

Against Lewis on ‘Desire as Belief’

Douglas Ian Campbell
The University of Canterbury

Abstract. David Lewis describes, then attempts to refute, a simple anti-Humean theory of desire he calls ‘Desire as Belief’. Lewis’ critics generally accept that his argument is sound and focus instead on trying to show that its implications are less severe than appearances suggest. In this paper I argue that Lewis’ argument is unsound. I show that it rests on an essential assumption that can be straightforwardly proven false using ideas and principles to which Lewis is himself committed.

1. Introduction

David Lewis (Lewis, 1988, 1996) describes, then attempts to refute, an anti-Humean theory of desire he calls ‘Desire as Belief’ (DAB). I will call Lewis’ argument against DAB the ‘updating argument’. The implications of the updating argument are often held to extend far beyond disputes over the nature of desire. For example, Graham Oddie says it constitutes ‘a massive problem for realism about value’ (Oddie, 1994, p. 453), while Ruth Weintraub says it ‘entails subjectivism about ethics’ (Weintraub, 2007, p. 119).

In this paper I show that the updating argument rests on an essential assumption that can be straightforwardly proven false using ideas and principles to which Lewis is explicitly committed and which are central to the updating argument itself. My position, in other words, is that in constructing the updating argument Lewis implicitly contradicts himself.

The error I identify in the updating argument appears to have gone unnoticed by the updating argument’s many other critics. These critics have, with a few exceptions, accepted the updating argument as sound, and focused instead on trying to show that its soundness can be tolerated. A common tack has been to contend that DAB is not in any case a plausible or attractive (or even, perhaps, coherent) version of anti-Humeanism, and then argue (or at least suggest) that certain more plausible versions of anti-Humeanism are invulnerable to the updating argument (Broome, 1991; Byrne & Hájek, 1997; Daskal, 2010; Hájek & Pettit, 2004; Price, 1989; Weintraub, 2007). This approach has the weakness that, while it might provide comfort to the anti-Humean, it does nothing to dispel the threats posed by the updating argument to value-realism and objectivism in ethics. The present paper disposes of all such threats.

§2 describes DAB. §3 outlines the updating argument. §4 exposes the error in the updating argument. §5 deals with possible Lewisian rejoinders. §6 wraps things up.

2. DAB

Following Lewis (Lewis, 1988, 1996), let ' A° ' (pronounced ' A halo') denote the claim *that A 's being true is good*. Harmlessly simplifying (although c.f. (Oddie, 2001)), Lewis ignores *degrees* of goodness, and in the interests of generality he doesn't specify what A 's being good consists in (instead leaving one free to plug in whichever analysis of goodness one wants). Let $V(A)$ denote the degree to which a rational agent values A being the case. Let $C(p)$ denote a rational agent's credence for p 's being true. DAB is the following claim:

$$\text{DAB:} \quad V(A) = C(A^\circ)$$

That is, DAB is simply the idea that the value assigned by a rational agent to some proposition A , and thus the extent to which she desires that A , is identical to her credence for A being good.¹ DAB is 'anti-Humean' in the sense that it entails a rational agent's desires logically supervene on her beliefs about what is good: i.e., that her desires and beliefs are not modally separable 'distinct existences'.

In the above formula, A is intended by Lewis to be a universally quantified variable ranging over all propositions. Following Lewis, I leave this implicit, omitting the quantifier. For simplicity I continue to suppress quantifiers until near the end of the paper.

3. Lewis' updating argument

The updating argument (Lewis, 1988, 1996) involves the following two claims:

$$\begin{array}{ll} \text{INV:} & V(A|A) = V(A) \\ \text{X:} & V(A|A) = C(A^\circ|A) \end{array}$$

INV says, in effect, that the value a rational agent assigns to a proposition, A , being true will not be affected by her learning that A is indeed true. Lewis provides a detailed argument for INV (Lewis, 1988, pp. 331-332). Recently, several authors have argued that INV is false (Bradley & List,

¹ Although DAB is intended as a theory of *desire*, Lewis casts it in terms of the V operator, and thus in terms of what the agent *values*. His tacit assumption is that an agent desires a given state of affairs just to the degree she values it. If this assumption appears dubious then it can simply be stipulated that $V(A)$ denotes strength of desire, rather than level of valuing.

