp the circle on the left my motor system will automatically pinch my finger and thumb closer together than if I reach to grasp the circle on the right (Franz, Bulthoff, and Fahle 2003). The belief that feeds into my motor system is the false, unruly belief, not the true, rationally accredited belief.

The phenomenon arises again in relation to racist and sexist prejudices. Few of us admit to being racist or sexist. On an intellectual level we may indeed sincerely believe, on the basis of compelling evidence, that superficial attributes like gender and skin colour are uncorrelated with human virtues like intelligence, trustworthiness and compassion. But Implicit Association Tests show that in most people these intellectually informed beliefs are accompanied by subconscious prejudices that cause discriminatory behaviour (Greenwald, McGhee, and Schwartz 1998). Moreover, recognizing one has these prejudices is of no significant help in eliminating them. They persist as unruly beliefs even when rationality expressly disavows and disowns them.

Other examples of unruly beliefs include the mental states that Tamar Gendler (2008a; 2008b) calls ‘aliefs’. For example, a prospective bungy jumper knows perfectly well that the bungy cord won’t snap, and that if he jumps then in a few seconds he will be lowered safely to the ground, happy and exhilarated. He knows, in short, that jumping is safe. Yet on a more visceral level he at the same time believes (or ‘alieves’) that jumping is *not safe*.[[9]](#footnote-9) Were it not for the latter belief, jumping would be easy. But jumping is not easy. The idea of it fills him with terror. The two contradictory beliefs battle within him as he tries to marshal his courage, the unruly belief that jumping is dangerous vying powerfully with the rationally endorsed belief that jumping is safe despite its lacking any rational foundation and despite his *knowing full well* it lacks any rational foundation.

The fifth and final ingredient of DTD is the idea that D-beliefs are similar to these other types of belief (perceptual beliefs, beliefs about the characters of others, and aliefs) in the respect that they too are often unruly. This idea is a useful addition to DTD because it helps explain how a mismatch between what an agent desires and what she believes to be ‘valuable’ *can be sustained over time*.

Consider, for example, the unwilling addict who believes that *D([I take heroin],1000)* while simultaneously believing that *E([I take heroin],-500)*. She can consistently believe both these propositions simultaneously only for so long as she doesn’t know that *D* and *E* are intensional siblings. But suppose we educate her on this point, teaching her that *E([I take heroin],-500)* entails *D([I take heroin],-500).* She is now in a position to infer that *D([I take heroin],-500)* is true, and thus that *D([I take heroin],1000)* is false. So it seems she should now stop believing the latter proposition, removing her desire and curing her addiction.

Clearly, however, this won’t work in practice. Heroin addictions can’t be cured by conceptual analysis, and for an obvious reason: viz., the addict’s desire for heroin is an *unruly desire*. It is cemented firmly in place. She knows it is a misplaced desire but rational reflection cannot shake or weaken it.

How are such unruly desires to be explained? DTD’s answer is simply that the unruliness of a desire is identical to the unruliness of a D-belief that the desire is identical too. The addict desires heroin because she believes that *D([I take heroin],1000)*. Her desire is unruly for the simple reason that the D-belief is unruly. The phenomenon of unruly desire is thereby explained as being a consequence of a more general phenomenon of unruly belief. The addict’s addiction is explained on the same model as, say, the Ebbinghaus illusion or involuntary racism.

Again, notice how economical DTD is. Unruly desires obviously exist, and their existence is plainly crucial to explaining how an agent like the addict can be motivated to perform actions she explicitly judges to be of great ‘disvalue’ (McNaughton 1988, 125). Unruly beliefs obviously exist too. On the theory that desires and beliefs are distinct existences these would amount to two independent phenomena requiring two independent explanations. DTD instead explains the former phenomenon as a special case of the latter.

To summarize, DTD consists of the idea that agents possess D-beliefs which: (i) motivate via the operation of MER2; (ii) constitute the reductive ground of desires; (iii) potentially possess various causal roles over and above their motivational role; (iv) have content about the ‘value’ of states of affairs; (v) are intensional; and (vi) are potentially unruly.

