

Desire as Belief II

DAVID LEWIS

1. Reason and passion

Hume wrote that “we speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them” (*Treatise*, Bk. II, Pt. III, Sect. III). What did he mean?

In the first place, Hume’s “passions” are sometimes none too passionate. He speaks of some passions as “calm”. We would do best to speak of all “passion”, calm and otherwise, as “desire”.

In the second place, we call someone “reasonable” in part because his desires are moderate and fair-minded. But when we do, I suppose we speak not strictly and philosophically. Strictly speaking, I take it that reason is the faculty in charge of regulating belief. And so I read Hume as if he had said that belief is the slave of desire. Our actions do, or they ought to, serve our desires according to our beliefs. More precisely, taking account of the fact that both belief and desire admit of degree, and not begrudging the usual idealizations that make the topic tractable, our actions serve our subjective expected values according to our subjective degrees of belief. For short: they serve our values according to our credences.

Values and credences belong to propositions: classes of maximally specific possibilities (perhaps egocentric and tensed). The value of a proposition A , written $V(A)$, is a real number; the credence of A , $C(A)$, is a non-negative real number; and the credence of the necessary proposition, $C(I)$, is 1. (We would do well to let these values and credences be nonstandard real numbers so that, for instance, the propositions corresponding to the maximally specific possibilities—call these point-propositions—might all get infinitesimal credence. For there well might be infinitely many point-propositions.) We assume the usual rules of finite (or *-finite) additivity for value and credence: when A_1, \dots are a partition of A ,

$$\begin{aligned}V(A) &= \sum_i V(A_i)C(A_i/A) \\C(A) &= \sum_i C(A_i)\end{aligned}$$

where $C(X/Y)$ abbreviates the quotient $C(XY)/C(Y)$. The additivity rule for value shows how belief serves desire: it generates an expected value for the less specific proposition A out of the values for the more specific cases A_1, \dots . Nobody doubts that belief and desire are entangled to this extent, whatever further entanglements there may or may not be.

(A famous difficulty need not concern us here. Suppose a certain action would serve as an effective means to your ends. Yet at the same time it would constitute evidence—evidence available to you in no other way—that you are predestined inescapably to some dreadful misfortune. Should you perform that action?—Yes; your destiny is not a consideration, since that is outside your control. Do you desire to perform it?—No; you want good news, not bad. Since our topic here is not choiceworthiness but desire, and since the two diverge, we adopt an “evidential” conception of expected value, on which the value of the useful action that brings bad news is low. Choiceworthiness is governed by a different, “causal”, conception of expected value.¹)

As an empiricist, Hume thinks that passions are where you find them. Desires are contingent. It is not contrary to reason—still less is it downright impossible!—to have peculiar and unusual desires, or to lack commonplace ones. It may be contrary to the laws of human nature, but those laws themselves are contingent regularities. Likewise there are no necessary connections between desire and belief. Any values can go with any credences. Or at any rate—remembering the entanglement of credences in the rule of additivity for values—any values for the point-propositions can go with any credences.

Neither is the rule of additivity for credences unHumean, even though it connects credences necessarily with other credences. By way of professing innocence, we could say roughly this: the credences of point-propositions, at any rate, are not necessarily connected. Any point-credences can go with any pattern of other point-credences. As for other propositions, their credences are mere patterns—namely, sums—of point-credences; and the necessary connection between a pattern and its elements is surely not a necessary connection between distinct existences.

(This is still not quite right. We chose to scale the credences in such a way that all the point-credences sum to 1. Likewise values are somehow scaled, though I had no need to say how. Either we must tolerate the necessities that arise from arbitrary choices of scale, or we must represent credences and values in a way that somehow abstracts from arbitrary choices of scaling. This need not concern us further.)