2009; Bradley & Stefánsson, 2016; Stefánsson, 2014). But here I will assume that INV is true, and show that there is a problem elsewhere in Lewis' argument.

X is clearly a claim that a proponent of DAB must accept: for if a rational agent has learnt that A obtains, then her credence for A being good, and thus – according to DAB – the degree to which she values A , will be updated by conditionalization on A . This is what X says.

DAB, INV and X together entail IND:

$$\text{IND: } C(A \uparrow A) = C(A^\circ)$$

Thus, a proponent of DAB must accept IND. However, suppose an agent's credence function is as depicted in the Venn diagram of Figure 1 (Lewis, 1996, p. 309).

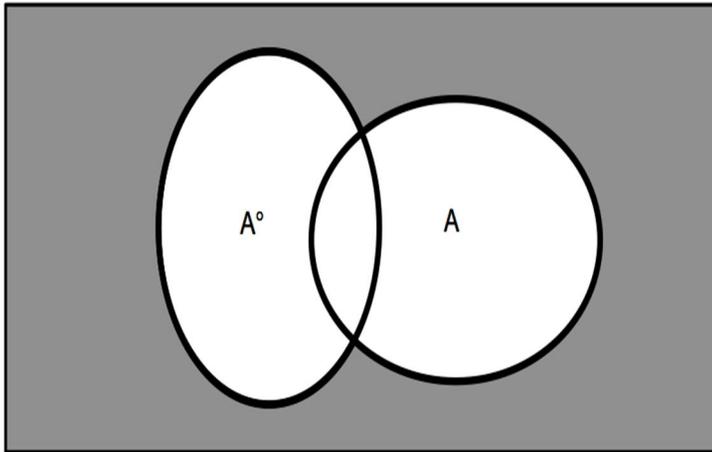


Figure 1. An agent's credence function for A , A° , and their negations.

Suppose that the agent learns (say) that $(A \vee A^\circ)$ is the case, and updates her credences accordingly. Let's call her updated credence function, C' . By the rules of Bayesian updating, C' is obtained by taking the credence represented by the shaded region of Figure 1 and redistributing it to the remaining regions so as to leave the relative sizes of these remaining regions unchanged. When this is done the A° -region grows in size. Thus $C'(A^\circ) > C(A^\circ)$. On the other hand, the proportion of the A -region which overlaps with the A° -region doesn't change. Thus $C'(A \uparrow A) = C(A \uparrow A)$. When these two results are combined with the assumption that C satisfies IND, we get:

$$C'(A \uparrow A) = C(A \uparrow A) = C(A^\circ) < C'(A^\circ),$$

and thus,

$$C'(A \circlearrowleft A) < C'(A \circlearrowright).$$

Notice that this means the agent's updated credence function, C' , *does not conform to IND*. And so even if an agent starts off using a credence function which satisfies IND, this property – of satisfying IND – will be unstable under Bayesian updating. It will, for instance, be lost as soon as the agent learns that $(A \vee A \circlearrowright)$ is the case and updates her credences accordingly. A proponent of DAB must – Lewis says – therefore deny that credences are to be adjusted by Bayesian updating. However Lewis insists, reasonably enough, that given the choice of repudiating DAB or repudiating Bayesian updating, we should repudiate the former, not the latter (Lewis, 1988, p. 325).

4. The error in Lewis' reasoning

My refutation of Lewis' updating argument has two steps. First, I prove that a result I call the 'conditionalization conjecture' (CC) must be true if DAB is true. Next I show that if CC is true, then Lewis' updating argument is unsound.