**3. DTD’s advantages over rival Desire-As-Belief theories**

D-beliefs are *evaluative beliefs*, for, as explained above, to believe that *D(ψ,x)* is to believe *ψ*’s being the case would be ‘valuable’ to degree *x*. DTD therefore implies *that certain evaluative beliefs* (namely, D-beliefs ) *can combine with means-end beliefs to produce motivating reasons* (as *per* MER2).

This idea, that some evaluative beliefs are motivationally ‘pushy’, is not novel. Such theories are divisible into three categories depending on how they relate ‘pushy’ evaluative beliefs to desires. (i) Desire-As-Belief theories (several of which will be described below) *identify* desires with pushy beliefs. DTD belongs to this category. (ii) ‘Hybrid’ theories deny that that desires are identical to pushy evaluative beliefs, and instead hold desires and pushy evaluative beliefs to be wholly distinct mental states *either* of which can play the role of motivator within means-end reasoning (Nagel 1970; McDowell 1978; Scanlon 1998; Shafer-Landau 2003).[[10]](#footnote-10) (iii) Other theories imply that *only* pushy evaluative beliefs can play the motivational role, and locate desires either upstream or downstream of means-end reasoning—e.g., by identifying them with *perceptions of value that feed into means-end reasoning* (Tenenbaum 2006), or with *the motivations to act that issue from means-end reasoning* (Dancy 1993).

Since DTD is a Desire-As-Belief theory, Desire-As-Belief theories will be the focus from now on. Extant theories of this kind include:

1. Humberstone (1987). For an agent to desire that *ψ* is for her to believe *that Dψ,* where *Dψ* means roughly ‘it is desirable that *ψ*’ or ‘would that it were that *ψ*’, but with *D*’s precise content being left unspecified.[[11]](#footnote-11)
2. McNaughton (1988, 112). For an agent to desire that *ψ* is for her to believe *that ψ’s being true is attractive*.
3. Alex Gregory (forthcoming). For an agent to desire to *ϕ* is for her to believe *that she has normative reason to ϕ*.

The DTD has several major advantages over these rival Desire-As-Belief theories.

*First advantage: explaining how evaluative beliefs motivate*

How do evaluative beliefs acquire the motivational ‘oomph’ that transforms them into desires? Humberstone, McNaughton and Gregory don’t address this question. Their theories say that some such beliefs *are* motivationally forceful, but they don’t explain how the forcefulness springs into being. DTD, in contrast, says exactly where a D-belief’s ‘oomph’ comes from—namely, MER2. MER2 is a simple rule. It is naturalistically ‘kosher’, in the sense that it is not mysterious how it could be implemented in the brain, alongside MAX, in such a way that D-beliefs and means-end beliefs combine to create instrumentally rational action. Any suspicion one might have had that there is something conceptually ‘fishy’ or incoherent in the very idea *of a belief being motivationally forceful* should be put to rest by MER2. (Why couldn’t MER2 be implemented in the brain? Why not indeed! It *could*. And if it *was* implemented in the brain, then D-beliefs *would* be motivationally forceful. Therefore, suspicions to the effect that it is somehow *just plain impossible* for a belief to be motivationally forceful are unfounded. Q.E.D.)

*Second advantage: answering the objection from modal separability*

An obvious general objection to Desire-As-Belief theories concerns the apparent possibility of an evaluative belief existing without the corresponding desire existing, or *vice versa* (Stocker 1979; Smith 1994). As run against DTD, this objection would involve defending one or both the following claims:

*Desire Without Belief* (DWB): it is possible for an agent to desire with strength *x* that *ψ* without believing that *D(ψ,x)*.

*Belief Without Desire* (BWD): it is possible for an agent to believe that *D(ψ,x)* without desiring with strength *x* that *ψ*.

The case of the unwilling addict might appear to support DWB, for the addict desires to take heroin despite believing this to be of strong ‘dis-value’. Similarly, the cases of the depressiveandthe amoralist might be cited as evidence for BWD, for the depressive and the amoralist believe an outcome to be ‘valuable’ without desiring it.

