Desire as Belief

DAVID LEWIS

1. *The anti-Humean challenge*

A Humean thesis about motivation says that we are moved entirely by desire: we are disposed to do what will serve our desires according to our beliefs. If there were no desires to serve, we would never be moved more to do one thing than another. Whatever might happen then would be entirely unmotivated. Here I shall uphold Humeanism against one sort of opponent.

Our Anti-Humean challenges us with this case. The Department must choose between two candidates for a job, Meane and Neiss. Neiss is your old friend, affable, sensible, fair-minded, co-operative, moderate, ... Meane is quite the opposite. But it is clear that Meane is just a little bit better at philosophy. Gritting your teeth and defying all desire, you vote for Meane, because you believe that Meane getting the job instead of Neiss would, all things considered, be good. Your belief about what's good has moved you to go against your desire to have Neiss for a colleague and to have nothing to do with Meane.

We Humeans reply that there are desires and there are desires. Some desires, for instance your desire to have Neiss for a colleague, are warm—you feel enthusiasm, you take pleasure in the prospect of fulfilment. Other desires, for instance your desire to hire the best available candidate, are cold. Nobody ever said that only the *warm* desires can move us. It is not so that you defied *all* desire when you voted for Meane. You were moved entirely by your desires, however the cold desire outweighed the warm one. We are within our rights to construe 'desire' inclusively, to cover the entire range of states that move us, including for instance the state that moved you to vote for Meane. Humeanism understood in this inclusive way is surely true—maybe a trivial truth, but a trivial truth is still a truth.

Let our Anti-Humean grant that the state that moved you was after all, inclusively speaking, a desire. He may insist, however, that it was *also* a belief: the belief (as he said before) that Meane getting the job would be good. Although it may be true—trivially, he sneers—that all motivation is by desire, it is *also* true that some motivation is by belief. Sometimes, what happens is that we do what will serve the good according to our beliefs about what would be good together with our other beliefs—no desire, other than desires which are identical with beliefs, need enter into it.

More cautiously, he might say that some beliefs are, at least, necessarily

Mind, vol. xcvii, no. 387, July 1988

conjoined with corresponding desires. If you believe that Meane getting the job would be good, then necessarily you desire that Meane get the job. This need not be your only relevant desire (as the story shows). It need not be your strongest desire (though in the story it was). But it must be there. It is just impossible to have a belief about what would be good and lack the corresponding desire.

If the belief and the desire are identical, a fortiori they are necessarily conjoined. Or the necessary connection might arise in some other way, even if the desire is in some way different from the belief. To cover both cases at once, let us take the necessary connection to be our Anti-Humean's main thesis, leaving identity as an optional extra.[1]

Let us suppose that what our Anti-Humean proposes is a necessary connection not with 'basic' desire for what is considered good 'in itself', but rather with desire that may be instrumental. For in the example, you did not desire Meane's appointment for its own sake; you were interested in the excellence he would add to the Department. And let us suppose that what our Anti-Humean proposes is a connection with an averaged desire that takes account of a range of cases, some better than others and some more likely than others. For in the example, it may be that you considered the case that Meane joined the Department but stopped doing good work, and also the case that he came and surpassed all that he had done before, and you decided what you thought about a probability-weighted average of these and other cases. Only on the basis of that average did you believe that Meane getting the job would be good. Only on the basis of that average did you desire that Meane get the job.

Our Anti-Humean has not yet offered any informative analysis of the content of beliefs about what would be good. Maybe he thinks this can be done, maybe not. But he says that we have one handle, at any rate, on the distinctive content of such beliefs: the proposition that Meane getting the job would be good is that proposition X, whatever it may be, such that believing X is somehow necessarily connected with desiring that Meane get the job.