Suppose that $A \wedge A \circlearrowright$ is true – which is to say that A is true, and that A is good. Then it follows, obviously, *that a good state of affairs obtains*. Let this proposition, that a good state of affairs obtains, be called H (for 'halo'). This being so, $H \circlearrowright$ says it would be good that a good state of affairs obtains, which is a trivial truth. On the other hand, $(\neg H) \circlearrowright$ says it would be good that a good state of affairs does not obtain, which is a trivial falsehood. A rational agent will therefore assign a credence of 1 to $H \circlearrowright$, and a credence of 0 to $(\neg H) \circlearrowright$. These results are recorded as follows:

$$C(H \circlearrowright) = 1 \quad (1)$$

$$C((\neg H) \circlearrowright) = 0 \quad (2)$$

If we assume DAB, then from (1) and (2) we can derive (3) and (4):

$$V(H) = 1 \quad (3)$$

$$V(\neg H) = 0 \quad (4)$$

According to Lewis (Lewis, 1988, p. 326, 1996, p. 303), the values assigned to outcomes by a rational agent will obey a principle of additivity, which he characterises as follows: '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' (Lewis, 1988, p. 326). In symbols:

$$V(X) = \sum_i (V(X \wedge E_i) C(E_i | X)) \quad (5)$$

Here X denotes any given proposition, and $\{E_1, \dots\}$ denotes any given set of mutually exclusive and jointly exhaustive propositions (i.e., any 'partition' of possibility space).

Now, let T represent some obvious tautology (say, ' $1=1$ '). Consider the partition $\{A, \neg A\}$, where A is some arbitrary proposition. Substituting T for X and $\{A, \neg A\}$ for $\{E_1, \dots\}$ within (5) gives (6):

$$V(T) = V(T \wedge A)C(A|T) + V(T \wedge \neg A)C(\neg A|T) \quad (6)$$

When (6) is simplified in light of T being a tautology (and hence a necessary truth), it turns into (7):

$$V(T) = V(A)C(A) + V(\neg A)C(\neg A) \quad (7)$$

Now, let's suppose the arbitrary selected proposition, A , that is mentioned in (7) is in fact identical to H . (7) then becomes (8):

$$V(T) = V(H)C(H) + V(\neg H)C(\neg H) \quad (8)$$

Substituting (3) and (4) into (8) yields (9):

$$V(T) = C(H) \quad (9)$$

Substituting (9) back into (7) gives us (10):

$$C(H) = V(A)C(A) + V(\neg A)C(\neg A) \quad (9)$$

Conditionalizing on A (and thus setting $C(A) = 1$ and $C(\neg A) = 0$) we obtain (10):

$$C(H|A) = V(A) \quad (10)$$

Assuming again that DAB is true, we can derive CC from (10):

$$\text{CC: } C(A^\circ) = C(H|A)$$

CC says that, for any proposition A , the credence that an agent assigns to A° is identical to the *conditional credence* she assigns to *there being a good outcome given that A comes true*. Notice that CC implies that $C(A^\circ)$ is not an *unconditional* credence (as surface appearances suggest), but that it is instead a *disguised conditional credence* – namely, the conditional credence of H given A .

So much for why, if DAB is true, CC must be true. I now explain why, if CC is true, then Lewis's updating argument against DAB rests on an illegitimate assumption.

As already explained, the updating argument rests on the assumption that credences are distributed as in Figure 1, with one 'dollop' of credence being assigned to A , another, overlapping 'dollop' of credence being assigned to A° , and the conditional credence assigned to $(A^\circ|A)$ being represented by the proportion of the A -region which overlaps with the A°

region. If credences are distributed in this way, then Lewis is quite right: Bayesian updating can in this case potentially change how much credence is assigned to A° without causing any corresponding change to how much conditional credence is assigned to $(A^\circ|A)$ (or *vice versa*) – with the result that IND will not be robustly satisfied.