I have already explained at length how DTD overcomes this objection. To recap, one ingredient of DTD is the idea that the D-concept is *intensional*. This enables me to concede to my opponent that the addict, depressive and amoralist have misaligned desires and beliefs. I can make this concession without conceding that there is any misalignment between what is desired and what is *D-believed*, by holding that the misaligned beliefs are instead *E-beliefs* (which are truth-conditionally equivalent to D-beliefs but motivationally inert because they don’t engage MER2). Another ingredient of DTD is the idea that D-beliefs are sometimes *unruly* and resistant to rational control. The hypothesis that the addict, depressive, and amoralist have unruly D-beliefs explains why they can’t bring their D-beliefs and E-beliefs into alignment via rational inference even when fully informed of all the relevant facts.

This response to the objection from modal separability is, I believe, far superior to the responses provided by other Desire-As-Belief theorists. Humberstone doesn’t deal with the objection at all. Gregory’s attempted response rests on the idea that the *strength* of a desire can potentially fail to align its *motivational force* (forthcoming, 11). For example, Gregory would claim of the depressive that: (i) since she believes she has strong normative reasons to get a job she has a strong desire to get a job; but (ii) this desire, although strong, is *motivationally* weak, and hence unable to spur her to action. However, as Sinhababu writes:

The distinction between desire and motivation doesn’t allow this move. Desire and motivation are different things, but the circumstances under which desires motivate us are well-understood. Strong desires cause action when combined with strong means-end beliefs. (forthcoming, 39)

Any Desire-As-Belief theory worthy of the name must be able to explain how the *motivational forcefulness* of a desire is determined by properties of the evaluative belief it is identical to. Gregory’s theory appears not to satisfy this criterion, since it explains only the (so-called) ‘strength’ of a desire, not its motivational forcefulness.

McNaughton’s (1988, 118–131) response to the modal separability objection is closer in spirit to mine. In explaining cases like the unwilling addict he notes that ‘Desires … do not come and go at our behest’ and that ‘They are not under the direct control of the will’ (p. 125). He is thereby noting, in effect, that the addict’s desires are *unruly*. However he does not suggest, as I do, that the unruliness of desire in to be explained as an instance of the more general phenomenon of *unruly belief*. Nor does he attempt to explain apparent cases of beliefs and desires being ‘separated’ (e.g., cases like the addict, depressive and amoralist) by suggesting that the beliefs in question are *intensional*.

*Third advantage: the desires of animals and infants*

Desire-As-Belief theorists face a dilemma. The general form of a Desire-As-Belief theory is this: *an agent desires that ψ (or desires to ϕ) iff it believes that \_\_\_)*. Different theorists substitute different things in place of ‘\_\_\_’. One option is to substitute an expression of natural language. McNaughton and Gregory take this approach, with McNaugton substituting ‘*ψ*’s being true is attractive’ and Gregory substituting ‘it has normative reason to *ϕ*’. This option is problematic because it implies an agent must lack desires if it lacks the concepts expressed by the phrase of natural language in question. There is surely ample reason to doubt that animals or human infants possess the concepts we express when we speak of a ‘normative reason’ or of a proposition’s coming ‘true’ being ‘attractive’. This being so, McNaughton’s and Gregory’s theories have difficulty explaining how animals or human infants can possess desires (as they plausibly do).

The other option is to substitute for ‘\_\_\_’ a logical construct that is *not* expressible in natural language. Humberstone takes this approach. He holds that an agent desires that *ψ* by virtue of believing that D*ψ*, and he maintains that the truth-conditions of the D-concept are (probably) *inexpressible* in natural language (1987, 50). The implication is that animals and infants who lack the concepts that we express with natural language do not therefore necessarily lack the D-concept (it not *being* one of these concepts).

Now perhaps the D-concept’s truth conditions are inexpressible, just as Humberstone suggests. (This is an issue on which, for the purposes of this paper, I am neutral.) But *if they are* inexpressible this is bad news for theorists seeking to understand desire, because it follows that attempts to *reductively analyse* D’s truth-conditional content, by identifying it with the truth-conditional content of some synonymous expression of natural language, are doomed to failure. This is an implication Humberstone (1987, 50) notes and endorses. For example, it would follow that no fully accurate precisification of GTD could ever be given, and that such theories as HTD and KTD, which attempt to provide such a precisification, are false.