Our Anti-Humean may say how intuitive it seems that a belief about what is good should be necessarily connected with desire, and how right it seems to explain your vote by saying that you believed that Meane getting the job would be good. We can counter in one of two ways. Maybe (1) a so-called 'belief about what would be good' is called a belief by courtesy, but rightly speaking it is not a belief at all but rather it is the

[1] For criticism and defence of several Anti-Humean views at least close to the Desire-as-Belief Thesis, see Michael Smith, 'The Humean Theory of Motivation', *Mind*, 1987, pp. 36–61; Philip Pettit, 'Humeans, Anti-Humeans, and Motivation', *Mind*, 1987, pp. 530–3; and Michael Smith, 'On Humeans, Anti-Humeans and Motivation: a Reply to Pettit', *Mind*, this issue. For critical discussion of an explicit desire-belief identity thesis, see J. E. J. Altham on 'besires': 'The Legacy of Emotivism' in *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*, ed. Graham Macdonald and Crispin Wright, Oxford, Blackwell, 1986, pp. 284–5.

corresponding desire; or maybe it consists of the desire plus something more. Then in any systematic treatment of belief and desire, we should not expect these beliefs-by-courtesy to function in the same way as beliefs rightly so-called. Or maybe (2) it is a genuine belief, and not *necessarily* connected with desire; but maybe it is contingently connected with desire and we can explain why it is that, quite often, beliefs about what would be good go hand in hand with the corresponding desires. Our Anti-Humean may reply that these explanations of what we say are strained compared with his own. We may reply that we find it hard to see how his could possibly be true. All this skirmishing is inconclusive.

2. *The collision*

Decision Theory is an intuitively convincing and well worked-out formal theory of belief, desire, and what it means to serve our desires according to our beliefs. It is of course idealized, but surely it is fundamentally right. If an Anti-Humean Desire-as-Belief Thesis collides with Decision Theory, it is the Desire-as-Belief Thesis that must go.[2] So now I shall display the collision.

It is fair to take a simple case; because if our Anti-Humean's thesis collides with Decision Theory only in simple cases, that is bad enough. (1) The Desire-as-Belief Thesis only applies to some desires—not including, for instance, your overpowered desire to have Neiss for a colleague. But if the thesis is right, surely it would be possible for some agent—say Frederic, that famous slave of duty—to have only the desires to which the thesis does apply. Let us suppose, for now, that Frederic is moved entirely by beliefs about what would be good; in other words, by desires necessarily connected to such beliefs. (2) Let us suppose, for now, that Frederic does not discriminate degrees of goodness. His desire that A is connected simply to his belief that A would be good—not to beliefs about just how good A would be. (3) Let us suppose that Newcomb-like problems do not arise, so that the 'causal' way of calculating expected value does not differ from the easier 'evidential' way. (4) Let us suppose that Frederic's system of beliefs and desires evolves in accordance with Richard Jeffrey's probability kinematics; and that the Desire-as-Belief Thesis continues to hold—as befits a necessary connection—after any such evolution.

Then the Desire-as-Belief Thesis says that Frederic desires things just when he believes they would be good. Or better, since we must

[2] Decision Theory treats belief and desire as matters of degree. Surely they admit of degree to a considerable extent, but we must of course grant that a thoroughly quantitative treatment is an idealization. The opposite idealization also is of interest, since the truth lies in between. How does the Desire-as-Belief Thesis fare under a thoroughly *non*-quantitative treatment? John Collins has investigated that question. See his article in this issue: 'Belief, Desire, and Revision', *Mind*, 1988, pp. 333–42.

acknowledge that desire and belief admit of degree, he desires things just to the extent that he believes they would be good. To any ordinary proposition A, there corresponds another proposition: Å, the proposition that it would be good that A. Frederic's expected value for A, which represents the degree to which he desires that A, equals the degree to which he believes that Å. And this is so not only for Frederic as he is at present, but also for Frederic after he evolves by probability kinematics. Now, what does Decision Theory say about Frederic's case?[3]