However, if CC is true, then Figure 1 misrepresents the logical relationship between A and A° . Whereas Figure 1 represents A and A° as being logically independent propositions, CC implies that they are not independent, and that A° is instead a conditional proposition, conditioned on A . If CC is true, then a rational agent's credences must be distributed not as they are in Figure 1, but rather as they are in Figure 2:

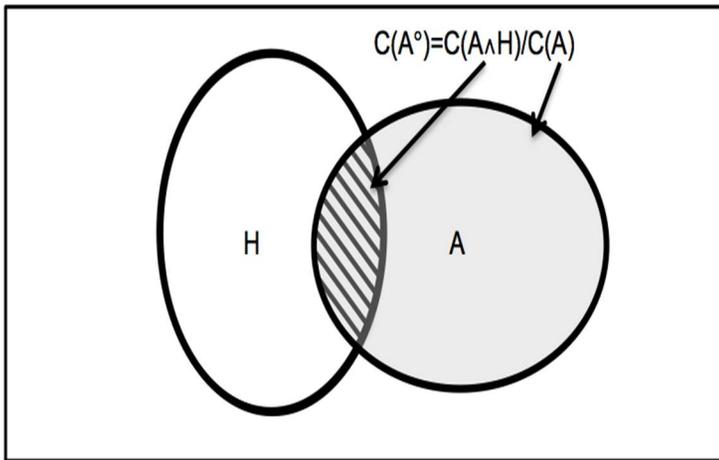


Figure 2. How credences should be assigned by a rational agent.

In Figure 2 one ‘dollop’ of credence is assigned to H , and another to A . The amount of credence assigned to A° , $C(A^\circ)$, is identified with $C(H|A)$, and thus with the proportion of A -region that overlaps with the H -region – or, that is, with the proportion of the shaded region that is crosshatched.

Now, notice two points. First, under this way of allocating credences, $C(A^\circ)$ is a conditional credence, and it is conditioned on A (for it is the conditional credence of H given A). Second, it is a universal rule that if a credence function is already conditioned on a certain outcome, *then updating it by conditionalizing it on the same outcome again has no effect*. (That is, for any two credence functions P and Q , if $P(x) = Q(x|a)$ – i.e., if P is already conditioned on a – then $P(x|a) = P(x)$ – i.e., conditionalizing P on a again changes nothing.) Putting these two points together, it follows that $C(A^\circ|A) = C(A^\circ)$ (for, since A° is already conditioned on A , conditioning on A again has no effect). Thus it follows that $C(A^\circ|A)$ will be represented in

Figure 2 in exactly the same way that $C(A^\circ)$ is represented – namely, by the proportion of the shaded region that is crosshatched. Since $C(A^\circ)$ and $C(A^\circ|A)$ are, for this reason, necessarily identical, it is impossible for Bayesian updating to produce any mismatch between them. That is, not only are $C(A^\circ)$ and $C(A^\circ|A)$ both represented in Figure 2 by the proportion of the shaded region that is crosshatched, *but they will both continue to be represented by this single element of the diagram even as the various regions in Figure 2 grow and shrink when forces of Bayesian updating are operating upon them.* Since $C(A^\circ)$ and $C(A^\circ|A)$ must, for this reason, always be identical, IND will always be satisfied.

Let's take stock. Lewis' updating argument against DAB allegedly proves that no credence function can robustly satisfy IND in the face of Bayesian updating. However, the updating argument assumes that credences are distributed as per Figure 1. I have shown that if DAB is true, then CC is true, and that if CC is true, then credences must instead be distributed as per Figure 2. Under this way of distributing credences, IND is necessarily satisfied. And so, in arguing against DAB, Lewis begs the question against DAB, by making an assumption that is inconsistent with the truth of DAB. The updating argument is unsound for this reason.