Is there any way of escaping this dilemma: of identifying desires with beliefs having a truth-conditional content expressed by some English sentence, *S*, but while also holding that desires can be possessed by creatures who lack the concepts expressed by *S*? Yes. DTD achieves this result by exploiting the idea that the D-concept is *intensional*. To see how this explanation works, imagine that *E* is some intensional sibling of *D*. The following theory will in this case successfully capture *D*’s truth conditional content: ‘*D(ψ,x)* is true iff *E(ψ,x)* is true’. Provided *E* is some concept that can be expressed in natural language, then this theory will be one we can potentially discover and understand. But from the fact that *D* and *E* are intensional siblings and share the same truth-conditional content, it nowise follows that an agent can possess *D* only if it also possesses *E*. It is instead possible that animals and infants who don’t possess the E-concept, and who are therefore in no position to understand or express the correct theory about *D*’s truth-conditions, might nevertheless possess the D-concept and have desires identical to D-beliefs.

*Fourth advantage: explaining variations in motivational force*

On Humberstone’s theory, one desires that *ψ* iff one believes that *Dψ*. How, on this view, are we to explain how different agents can desire that *ψ* *with different strengths*? Humberstone doesn’t say, but a natural suggestion might be that the strength of an agent’s desire that *ψ* is proportionate to the strength (i.e., certitude) of her belief that *Dψ*. However this suggestion runs afoul of an argument against Desire-As-Belief due to Nick Zangwill (2008). Zangwill imagines a scenario involving two agents, both of whom believe *equally strongly* that taking bribes is immoral, but the first of whom ‘cares less about morality’ than the second. Hence the first accepts a bribe while the second refuses. In order to explain the difference in these agents’ motivations it seems Humberstone must posit a difference in their D-beliefs, but, by hypothesis, both agents believe, equally strongly, that *D*[I don*'*t take bribes]. This appears to leave Humberstone no choice but to deny that Zangwill’s scenario is possible in the first place.

DTD, in contrast, easily accommodates Zangwill’s case. As we have seen, DTD explains variations in the motivational force of desires in terms of variations in the value of the scalar variable, *x*, within a belief that *D(ψ,x)*. It thus implies that *A* and *B* might, while both believing with equal certitude that taking bribes is of ‘dis-value’, nevertheless disagree as to *how dis-valuable* it is, and be motivated differently as a result. E.g., *A* might be believe that *D([I take bribes],-100)* is true, while *B* might believe, albeit with *equal certitude*, that *D([I take bribes],-200)* is true. *B* will in this case be more strongly motivated than *A* not to take bribes, notwithstanding their being equally confident that taking bribes is ‘dis-valuable’.[[12]](#footnote-12)

*Fifth advantage: explaining the multifarious causal powers of desire*

Neil Sinhababu (forthcoming) attacks Desire-As-Belief theories on the basis that desires and beliefs are radically unalike. ‘Understanding desire as a kind of belief with a few properties unusual for beliefs is,’ he says, ‘almost as implausible as understanding desire as a kind of cheese with a few properties unusual for cheeses’ (p. 40). The properties of desire he has in mind are causal powers such as: (i) being motivationally efficacious; (ii) helping direct attention; and (iii) triggering hedonically charged feelings, like gratification and disappointment, when the subjective expectation of their being satisfied rises or falls.

As explained in §2, DTD meets this objection head on by incorporating the idea that (human) D-beliefs have the various causal powers of desire which Sinhababu mentions. Does their having these causal powers make D-beliefs *unusual* among beliefs? Yes! Should the proponent of DTD be embarrassed by this fact? No! To the contrary, it would be embarrassing if D-beliefs *did not* have these causal powers, for then the theory that desires are identical to D-beliefs would be manifestly false. A D-belief might stand to other beliefs rather as a platypus stands to other mammals, but a platypus *is* a mammal for all that it lays eggs and has a bill, and, similarly, a D-belief *is* a belief for all that it has causal powers that other beliefs don’t have. It is a belief for the simple reason that it has *truth-conditions*—namely, the truth-conditions loosely described by GTD (and perhaps more precisely described by HTD or KTD). It nowise follows from D-beliefs being *peculiar* beliefs that they are *not* beliefs.