At any moment, Frederic has a *credence* function C. It measures the degree to which he believes various propositions. A conditional credence C(A/E) is defined as a quotient of unconditional credences C(AE)/C(E); and whenever I write a conditional credence, I am imposing a tacit restriction to cases in which the denominator is positive. Credence obeys a principle of additivity: for any proposition A and any partition $E_1, \ldots, E_n$ (a partition being a set of mutually exclusive and jointly exhaustive propositions),

$$(1) \qquad C(A) = \Sigma_i\, C(AE_i) = \Sigma_i\, C(A/E_i)C(E_i).$$

At any moment, Frederic also has an (evidential) expected *value* function V. It measures the degree to which he desires that various propositions be true. Value obeys its own principle of additivity: for any proposition A and any partition $E_1, \ldots, E_n$,

$$(2) \qquad V(A) = \Sigma_i\, V(AE_i)C(E_i/A) = \sum_i \frac{V(AE_i)C(A/E_i)C(E_i)}{C(A)}.$$

Thus the value of a proposition that might come true in several alternative ways is an average of the values of those several alternatives, weighted by their conditional credences.

Now suppose that Frederic's state changes by probability kinematics, starting from an initial state given by the credence and value functions C and V. The change is given by three things. First, we have an *originating partition* $E_1, \ldots, E_n$; these are the propositions subject to exogenous change. Next, we have numbers $d_1, \ldots, d_n$ which measure the *distribution* of change over the members of this partition. Some of the d_i's are positive; these sum to 1; and when d_i is positive, the credence of E_i is raised proportionally to d_i. Other d_i's are negative; these sum to -1; and when d_i is negative, the credence of E_i is lowered proportionally to d_i. Still other d_i's may be zero, in which case the credence of the corresponding E_i's is unchanged. Finally, we have a positive number x which measures the

[3] I follow the exposition of Decision Theory and probability kinematics in Richard C. Jeffrey, *The Logic of Decision*, 2nd edn., London, University of Chicago Press, 1983, except that, unlike Jeffrey, I split the increments that specify an exogenous change into 'distribution' and 'amount'. I speak of the bearers of credence and expected value as 'propositions'; for present purposes, it does not matter that they might be egocentric propositions, or that they might be taken as sentences.

amount of exogenous change. The credence of each E_i changes, up or down as the case may be, by the amount $d_i x$. All other changes in credence are driven by this exogenous change: the credence of any other proposition conditional on any one of the E_i's remains unchanged. So if C_x is the new credence function, we have $C_x(E_i) = C(E_i) + d_i x$ for each E_i; and we have $C_x(A/E_i) = C(A/E_i)$ for any A and E_i. It follows, using additivity for credences, that

$$(3) \quad C_x(A) = C(A)[1+px], \qquad \text{where } p = \sum_i \frac{C(A/E_i)d_i}{C(A)}.$$

Likewise for Å, the proposition that it would be good that A:

$$(4) \quad C_x(\mathring{A}) = C(\mathring{A})[1+qx], \qquad \text{where } q = \sum_i \frac{C(\mathring{A}/E_i)d_i}{C(\mathring{A})}.$$

When credences change by probability kinematics, expected values may change also, but only in response to the exogenous redistribution of credence over the E_i's. If a proposition A is compatible with several different E_i's, then the redistribution may change the conditional credences that A will come true in good or bad ways, and thereby affect the expected value of A. But this cannot happen if A is compatible with only one of the E_i's. In that case, its value remains unchanged. (Call this the *Invariance Assumption.*) In particular, the value of any conjunction AE_i remains unchanged: if V_x is the new value function, $V_x(AE_i) = V(AE_i)$. It follows, using additivity for values, that

$$(5) \quad V_x(A) = V(A)\frac{[1+rx]}{[1+px]}, \qquad \text{where } r = \sum_i \frac{V(AE_i)C(A/E_i)d_i}{V(A)C(A)}.$$

Now take any originating partition and any distribution. Hold them fixed and let x vary. We assume that x can indeed vary, at least within some limited range: the partition and distribution never determine the exact amount of change. The Desire-as-Belief Thesis, applied to Frederic's old and new states, says that

$$(6) \quad C(\mathring{A}) = V(A),$$

$$(7) \quad C_x(\mathring{A}) = V_x(A).$$

From (4), (6), and (7) we have

$$(8) \quad V_x(A) = V(A)\,[1+qx].$$

Now we see the problem: according to (8) the expected value of A goes by a linear function of x, whereas according to (5) it goes by a quotient of linear functions. So the linear function and the quotient must somehow stay equal throughout some range of values of x. How is that possible?