Thus Humeanism takes point-values and point-credences to be “loose and separate”, unconstrained by necessary connections. If there are uni-

¹ See, *inter alia*, Lewis (1981).

versally shared desires, that is a contingent matter. If there are universal correlations between certain beliefs and desires, that too is a contingent matter. Someone might have no desire at all for joy, knowledge, or love. Someone might believe just what you and I believe, and still have no desire at all for joy, knowledge, or love. Indeed, someone might believe just what G. E. Moore believed about the simple, non-natural properties of these things and still have no desire for them.

2. How Humeanism might be false

You might fear that anti-Humean moral science would have to rest on anti-Humean metaphysics of modality. The necessity whereby we cannot lack certain desires, or whereby our desires cannot fail to be suitably aligned with our beliefs, would then be necessity *de re*. It would be like the necessity that theorists of “strong laws” discern in the laws of nature, or in the alignment between the laws of nature and certain remarkable relations of universals.²

But there is an alternative. The necessity that supposedly governs desire might be a merely verbal, or conceptual, necessity. So those of us who follow Hume unswervingly in rejecting *de re* necessary connections in nature—“strong laws” or whatnot—still can afford to be open-minded about anti-Humean moral science.

It is a familiar idea that theoretical terms introduced in scientific theories denote the occupants of roles set forth in the term-introducing theory. Mass is that which occupies—perfectly, or near enough—the mass-role set forth in classical mechanics. Phlogiston would have been that which occupied the phlogiston-role set forth in obsolete chemistry. Now that we think there is nothing that does what phlogiston was said to do, or even comes close, we conclude that there is no such thing as phlogiston.

It is also a familiar idea that tacitly known folk theories may introduce terms in much the same way that scientific theories do; and, in particular, that our ordinary mental vocabulary consists of the theoretical terms of commonsensical “folk psychology”.³ Belief and desire, among others, are the states that occupy certain folk-psychological roles. And again, when it comes to occupying a role and thereby deserving a name, near enough is good enough. Folk psychology needn’t be flawless!

² See Armstrong (1983) for a defence of one such theory and discussion of others.

³ See, *inter alia*, Lewis (1972).

A less familiar, but promising, idea is that the “theory theory” applies also to our ethical vocabulary (Railton 1992; Jackson and Pettit 1995). Schisms within folk ethics are of course an obstacle. But perhaps there is more common ground than meets the eye of us professional controversialists. Or perhaps there is a trajectory toward greater consensus, and we can take the term-introducing theory to be the not-yet-seen (and perhaps never-to-be-seen) limit of that trajectory. Or if all else fails we can go relativist: there are as many disambiguations of our ethical terms as there are irreconcilably conflicting versions of folk ethics.

Now suppose that folk moral science is an inseparable mish-mash of psychology and ethics. Its theoretical vocabulary is in part psychological, in part ethical. Its tacitly known postulates include some that say what is universally desired, or that say how our desires are aligned with our beliefs. Both psychological and ethical vocabulary appear in these postulates. Further, these postulates specify an important part of the theoretical roles that define theoretical terms. Conforming to them plays a big part in determining whether states occupy the roles, and deserve the names, of belief and desire.

The upshot might be that if someone disdained joy, knowledge, and love; or if he did so despite believing just what you and I believe; or if he did so despite believing just what Moore believed about the simple, non-natural properties; or if ... ; then his states would not after all occupy the roles and deserve the names of belief and desire (and disdain). The description of the case is subtly contradictory. That is how Humean moral science might be false, and how some anti-Humean theory of Desire by Necessity or Desire as Belief might be true; and without benefit of any *de re* necessity in nature.

I understand the hypothesis that Humeanism might be false in the way just explained. But I do not believe it. For when I consider stories in which supposedly necessary desires go missing, or in which supposedly necessary alignments of desire with belief go haywire, I find I am not at all inclined to doubt that the so-called “beliefs” and “desires” in the story are rightly so called.