Might Humberstone, McNaughton and Gregory reply to Sinhababu along similar lines? Humberstone might do so, since he can hold, just as I do, that the D-concept is peculiar among concepts in being phylogenetically ancient and in having had multifarious causal powers hardwired into it by evolution. McNaughton and Gregory, in contrast, are poorly placed to respond adequately to Sinhababu’s argument, it being far from obvious why a belief *that ψ is attractive* (McNaughton) or *that one has normative reason to ϕ* (Gregory) would have causal powers so sharply different from those of ordinary, common-or-garden beliefs.

**4. Conclusion**

There are things we label ‘desires’. DTD implies these things can also be labelled ‘D-beliefs’. The discovery that two labels attach to one thing is often of far-reaching scientific or philosophical importance—as, for example, when it was learned that ‘heat’ and ‘rapid molecular vibration’ are two names for one phenomenon. Major implications of the theory that desires are nothing but D-beliefs include these: (i) in so far as it suggests Attitudinal Monism is true, it paves the way for the attitudinal half of the problem of naturalizing the propositional attitudes to be dissolved (see §1); (ii) in showing how beliefs can play the role of motivator in means-end reasoning (via the operation of MER2) it paves the way to a cognitivist internalist solution to Smith’s ‘moral problem’ (again, see §1); (iii) it implies that the faculty of instrumental reason will, in governing what we believe, thereby automatically govern what we desire into the bargain, so securing a major advance in theoretical economy by implying that practical reason needn’t contain an additional process dedicated to governing desire (see §2); and (iv) it implies that the phenomenon of unruly desire can be explained as being an immediate consequence of the more fundamental phenomenon of unruly belief, thereby securing us another (albeit closely related) advance in theoretical economy (again see §2).

Given DTD has both these capacities to solve major philosophical problems and these advantages of theoretical economy, those who would reject it need *good reasons* for doing so. What good reasons might they cite? Not, it seems, Smith’s ‘directions of fit’ objection, or the ‘modal separability’ objection, or Zangwill’s objection from motivational strength, or Sinhababu’s objection from the causal powers of desire, or the objection that the very idea of a belief being motivationally forceful is somehow just plain incoherent—because all these objections have been considered above and shown to have no teeth against DTD.[[13]](#footnote-13) If there are good reasons why DTD should be repudiated I believe they have yet to be articulated.

**Acknowledgements**

Many thanks to Tim Bayne, Susan Dimock, Carolyn Mason, Elijah Millgram, Michael-John Turp, and especially Neil Sinhababu, for comments and suggestions.

**Bibliography**

Anscombe, Elizabeth. 1957. *Intention*. Oxford: Basil Blackwel.

Bentham, Jeremy. 1996. *An Introduction to the Principles of Morals and Legislation: The Collected Works of Jeremy Bentham*. Oxford University Press Uk.

Bromwich, Danielle. 2010. “Clearing Conceptual Space for Cognitivist Motivational Internalism.” *Philosophical Studies* 148 (3): 343–67.

Charlton, William. 1988. *Weakness of Will*. B. Blackwell.

Dancy, Jonathan. 1993. *Moral Reasons*. 178. Blackwell.

Daskal, Steven. 2010. “Absolute Value as Belief.” *Philosophical Studies* 148 (2): 221–29.

Frankfurt, Harry G. 1971. “Freedom of the Will and the Concept of a Person.” *Journal of Philosophy* 68 (1): 5–20.

Franz, VH, HH Bulthoff, and M Fahle. 2003. “Grasp Effects of the Ebbinghaus Illusion: Obstacle Avoidance Is Not the Explanation.” *Experimental Brain Research* 149 (4): 470–77. doi:10.1007/s00221-002-1364-6.