Here is one way: $p = 0$ and $q = 1$. But then the credence of A must be constant. Here is another way: $q = 0$ and $p = 1$. But then the value of A must be constant. These are the only possibilities: because from (5) and (8) we have

$$(9)\quad pqx^2+[p+q-1]x = 0,$$

which cannot hold for more than a single value of x unless both coefficients are zero. So a change by probability kinematics, no matter what the partition and distribution and amount, cannot change both the credence and also the expected value of any proposition A. That is to say that Frederic cannot simultaneously change both his opinion about whether A and his desire about whether A.

This is quite wrong. By imposing the Desire-as-Belief Thesis as a new constraint on Decision Theory, we have *over*constrained it, and made it exclude what can perfectly well happen. Example. Frederic knows that Stanley has often escaped the anger of the pirates by claiming to be an orphan. He now learns that Stanley is in fact no orphan. This discovery has two effects. Frederic reckons that what he can find out, the pirates also will soon find out (perhaps because he will be duty-bound to tell them himself); and so he thinks that the pirates will soon be very angry with Stanley for deceiving them. In addition, he thinks that Stanley will deserve their anger; he believes at least somewhat more than he did before that it would be good for the pirates to be angry with Stanley; and so (in his moralistic way) he desires at least somewhat more than he did before that the pirates be angry with Stanley. Where A is the proposition that the pirates will soon be angry with Stanley, the discovery that Stanley is no orphan brings both a change in the credence of A and also a change in the credence of Å and the expected value of A.

I conclude that our Anti-Humean's Desire-as-Belief Thesis is in bad trouble.

3. *Does the argument prove too much?*

You may think (as I did) that my argument against the Desire-as-Belief Thesis has to be wrong, because it proves too much. For it does not just refute the Anti-Humean's grand Desire-as-Belief Thesis; it refutes also the supposition that some modest, contingent equation of desire with belief might hold in some special case. But consider this special case. Suppose Frederic single-mindedly pursues one goal: that the proposition G be true. (For instance, G might be the proposition that he never ever fails in his duty. The example requires that his goal is perfection—a miss is as good as a mile.) Then, scaling V to the unit interval, we have that for any A, $V(AG) = 1$ and $V(A\bar{G}) = 0$, so

$$(10)\quad V(A) = V(AG)C(G/A)+V(A\bar{G})C(\bar{G}/A) = C(G/A).$$

Likewise for any later state of Frederic given by C_x and V_x,

(11) $V_x(A) = C_x(G/A)$.

For this special case, we have managed at least to equate desire with *conditional* credence. It may seem that we can do better. For any proposition A, let $\mathring{A}$ be the proposition that A conduces to G: $A \rightarrow G$, in some appropriate sense of the conditional arrow. It seems that Frederic's desire that A should always equal his degree of belief that A conduces to achieving his goal G. And not just in his present state given by C and V, but also in any new state given by some C_x and V_x. So we get back the supposedly refuted Anti-Humean equations,

(12) $V(A) = C(A \rightarrow G) = C(\mathring{A})$,

(13) $V_x(A) = C_x(A \rightarrow G) = C_x(\mathring{A})$,

this time not from some grand Desire-as-Belief Thesis but just from plausible-sounding assumptions about the case of the single-minded Frederic. Why not?

Answer: because, *pace* intuition, an 'appropriate sense of the arrow' just does not exist. Taking (10) and (12) together, or (11) and (13), we find ourselves dealing with the dreaded 'probability conditional', a supposed connective which makes probabilities of conditionals equal the corresponding conditional probabilities:

(14) $C(A \rightarrow G) = C(G/A)$,

(15) $C_x(A \rightarrow G) = C_x(G/A)$.