5. How might Lewis respond?

CC tells us that for any credence function, C , and proposition, A , $C(A^\circ) = C(H|A)$. The implication is that the halo function, $^\circ$, is a propositional function that accepts a proposition, p , as input, and produces another proposition, p° , as output, such that a rational agent's credence for p° should always match her conditional credence for H , given p . But is a propositional function that meets these specifications even logically possible in the first place? We need look no further than some of David Lewis's own earlier work in order to find an argument that purports to show that such a function is *not* logically possible.² In his (Lewis, 1976, 1986) Lewis aims to demonstrate that there can be no systematic way of mapping a pair of propositions, q and r , onto a proposition, s , such that $C(s)$ should match $C(q|r)$. Lewis might respond to my demonstration that DAB entails CC by harking back to this earlier work. Specifically, he could argue that if DAB entails CC, then it follows from his earlier work that the halo function cannot exist, and thus that DAB is incoherent (because it posits the halo function).

Although Lewis could respond in this way, there are two points to notice. First, although this would still give Lewis an argument against DAB, it

² Many thanks to an anonymous reviewer for pointing me to this work of Lewis.

would be a different argument than the one he in fact uses, the updating argument, which I have shown to be unsound. Second, it is controversial whether Lewis' earlier results are correct (see, e.g., Edgington, 2014; Milne, 1997). Hence they provide an insecure foundation for a revised argument against DAB.³

Another reply Lewis might make to my rebuttal of his updating argument can be guessed based on his response to a related argument by Huw Price. Price (1989) shows that Lewis' updating argument can't be used to refute the following anti-Humean rival to DAB, which Lewis (Lewis, 1996) calls 'Desire as Conditional Belief' (DACB):

$$\text{DACB:} \quad V(A) = C(A \mid A)$$

Lewis' (1996) response to Price is rather involved. Here is my reconstruction of it. Lewis begins by, in effect, distinguishing two claims to which the Humean is committed:

H1. Facts about what a rational agent desires are not fixed or determined by facts about what she believes.

H2. Agents that are alike in being rational, that have the same priors, and that are privy to the same empirical data about the world, can potentially be unlike in what they value and desire.

He then distinguishes two versions of anti-Humeanism, which I will call 'strong' and 'weak'.⁴ The strong anti-Humean denies both H1 and H2, while the weak anti-Humean denies only H1. Lewis represents the strong anti-Humean's claim as follows:

$$\text{DBN:} \quad V(A) = C(G \mid A)$$

Here G amounts to the proposition *that something objectively good happens* (Lewis, 1996, p. 307). Since DBN is intended to express strong anti-Humeanism, it should contradict both H1 and H2, and indeed it does. It contradicts H1 because it implies that facts about what a rational agent values, and thus desires, are fixed by her credences, and hence by her beliefs. It contradicts H2 because it implies that rational agents are all alike in having desires geared toward the same ultimate, objective good, embodied by G .

Lewis is surprisingly non-hostile to strong anti-Humeanism as encapsulated in DBN. He says little about it – only that it is a

³ My own view is that Lewis's (1976) result is definitely erroneous, but I will defend this claim elsewhere.

⁴ Lewis (1996) instead calls these positions 'Desire by Necessity' and 'Desire as Belief'. However, these terms are ambiguous since he uses them *both* as names of the two anti-Humean positions in question *and* as the names of formula that purportedly express these anti-Humean positions. Moreover, the terms are not especially clear – so I don't use them.

'comparatively simple and unproblematic version' of anti-Humeanism (Lewis, 1996, 307). He makes no attempt to refute it. His lack of interest in refuting strong anti-Humeanism is presumably to be explained in terms of his being a sufficiently committed Humean to think strong anti-Humeanism has no real chance of being true (due to its radical denial of H2). Weak anti-Humeanism poses a much more credible threat to Humeanism.

Lewis proceeds to critique various possible formulations of weak anti-Humeanism, including both his own DAB and Price's DACB. As we have seen, he holds that DAB is untenable because it falls prey to the updating argument. His response to Price, in contrast, is that DACB turns out, upon analysis, *to be re-expressible as DBN*. His complaint against DACB is therefore that *it is not a version of weak anti-Humeanism after all*. It is just strong anti-Humeanism in disguise. In Lewis' words, DACB is 'a form of anti-Humeanism, sure enough, but not the right form of anti-Humeanism' (Lewis, 1996, p. 313).

