# Desire-as-belief revisited

Richard Bradley and Christian List

June 30, 2008

## 1 Introduction

On Hume's account of motivation, beliefs and desires are very different kinds of propositional attitudes. Beliefs are cognitive attitudes, desires emotive ones. An agent's belief in a proposition captures the weight he or she assigns to this proposition in his or her cognitive representation of the world. An agent's desire for a proposition captures the degree to which he or she prefers its truth, motivating him or her to act accordingly. Although beliefs and desires are sometimes entangled, they play very different roles in rational agency.[1]

In two classic papers (Lewis 1988, 1996), David Lewis discusses several challenges to this Humean picture, but ultimately rejects them. We think that his discussion of a central anti-Humean alternative – the *desire-as-belief thesis* – is in need of refinement. On this thesis, the desire for proposition $p$ is given by the belief that $p$ is desirable. Lewis claims that '[e]xcept in trivial cases, [this thesis] collapses into contradiction' (Lewis 1996, p. 308). The problem, he argues, is that the thesis is inconsistent with the purportedly plausible requirement that one's desire for a proposition should not change upon learning that the proposition is true; call this the *invariance requirement*.

In this paper, we revisit Lewis's argument. We show that, if one carefully distinguishes between non-evaluative and evaluative propositions, the desire-as-belief thesis can be rendered consistent with the *invariance requirement*. Lewis's conclusion holds only under certain conditions: the desire-as-belief thesis conflicts with the invariance requirement if and only if there are certain correlations between non-evaluative and evaluative propositions. But when there are such correlations, we suggest, the invariance requirement loses its plausibility. Thus Lewis's argument against the desire-as-belief thesis appears to be valid only in cases in which it is unsound.

[1] On the standard picture, if a proposition can be expressed as a disjunction of several other mutually exclusive propositions, the agent's desire for it is the sum of the agent's desires for each of the component disjuncts, weighted by his or her beliefs in them. If a proposition is maximally specific, however – i.e., it is true in one and only one possible world – there is no entanglement at all: the agent's desire for such a proposition is independent of his or her beliefs.

## 2 Lewis's argument restated

Let $\Omega$ be the set of all relevant possible worlds; $\Omega$ is assumed to be non-empty and, for simplicity, countable. A *proposition* is a subset $p \subseteq \Omega$. Thus the set of all possible propositions is given by the power set $\mathcal{P}(\Omega)$. An agent's beliefs are represented by a probability function $P : \mathcal{P}(\Omega) \to [0, 1]$ with standard properties, and his or her desires by a utility function $U : \mathcal{P}(\Omega) \to \mathbf{R}$. Further, for any proposition $p$, let $P_p$ denote the agent's revised probability function obtained by Bayesian updating after learning that $p$ – i.e., $P_p(\bullet) = P(\bullet|p)$ – and let $U_p$ to denote the corresponding utility function.[2]

To formulate the desire-as-belief thesis, Lewis introduces a 'halo' operator that assigns to each proposition $p$ the corresponding proposition that $p$ is desirable, denoted $p^\circ$. The *desire-as-belief thesis* states that the desire for $p$ is given by the belief in $p^\circ$. Formally,

$$\text{for any } p,\ U(p) = P(p^\circ). \tag{1}$$

Why does Lewis think this thesis 'collapses into contradiction'? Lewis argues that it conflicts with the *invariance requirement* that the agent's desire for $p$ should be unaffected by learning that $p$. Formally,

$$\text{for any } p,\ U_p(p) = U(p). \tag{2}$$

To see the conflict between (1) and (2), notice that (1), understood as a requirement both before and after learning that $p$, and (2) jointly imply that

$$\text{for any } p,\ P_p(p^\circ) = U_p(p) = U(p) = P(p^\circ),$$

and thus

$$\text{for any } p,\ P(p^\circ|p) = P(p^\circ). \tag{3}$$

Claim (3) states that $p^\circ$ and $p$ are probabilistically independent from each other; call this the *independence requirement*. This requirement, however, is not only violated by many probability functions $P$, but even when it is satisfied by $P$, it is usually easy to find another proposition $q \subseteq \Omega$ such that it is violated by the agent's revised probability function after learning that $q$. For example, when $P(p^\circ)$ and $P(p)$ are both above 0 but below 1, then $q = \neg(p \wedge p^\circ)$ is such a proposition: assuming $P(p^\circ|p) = P(p^\circ)$, we have $P_q(p^\circ|p) = 0$ while $P_q(p^\circ) > 0$, and hence $P_q(p^\circ|p) \neq P_q(p^\circ)$.