(It may be otherwise with still weirder psychological fantasies. When Anscombe tells of the man who desires a saucer of mud, though he has no idea what would be good about having it, the story does seem not altogether intelligible; likewise when Goodman tells the story of the man who expects the future to resemble the past only in respect of gruesome disjunctive properties (Anscombe 1958; Goodman 1955). But what sort of unintelligibility am I detecting? Is it semantic anomaly, the incorrect applying of names to things that could not deserve those names, as when ideas are said to be green? Or is it rather the frustration of my best efforts

at empathetic understanding? I do not know. I do not know how to find out.)

So I am doubtful about all versions of anti-Humeanism. But my doubts rest on intuitions that might be easy to controvert. And besides, these theories offer a rich reward: objective ethics. If there are some things we desire by necessity, we surely would want to say that these things were objectively desirable. Or if there were some propositions, belief or disbelief in which was necessarily connected with desire, some of them presumably would be true; then we surely would want to say that the true ones were the objective truth about ethical reality.

Why care about objective value or ethical reality? The sanction is that if you do not, your inner states will fail to deserve folk-theoretical names. Not a threat that will strike terror into the hearts of the wicked! But who ever thought that philosophy could replace the hangman?

3. Desire by necessity

We can go no further talking about anti-Humeanism in general. It is time to examine various versions. A systematic survey of all possible versions, including versions not yet invented, would be nice. But we shall have to settle for less.

Desire by Necessity is a comparatively simple and unproblematic version. In its simplest form, it says that necessarily and regardless of one's credence distribution, certain point-values must be high and the rest low. Scale these as 1 and 0. Let G be the union of point-propositions with necessarily high value: the objectively desirable point-propositions—for short, the *good* ones. Then for any proposition A and any credence distribution C (provided that $C(A)$ is positive, a restriction we shall henceforth leave tacit),

$$V(A) = C(G/A).$$

Refinements are obvious, but we need not consider them in detail. (1) We could have more than just two degrees of objective value for point-propositions. (2) We could distinguish different components of the value of a point-proposition, pertaining to different objective values. (3) We could allow contingent, Humean desires alongside the necessary, unHumean ones—a half-Humean mixed theory.

4. *Desire as belief revisited*

We turn next to versions of anti-Humeanism on which desires are said to be contingent, but necessarily aligned with suitable beliefs.⁴ These form a more varied family of theories. One of them, at least, is definitely untenable. Others are not what they seem to be. Maybe some tenable version of anti-Humeanism falls unproblematically into this class. But if so, I do not know what it is.

In the paper to which this one is a sequel, I examined and refuted one especially simple theory in this family (Lewis 1988). (Would-be anti-Humeans hastened to inform me that the refuted theory was but one possible version of anti-Humeanism—something I myself had said at the outset!) I shall call this simple theory “Desire as Belief”—for short, DAB—without any qualifying adjective. DAB says that there is a certain function (call it the “*halo*” function) that assigns to any proposition A a proposition A° (“ A -halo”) such that, necessarily, for any credence distribution C ,

$$(DAB) \quad V(A) = C(A^\circ).$$

We might want to say that A° is the proposition that A is, or would be, objectively desirable—that is, good. Necessarily, and regardless of one’s credence distribution, one must desire A exactly to the extent that one believes it to be good.

This version of anti-Humeanism is untenable. Except in trivial cases, it collapses into contradiction. Credences and expected values respond differently to redistribution of credence with point-values held constant. Suppose the DAB equation holds under a given credence distribution; it will cease to hold under almost all redistributions of credence.

The refutation by redistribution of credence given in my previous paper was needlessly complicated.⁵ To make it simple, and at the same time to make it obvious where the blame falls for the collapse, we note that DAB can be equivalently restated as a pair of equations: necessarily, for any A and C ,

$$(DACB) \quad V(A) = C(A^\circ/A)$$

$$(IND) \quad C(A^\circ/A) = C(A^\circ).$$

To derive DACB, we recall that DAB is supposed to continue to hold under redistributions of credence, and we redistribute by conditionalizing

⁴ Instead of speaking as I do of desires necessarily connected to beliefs, you might prefer to speak of beliefs that function as if they were desires; or of states that occupy a double role, being at once beliefs and desires. I take these descriptions to be equivalent.