Now, in my above rebuttal of the updating argument I showed that CC is true, and when DAB and CC are put together they entail a theory strikingly similar to DBN. (More on this shortly.) This being so, Lewis might reply to me much as he replies to Price. Specifically, he might contend that I have rescued DAB from the jaws of the updating argument only by showing that DAB amounts to a form, not of *weak* anti-Humeanism, but of DBN, and thus of *strong* anti-Humeanism. In other words, Lewis might, while conceding that I have refuted his updating argument (a major concession on his part), claim that I have nevertheless played into his hands by providing an alternative proof of the result he is after: viz., that DAB is not a tenable form of *weak* anti-Humeanism.

My counter-reply to this potential reply of Lewis' is to deny that DAB and CC together entail a form of strong anti-Humeanism. (I believe Price could respond similarly with regards DACB.)

DAB and CC together entail a theory I will call 'DAB⁺':

DAB:	$V(A)=C(A^\circ)$
CC:	$C(A^\circ)=C(H A)$
DAB ⁺ :	$V(A)=C(H A)$

DAB⁺ and DBN certainly *appear* to be very similar theories. Indeed, they appear to share precisely the same logical form, with the only difference between them being that where DAB⁺ references *H*, DBN instead references *G*. But this apparent similarity is, I will now show, merely superficial. When quantifiers are brought into the open, the two views are exposed as having *different logical forms*.

To see this, let's start with DAB. Lewis proposes DAB as a formulation of *weak*, rather than *strong*, anti-Humeanism. This raises a puzzle. On the

face of it, since DAB says simply that $V(A)=C(A^\circ)$, DAB would seem to imply that if two rational agents share all the same beliefs, and thus assign the same credences to all propositions, then they will necessarily share the same values, and so share the same desires. Thus DAB would appear to be inconsistent with the possibility of a pair of equally well-informed rational agents with the same priors diverging in what they desire. If this were right then DAB would entail the falsity of H2 as well as the falsity of H1, and so it would be a form of *strong* anti-Humeanism, not of *weak* anti-Humeanism. The solution to this puzzle lies in recognizing that the halo function is intended by Lewis to be *agent-relative*. Different rational agents can potentially have different halo functions. That is, it can potentially be the case that for one agent, $A \circ=P$, while for a second agent, $A \circ=Q$, where P and Q are different propositions. If these two agents share all the same beliefs, then they will agree in the credences they assign to P and Q respectively. But they will nevertheless differ in how much they value, and thus desire, A 's being the case, for where the first agent is concerned, $V(A)=C(A^\circ)=C(P)$, while for the second agent, $V(A)=C(A^\circ)=C(Q)$.

Recall that H is characterized by reference to the halo function (since H is defined such that (1) and (2) come out as true). This being so, it follows from the fact that Lewis' halo function is agent-relative that H will be agent-relative too. With this in mind, consider the following two rival ways in which DAB^+ might be formulated when quantifiers are brought into the open:

$$DAB^+1: \quad \exists H \forall x \forall A: V_x(A)=C_x(H|A)$$

$$DAB^+2: \quad \forall x \exists H \forall A: V_x(A)=C_x(H|A)$$

Here $V_x(A)$ denotes the value assigned by a rational agent, x , to an outcome, A . Similarly, C_x denotes the credence function used by x . Notice that, because of where the existential quantifier is positioned, DBN^+1 implies that there is a single proposition, H , which serves as a universal 'yardstick' against which *every* rational agent measures the values of outcomes. On the other hand, DBN^+2 is consistent with the possibility of different agents using different versions of H as their respective yardsticks for measuring value. Clearly the idea that the halo function, and thus the identity of H , can *vary* between rational agents is correctly captured by DBN^+2 , but not by DBN^+1 . We should therefore understand DBN^+ as being equivalent to DBN^+2 , not as being equivalent to DBN^+1 .