Gendler, Tamar Szabó. 2008a. “Alief and Belief.” *Journal of Philosophy* 105 (10): 634–63.

———. 2008b. “Alief in Action (and Reaction).” *Mind and Language* 23 (5): 552–85.

Greenwald, AG, DE McGhee, and JLK Schwartz. 1998. “Measuring Individual Differences in Implicit Cognition: The Implicit Association Test.” *Journal of Personality and Social Psychology* 74 (6): 1464–80. doi:10.1037/0022-3514.74.6.1464.

Gregory, Alex. forthcoming. “Might Desires Be Beliefs About Normative Reasons?” In *The Nature of Desire*, edited by J. Deonna and F. Lauria. Oxford University Press.

Humberstone, I. L. 1987. “Wanting as Believing.” *Canadian Journal of Philosophy* 17 (March): 49–62.

———. 1992. “Direction of Fit.” *Mind* 101 (401): 59–83.

Hume, David. 2000. *A Treatise of Human Nature*. Oxford University Press.

Kriegel, Uriah. 2012. “Moral Motivation, Moral Phenomenology, And The Alief/Belief Distinction.” *Australasian Journal of Philosophy* 90 (3): 469–86.

Lewis, David. 1988. “Desire as Belief.” *Mind*, New Series, 97 (387): 323–322.

———. 1996. “Desire as Belief II.” *Mind* 105 (418): 303–13.

Little, Margaret Olivia. 1997. “Virtue as Knowledge: Objections From the Philosophy of Mind.” *Noûs* 31 (1): 59–79.

Locke, John. 1995. *An Essay Concerning Human Understanding*. Oxford University Press.

McDowell, John. 1978. “Are Moral Requirements Hypothetical Imperatives?” *Proceedings of the Aristotelian Society* 52: 13–29.

McNaughton, David. 1988. *Moral Vision*. Oxford: Blackwell.

Mele, Alfred R. 1998. “Motivational Strength.” *Noûs* 32 (1): 23–36.

Nagel, Thomas. 1970. *The Possibility of Altruism*. 86. Oxford Clarendon Press.

Park, Seungbae. 2013. “Against the Besire Theory of Moral Judgment.” *Organon F: Medzinárodný Asopis Pre Analytickú Filozofiu* 20 (1): 5–17.

Pettit, Philip. 1987. “Humeans, Anti-Humeans, and Motivation.” *Mind* 96 (384): 530–33.

Platts, M. 1979. *Ways of Meaning*. London: Routledge and Kegan Paul.

Price, Huw. 1989. “Defending Desire-as-Belief.” *Mind*, New Series, 98 (389): 119–27.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Vol. 66. 1. Belknap Press of Harvard University Press.

Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. 4. Oxford University Press.

Sinhababu, Neil. forthcoming. *Humean Nature: How Desire Explains Action, Thought, and Feeling*. Oxford: Oxford University Press.

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

Stocker, Michael. 1979. “Desiring the Bad: An Essay in Moral Psychology.” *Journal of Philosophy* 76 (12): 738–53.

Tenenbaum, Sergio. 2006. “Direction of Fit and Motivational Cognitivism.” In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 235–64. Oxford University Press.

Weintraub, Ruth. 2007. “Desire as Belief, Lewis notwithstanding.” *Analysis* 67 (2): 116–22.

Zangwill, Nick. 2008. “Besires and the Motivation Debate.” *Theoria* 74 (1): 50–59.