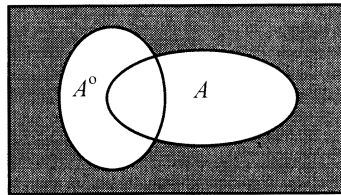
⁵ See Arló Costa, Collins, and Levi (1995) for a refutation simpler than my previous one, but somewhat different from the one given below.

on A .⁶ (That is, we put all the credence on point-propositions within A , but we do not alter point-values or ratios of point-credences within A .) IND follows immediately from DAB and DACB. Conversely, DAB follows from DACB and IND.

Whereas DAB equated values to unconditional credences, DACB equates them to conditional credences. But according to IND this difference does not matter, because the unconditional and conditional credences are always equal. A° and A are probabilistically independent with respect to C , and they remain independent under any redistribution of credence (provided that the credence of A remains positive so that the conditional credence does not go undefined).

Now it is IND, unabetted by DACB, that leads to contradiction. Take any A and C such that $C(A)$ and $C(A^\circ/A)$ are positive, and such that $C(A)$ and $C(A^\circ/A)$ are less than 1. If there are no such A and C , the case is trivial. (We shall take a closer look at the trivial cases in §5 below.)

It follows from IND and our stipulations on A and C that all four of the propositions $(A \wedge A^\circ)$, $(A \wedge \neg A^\circ)$, $(\neg A \wedge A^\circ)$, and $(\neg A \wedge \neg A^\circ)$ have positive credence. Then there are various redistributions of credence, by conditionalizing and otherwise, that will make IND go from true to false. (For instance, if we redistribute credence from the shaded into the unshaded region while leaving ratios of point-credences in the unshaded region unchanged—at the extreme, if we conditionalize on $(A \vee A^\circ)$ —then $C(A^\circ)$ increases while $C(A^\circ/A)$ stays the same.) This contradicts the claim that IND is preserved under redistributions. DAB can hold only in trivial cases. This completes our refutation.



⁶ I assume here that *one* way to revise credences is by conditionalizing, and that DAB will continue to hold after any such revision; I do not assume that credences may never be revised in any other way. Nor, *pace* Graham Oddie, was it “a fundamental assumption” (1994, p. 466) of my previous refutation that revisions of credence must invariably go by conditionalizing; or even that they must invariably go by the sort of generalized conditionalizing that Richard Jeffrey has described under the name of “probability kinematics”. Maybe Oddie is right that there are other ways for credences to be revised, at least when they are the credences of tensed propositions. (Before I turned out the light, I saw that it was just minutes before midnight. In the course of a long and sleepless night, I undergo a redistribution of credence from the proposition that it is now before midnight to

5. Desire as conditional belief

It was IND that did the dirty work. DACB had no hand in it. So the obvious line of retreat, after the downfall of DAB, is to keep DACB and junk IND. Exactly this theory of Desire as Conditional Belief has been defended by Huw Price (1989). But not in quite the way that I have presented it: Price presents DACB not as a consequence or fragment of DAB, but as a superior rival to it.

Superior in two ways. In the first place, DACB is immune to refutation by redistribution of credence. With my previous complicated refutation, this took some proving; as for the present simpler refutation, we need only recall that the contradiction was derived not from DACB but from IND.

In the second place, Price argues that “whenever it makes a difference, we should assess a possible outcome under the hypothesis that it is the actual outcome” (1989, p. 122). Well, not always. Not, for instance, when thinking how pleased we should be that a certain undesired outcome has turned out not to be actual. So I question Price’s second reason for preferring DACB to DAB. No matter—his first reason is reason enough.