Let's now turn to DBN , two rival formulations of which are as follows:

$$DBN1: \quad \exists G \forall x \forall A: V_x(A)=C_x(G|A)$$

$$DBN2: \quad \forall x \exists G \forall A: V_x(A)=C_x(G|A)$$

DBN is, of course, intended by Lewis to encapsulate *strong* anti-Humeanism, with the idea being that G represents an *objective* good, that is

the same for all rational agents and cannot vary between rational agents. This being so, it is obvious that DBN is to be correctly understood as saying the same thing as DBN1, not DBN2.

Thus when quantifiers are brought into the open DAB^+ is revealed as being equivalent to DAB^{+2} , while DBN is revealed as being equivalent to DBN1. Crucially, DAB^{+2} and DBN1 *do not* share the same logical form. Hence, initial appearance to the contrary notwithstanding, DAB^+ and DBN do not share the same logical form either. They *appear* to share the same logical form only because of a scope ambiguity. When this ambiguity is resolved by bringing quantifiers into the open, it becomes clear that DAB^+ is consistent with rational agents *differing* among each other in what they ultimately value, as *per* weak anti-Humeanism, while DBN instead entails that rational agents must be *alike* in what they ultimately value, as *per* strong anti-Humeanism. The potential objection here being considered – that by showing DAB entails DAB^+ I have shown DAB to be a form of strong anti-Humeanism – is thus without foundation.

6. Conclusion

Lewis' updating argument is, I believe, decisively refuted by the above proof of CC. If this is right then anti-Humeans (weak anti-Humeans included), value-realists, and objectivists about ethics have nothing to fear, at least where this particular threat to their philosophies is concerned.

References

- Bradley, R., List, C. (2009). Desire-as-Belief Revisited. *Analysis*, 69(1), 31–37.
- Bradley, R., Stefánsson, H. O. (2016). Desire, Expectation, and Invariance. *Mind*, 125(499).
- Broome, J. (1991). Desire, Belief, and Expectation. *Mind*, 100(2), 265–267.
- Byrne, A., Hájek, A. (1997). David Hume, David Lewis, and Decision Theory. *Mind*, 106(423), 411–428.
- Daskal, S. (2010). Absolute Value as Belief. *Philosophical Studies*, 148(2), 221–229.
- de Finetti, B. (1995). The Logic of Probability. *Philosophical Studies*, 77(1), 181–190.
- Edgington, D. (2014). Indicative Conditionals. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2014). Metaphysics Research Lab, Stanford University. Retrieved from <https://plato.stanford.edu/archives/win2014/entries/conditionals/>

- Hájek, A., Pettit, P. (2004). Desire Beyond Belief. *Australasian Journal of Philosophy*, 82(1), 77–92.
- Lewis, D. (1976). Probabilities of Conditionals and Conditional Probabilities. *Philosophical Review*, 85(3), 297–315.
- Lewis, D. (1986). Probabilities of Conditionals and Conditional Probabilities II. *Philosophical Review*, 95(4), 581–589.
- Lewis, D. (1988). Desire as Belief. *Mind*, 97(387), 323–322.
- Lewis, D. (1996). Desire as Belief II. *Mind*, 105(418), 303–313.
- Milne, P. (1997). Bruno de Finetti and the Logic of Conditional Events. *The British Journal for the Philosophy of Science*, 48(2), 195–232.
- Oddie, G. (1994). Harmony, Purity, Truth. *Mind*, 103(412), 451–472.
- Oddie, G. (2001). Hume, the BAD Paradox, and Value Realism. *Philo*, 4(2), 109–122.
- Price, H. (1989). Defending Desire-as-Belief. *Mind*, 98(389), 119–127.
- Stefánsson, H. O. (2014). Desires, Beliefs and Conditional Desirability. *Synthese*, 191(16), 4019–4035.
- Weintraub, R. (2007). Desire as Belief, Lewis Notwithstanding. *Analysis*, 67(2), 116–122.