1. This formulation of the inconsistent triad is adapted from (McNaughton 1988, 23), (Smith 1994, 12 & 119–120) and (Kriegel 2012). [↑](#footnote-ref-1)
2. I use [n] to denote the proposition expressed by sentence number n and […\*…] to denote the proposition expressed by the sentence ‘…\*…’. [↑](#footnote-ref-2)
3. Adapted from Smith (1994, 93). Smith’s formulation does not account for variations in a desire’s *motivational* *strength*. For a discussion of what motivational strength is and why an adequate theory of desire must explain it, see (Mele 1998). [↑](#footnote-ref-3)
4. In the examples that follow I will simplify by supposing that the scaler value is some real number, like 100 (big) or 1000 (extremely big). In reality it is most unlikely the scale used by the human mind allows for the exactitude real numbers permit. [↑](#footnote-ref-4)
5. MER2 needs to be modified as *per* decision theory in order to be able to cope with beliefs held by an agent with varying degrees of *credence*. I leave this complication aside (as do proponents of MER1). [↑](#footnote-ref-5)
6. If an agent believes that *D(ψ,x)* where *x* is some *negative value*, then IDENTITY implies that the agent will have a *negative desire* that *ψ*, or, in other words, a desire that ¬*ψ*. If an agent has such a negative desire and also believes that by *φ*-ing she will make it the case that *ψ*, then MER2 will issue in a motivating reason to *φ* that is of *negative strength*—i.e., a motivating reason to *refrain* from *φ*-ing. [↑](#footnote-ref-6)
7. Sinhababu and I agree that some causal powers of desires are essential and others are accidental, but disagree as to *which* are essential. I hold that the *only* causal power essential to desire is (MER2-based) motivational role. Sinhababu holds that *both* the motivational role *and* the power of causing hedonically charged feelings are essential (forthcoming, 30). This difference is merely terminological. However, my (more liberal) usage of ‘desire’ makes more sense provided the phenomenon one wants to refer to is the *entire class* of intentional mental states that serve as motivators in means-end rationality, not just the subcategory of this class which also cause hedonic feelings in the right way. [↑](#footnote-ref-7)
8. Here, ‘pleasures’ and ‘pains’ obviously include not just bodily pleasures and pains but also the positive and negative hedonic valances that tinge many affects and emotions. A more sophisticated version of HTD would allow for temporal discounting. [↑](#footnote-ref-8)
9. Gendler (2008a; 2008b) would say that the bungy jumper *believes* jumping to be safe but *alieves* it to be dangerous, with ‘alief’ being some sort of poor cousin of belief. However, her definition of ‘alief’ leaves it unclear what differentiates aliefs from beliefs (Sinhababu forthcoming). As I use the term ‘belief’, Gendler’s ‘aliefs’ count as a sub-species of belief. (Perhaps what she means by ‘alief’ matches what I mean by ‘unruly belief’. I’m not sure.) [↑](#footnote-ref-9)
10. A serious objection to hybrid theories is that they multiply states capable of playing the motivational role beyond necessity (Dancy 1993; McNaughton 1988, 130). Neither McDowell nor Shafer-Landau commit wholeheartedly to the hybrid view. Both leave the door open to Desire-As-Belief being correct instead (McDowell 1978, 24–25; Shafer-Landau 2003, 134). [↑](#footnote-ref-10)
11. DTD might be thought of as a *logical descendent* of Humberstone’s theory (albeit I developed DTD and the notion of the D-concept before reading Humberstone’s work). Both Humberstone and I posit a D-concept. We both identify desires with D-beliefs, and both remain officially neutral as regards its truth-conditional content. Beyond this, however, the similarities between our theories are slight. [↑](#footnote-ref-11)
12. Perhaps Zangwill would respond by insisting that—at least as he conceives of *A* and *B*—not only do they both believe with equal strength that taking bribes is ‘bad’, but they also agree as to *how bad* it is. But the scenario then becomes similar to that of the depressive, and can be dealt with along the same lines. E.g., we may suppose that *A* and *B* both believe, with equal certitude, that *E([I take bribes],-200)*. Hence both believe equally strongly that taking bribes is ‘very dis-valuable’. But whereas *B* also believes that *D([I take bribes],-200)*, *A* instead believes that *D([I take bribes],-100)*. *A* is therefore less strongly motivated than *B* to refuse bribes. [↑](#footnote-ref-12)
13. I have omitted discussion of Lewis’ (1988; 1996) decision-theoretic objection against Desire-As-Belief theories because, albeit I think Lewis’ argument is severely flawed, explaining why is an article-length task in its own right. However for resent critiques of Lewis, see (Weintraub 2007) and (Daskal 2010). [↑](#footnote-ref-13)