To understand DACB better, we must learn what it is trying to tell us about the “halo” function: the mapping from A to A° .

Initial Lemma. Whenever $C(AB)$ is positive, $C(A^\circ/AB) = V(AB) = C(B^\circ/AB)$. *Proof.* DACB continues to hold under redistributions of credence, and in particular under redistributions by conditionalizing on A or on B . Conditionalizing on B turns $C(A^\circ/A)$ into $C(A^\circ/AB)$ and $V(A)$ into $V(AB)$. Conditionalizing on A turns $V(B)$ into $V(AB)$ and $C(B^\circ/B)$ into $C(B^\circ/AB)$. So both our new equations come from instances of DACB by conditionalizing.

Upward Lemma. When A is nonempty, and I is the necessary proposition, $(A^\circ \wedge A) = (I^\circ \wedge A)$. *Proof.* If not, we could distribute credence in such a way as to make $C(A)$ positive, and also make $C(A^\circ \wedge A)$ and $C(I^\circ \wedge A)$ unequal. That would make $C(A^\circ/AI)$ and $C(I^\circ/AI)$ unequal, thereby falsifying an instance of the Initial Lemma.

Downward Lemma. When W is a point-proposition, $V(W) = 1$ if W is included in I° , $V(W) = 0$ otherwise. *Proof.* Let C be any distribution that gives W positive credence. Taking A as W and B as I , and dropping I whenever it appears as a conjunct, the right hand equation of our Initial Lemma gives us that $V(W) = C(I^\circ/W)$. If W

the proposition that it is now after midnight. It is far from obvious that this revision goes by probability kinematics, let alone by conditionalizing.) But that fact, if fact it be, does nothing to rescue DAB from either my present or my previous refutation.

is included in I° , $C(I/W) = 1$. If not, then since W is a point proposition, W is included in $\neg I^\circ$; in which case $C(I/W) = 0$.

Desire as Conditional Belief is now unmasked. It is not, despite superficial appearances, a theory of contingent desire necessarily aligned with belief. Rather, it is the very same theory of Desire by Necessity that we have already examined—except that the union of necessarily desired point-propositions, formerly called G , is now renamed I° . Point-propositions have value 1 if they fall within I° , value 0 otherwise. So for any A and C , if $C(A)$ is positive, we have $V(A) = C(I^\circ/A)$.

(When A is a proposition other than I , we did not settle, cannot settle, and need not settle exactly what proposition A° is. All that matters to the value of A is the part of A° that lies within A ; and within A , A° and I° coincide. A° might contain all, some, or none of the point-propositions that lie outside A . On that question, DACB plays it safe by giving us no information—unlike DAB, which gave us more information than consistency would allow!)

(We noted that if there are no A and C such that $C(A)$ and $C(A^\circ/A)$ are positive, and such that $C(A)$ and $C(A^\circ/A)$ are less than 1, then the case is trivial. We can now characterize the trivial cases. They are those in which, no matter how we choose A and C , we cannot give positive credence to all three of $\neg A$, $(A \wedge A^\circ)$, and $(A \wedge \neg A^\circ)$. There are three ways that could happen. (1) The space might be too small: we might not have three different point-propositions, but only one or two. (2) I° might be empty, giving all propositions a value of 0. (3) I° might be I , giving all propositions a value of 1.)