If we had a probability conditional, we could uphold a modest equation of desire with belief, at least in this special case; we have seen why even this modest equation collides with Decision Theory; therefore we do not have a probability conditional. That was known already: certain trivial cases aside, we cannot give a sense to the arrow such that (14) will hold for all C, A, and G. Even if we fix C, and fix an originating partition and distribution, we still cannot give a sense to the arrow such that (15) holds for all x, A, and G.

You might protest that we need nothing so ambitious. For present purposes, the only conditionals that matter are those that have the one fixed consequent G that specifies Frederic's single goal. It is enough if we can fix C, fix the partition and distribution, and also fix G, and then make (14) and (15) hold for all x and A. But what our present argument shows, when applied to the case of the single-minded Frederic, is that we cannot even do that well. In view of earlier negative results against the probability conditional, that should come as no big surprise.[4]

[4] For these negative results, see my 'Probabilities of Conditionals and Conditional Probabilities', *The Philosophical Review*, 1976, pp. 297–315; and 'Probabilities of Conditionals and Conditional

4. *Two simplifications removed*

I said that if our Anti-Humean's thesis collides with Decision Theory only in simple cases, that would be bad enough. Then I went on to simplify by supposing that Frederic was moved entirely by desires to which the thesis applies, and that he did not discriminate degrees of goodness. You might doubt that trouble in this very special case really is bad enough to matter, if it does not arise in less peculiar cases as well. I disagree; but rather than dispute the question, we can do the calulation over with the two simplifications removed.

Suppose, then, that in addition to the component V of expected value that obeys the Desire-as-Belief Thesis, there is also an 'ordinary' component V_o that does not. The total value of a proposition A is $V(A)+V_o(A)$. Assume still that total value obeys additivity, and also obeys the rule that when a change originates in $E_1, \ldots, E_n$ the value of any conjunction AE_i is unchanged. Let us make the same assumptions for V_o: they would hold if V_o were the whole of total value, and why should the behaviour of V_o change just because we add another component to it? Then V, regardless of whether it is the whole or merely a component of total value, also obeys the two assumptions. Therefore (5) still holds.

Now suppose that Frederic thinks of goodness as something that admits of degree. Let $g_1, \ldots, g_m$ be the degrees of goodness that he discriminates. (If you do not want to assume that there are only finitely many, let our sums be infinite and our probabilities infinitesimal, or switch to integrals and probability densities.) Let $\mathring{A}_j$ be the proposition that it would be good to degree g_j that A. Instead of (4) we have for each of the degrees

$$(16)\quad C_x(\mathring{A}_j) = C(\mathring{A}_j)[1+q_j x], \qquad \text{where } q_j = \sum_i \frac{C(\mathring{A}_j/E_i)d_i}{C(\mathring{A}_j)}.$$

The Desire-as-Belief Thesis now takes the form

$$(17)\quad \Sigma_j\, C(\mathring{A}_j)g_j = V(A),$$

$$(18)\quad \Sigma_j\, C_x(\mathring{A}_j)g_j = V_x(A).$$

Probabilities II', *The Philosophical Review*, 1986, pp. 581–9 (with errata noted in the following two issues).

The present argument is very similar to a direct proof that we cannot make (14) hold for fixed C, fixed partition and distribution, and all x, A, and G. That proof, restricted to the case of two-membered originating partitions, appears in the second of the papers just cited.

For the direct proof, we look only at a single conditional—generality over consequents is not much used. But it is used in this way: since we are free to choose G, we may choose it as the disjunction of those members E_i of the originating partition for which d_i is positive. That choice affords a subsidiary argument to eliminate the case that $q = 0$ and $p = 1$. With G fixed in advance, on the other hand, we can assume nothing about how G is related to the originating partition.