6. *Desire as belief restricted*

We could keep the original DAB equation, but allow it to apply only to point-propositions: for all C and W

$$\text{(DABR)} \quad \text{If } W \text{ is a point-proposition and } C(W) \text{ is positive,} \\ V(W) = C(W^\circ)$$

Conditionalizing on W we have

$$\text{(DACBR)} \quad V(W) = C(W^\circ/W) = \begin{cases} 1 & \text{if } W \text{ is included in } W^\circ \\ 0 & \text{otherwise} \end{cases}$$

From DABR and DACBR we have

$$\text{(INDR)} \quad C(W^\circ) = C(W^\circ/W) = \begin{cases} 1 & \text{if } W \text{ is included in } W^\circ \\ 0 & \text{otherwise} \end{cases}$$

INDR, unlike IND, does not collapse into contradiction. Instead, since it holds for all credence distributions, we have

$$W^\circ = \begin{cases} I & \text{if } V(W) = 1 \\ \emptyset & \text{if } V(W) = 0. \end{cases}$$

So we're back once more to a disguised version of Desire by Necessity, with a new way to characterize the values of point-propositions in terms of the restricted halo function.

7. *Inconstancy*

Hitherto we have hoped for one fixed halo function that would continue to satisfy the conditions we imposed on it under all redistributions of credence. We have not supposed that each agent might have his own personal halo function in the same way that he has his own personal credence and value functions. But if instead we only require that for any given credence and value functions C and V , there exists a halo function chosen ad hoc to satisfy the desired conditions with respect to that particular pair of C and V , then our task is almost trivial.⁷ We need only require that C and V have the right ranges of values: for any A there exists some B such that $V(A) = C(B)$. (This means that values must be bounded above and below, and must be suitably scaled.) Then, appealing to the Axiom of Choice, there is indeed a halo function such that for all A , $V(A) = C(A^\circ)$.

It's too easy, and no anti-Humean should celebrate such an easy victory. The DAB equation holds not in virtue of any interesting relationship between a desired and a believed proposition, but only in virtue of what it takes for a proposition to deserve the name " A° "—and what it takes is nothing more or less than the right credence, one that equals $V(A)$. There is nothing at all anti-Humean about this little trick. Further, there is nothing that should make us want to say that A° is the proposition that A is objectively good.

⁷ John Collins has noted that if we let the inconstant halo function depend only on V and not also on C , our task may not be so trivial. I have no results to offer about this version of Inconstancy; except only that even if it succeeded it would not deliver objective ethics.

8. Conclusion

We have examined four theories that sought to implement the anti-Humean idea that desires and beliefs are necessarily aligned. One collapses into either triviality or contradiction. Two more collapse into Desire by Necessity—a form of anti-Humeanism, sure enough, but not the right form of anti-Humeanism. Another is not really anti-Humean at all. We could keep trying, but the prospects of success have begun to look dim.

Department of Philosophy
Princeton University
Princeton, NJ 08544
USA

DAVID LEWIS

REFERENCES

- Anscombe, G.E.M. 1958: *Intention*. Oxford: Blackwell.
- Arló Costa, H., Collins, J. and Levi, I. 1995: "Desire-as-Belief Implies Opinionation or Indifference". *Analysis* 55, pp. 2–5.
- Armstrong, D.M. 1983: *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Goodman, N. 1955: "The New Riddle of Induction", in Goodman: *Fact, Fiction, and Forecast*. Cambridge MA: Harvard University Press, 1955.
- Hume, D. 1739/40: *A Treatise on Human Nature*. L. A. Selby-Bigge (ed.), Oxford: Oxford University Press (1888).
- Jackson, F. and Pettit, P. 1995: "Moral Functionalism and Moral Motivation". *The Philosophical Quarterly* 45, pp. 451–72.
- Lewis, D. 1972: "Psychophysical and Theoretical Identifications". *Australasian Journal of Philosophy* 50, pp. 249–58.
- 1981: "Causal Decision Theory". *Australasian Journal of Philosophy* 59, pp. 5–30.
- 1988: "Desire as Belief". *Mind* 97, pp. 323–32.
- Oddie, G. 1994: "Harmony, Purity, Truth". *Mind* 103, pp. 451–72.
- Price, H. 1989: "Defending Desire-as-Belief". *Mind* 98, pp. 119–27.
- Railton, P. 1992: "Nonfactualism About Normative Discourse". *Philosophy and Phenomenological Research* 53, pp. 961–8.