Instead of (8), (16)-(18) give us

$$(19) \quad V_x(A) = V(A)[1+qx], \qquad \text{where now } q = \sum_j \frac{C(\mathring{A}_j)q_jg_j}{V(A)},$$

and from (5) and (19) we have (9) as before, except for the redefinition of q just noted. So again there are two alternatives. Either $p = 0$ and $q = 1$, in which case the credence of A must be constant; or else $q = 0$ and $p = 1$, in which case the component of the value of A that obeys the Desire-as-Belief Thesis must be constant. That is to say that Frederic cannot simultaneously change both his opinion about whether A and the component of desire that derives from his opinions about how good it would be that A. Again this is wrong; and for an example of what it wrongly excludes, we need only modify the story of Frederic. Let Frederic now have an ordinary desire that the pirates not be angry with Stanley; and let him discriminate at least two degrees of goodness, one for no anger and one for deserved anger.

5. *Invariance defended*

When he sees how his Desire-as-Belief Thesis collides with standard Decision Theory, our Anti-Humean might well hope to avert the collision by some not-too-radical amendment to Decision Theory. One assumption in particular is the main candidate for discarding. Recall how we derived (5), which specified how expected values change in response to an exogenous redistribution of credence over a partition $E_1, \ldots, E_n$. First we assumed Invariance: no change in the value of any proposition that is compatible with just one of the E_i's, and therefore no change in the value of any conjunction AE_i. Then we used additivity for expected values to calculate the new value of a proposition A.

Let our Anti-Humean propose to discard Invariance. This blocks our derivation of (5), and thereby blocks the collision between the Desire-as-Belief Thesis and Decision Theory. What is more, this is not just blind tinkering. It seems to make sense by the Anti-Humean's lights. There are *two* ways, so he says, that redistribution of credence may change the value of a proposition A. One way is that it may change the conditional credences of the various AE_i's, some of which may be better than others. The other way is that it may change the values of the AE_i's themselves. For each AE_i, he says, we have the proposition that it would be good that AE_i; and why should not the redistribution change the credences of these propositions? Invariance says that only the first kind of change happens, never the second. Why should he accept that?

So far, so good; but discarding Invariance turns out to lead to an unintelligible consequence. It is therefore not an acceptable way to rescue the Desire-as-Belief Thesis.

We note first that it is impossible to discard Invariance only as applied to the AE_i's themselves. For each AE_i may be further partitioned into subcases $AE_iF_1, \ldots, AE_iF_k$, in such a way that each of these subcases is maximally specific in all respects relevant to its value. Since the ratios of probability between the subcases of any single AE_i do not change, the value of AE_i is determined by the values of its subcases. So in order to discard Invariance for AE_i as a whole, we have to discard Invariance also for at least some of the subcases. This means that a change in credence will affect the value of some proposition AE_iF_h that is maximally specific in all respects relevant to its value. How is that possible?

If AE_iF_h were maximally specific merely in all 'factual' respects relevant to its value, and if the Desire-as-Belief Thesis were true, then it would be no surprise if a change in belief changed our minds about how good it would be that AE_iF_h, and thereby affected the value of AE_iF_h. But the subcase was supposed to be maximally specific in *all* relevant respects—and that includes all relevant propositions about what would and would not be good. The subcase has a maximally specific hypothesis about what would be good built right into it. So in assigning it a value, we do not need to consult our opinions about what is good. We just follow the built-in hypothesis.

(Example. How good would it be if, first, pain were the sole good, and second, we were all about to be in excruciating and everlasting pain?—I have to say that this would be good, and so I value the case highly. My opinion that in fact pain is no good does not affect my valuing of the hypothetical case in which, *ex hypothesi*, pain is good. My opinion does cause me to give the case negligible credence, of course, but that is different from affecting the value.)

It is unintelligible how a shift in opinions about what is good could affect the value of any of the maximally specific AE_iF_h's, since these have hypotheses about what's good already built into them. But if not, then the AE_iF_h's should obey Invariance: there is no way left for a change of credence (originating in $E_1, \ldots, E_n$) to affect their value. It follows that the AE_i's also obey Invariance.[5]

Department of Philosophy
Princeton University
Princeton, NJ 08544
USA

DAVID LEWIS

[5] I am indebted to several people for helpful discussion or correspondence; especially Simon Blackburn, John P. Burgess, John Collins, Frank Jackson, D. H. Mellor, Philip Pettit, and Michael Smith.