Thus (1) and (2), both understood as requirements both before and after Bayesian updating, are mutually inconsistent, except in trivial cases. Given the invariance requirement, the desire-as-belief thesis therefore leads to a contradiction.

[2] Suitable provisions are needed if we allow the possibility of conditionalizing on zero-probability propositions. We set these technicalities aside for the present purposes.

## 3 Why Lewis's argument is too quick

It is important to note that, given the desire-as-belief thesis, the invariance requirement not only implies the independence requirement, but is also implied by it. The invariance requirement holds if and only if there are no correlations between proposition $p$ and the proposition that $p$ is desirable. Formally, given (1), understood as a requirement both before and after updating, (2) and (3) are equivalent, as Lewis recognizes. But does this observation help us to defend the desire-as-belief thesis against Lewis's argument? After all, we have seen that (3) is not only violated by many probability functions, but that Bayesian updating can also turn a probability function that satisfies it into one that violates it.

By extending Lewis's analysis in a way that allows us to distinguish between non-evaluative and evaluative propositions, we now show that there exists a class of well-defined probability functions and associated utility functions satisfying the desire-as-belief thesis and the invariance requirement, each understood as requirements before and after admissible instances of Bayesian updating. We do not suggest that the distinction between non-evaluative and evaluative propositions involved in our construction can always be upheld, but its well-definedness is enough to show that, contrary to Lewis's claim, there do exist non-trivial cases in which the desire-as-belief thesis is consistent with the invariance requirement.

The key idea underlying our construction is to express the set $\Omega$ of all relevant possible worlds as a Cartesian product of the set $\Phi$ of all possible non-evaluative states and the set $\Psi$ of all possible evaluative states. Formally, $\Omega = \Phi \times \Psi$. Thus each possible world $\omega \in \Omega$ is an ordered pair $(\phi, \varphi)$ of a non-evaluative state $\phi \in \Phi$ and an evaluative state $\varphi \in \Psi$. Interpretationally, $\phi$ could capture the totality of physical facts holding in that world, and $\varphi$ the totally of normative facts (e.g., ought facts or goodness facts).

A few words of justification may be useful. On some meta-ethical theories, the evaluative facts supervene on the non-evaluative ones, meaning that once $\phi$ is fixed, so is $\varphi$; and hence the non-evaluative states in $\Phi$ cannot be freely combined with the evaluative ones in $\Psi$. But since we do not generally know the correct normative theory, we may still consider a whole range of different evaluative states epistemically possible, given the same non-evaluative state. Therefore, when $\Omega$ is interpreted as the set of epistemically possible worlds – which seems appropriate in a theory of rational agency – it may well make sense to express $\Omega$ as the Cartesian product of a set of non-evaluative states and a set of evaluative ones.

As before, a proposition is a subset $p \subseteq \Omega$. We call $p$ *purely non-evaluative* if $p = p_\Phi \times \Psi$ for some $p_\Phi \subseteq \Phi$ and *purely evaluative* if $p = \Phi \times p_\Psi$ for some $p_\Psi \subseteq \Psi$. A purely non-evaluative proposition has no evaluative implications: its acceptance in no way narrows down the possible evaluative states; and a purely evaluative proposition similarly has no non-evaluative implications. An agent's beliefs over the non-evaluative states can be represented by a probability function $P_\Phi : \mathcal{P}(\Phi) \to [0, 1]$, his or her beliefs over the evaluative states by a probability function $P_\Psi : \mathcal{P}(\Psi) \to [0, 1]$.

We now state a condition on the agent's overall probability function

$P : \mathcal{P}(\Omega) \to [0,1]$ that is sufficient for the satisfaction of (a variant of) (1), (2) and (3) above and guarantees the satisfaction of these requirements even after admissible instances of Bayesian updating. The condition, called *multiplicative decomposability*, requires that the probability of each world is the product of the probability of its non-evaluative state and the probability of its evaluative state. Formally,

$$\text{for any } \omega \in \Omega,\ P(\omega) = P_\Phi(\phi) \times P_\Psi(\varphi), \text{ where } \omega = (\phi, \varphi). \tag{4}$$

It is easy to see that, under (4), any given pair $P_\Phi$ and $P_\Psi$ of non-evaluative and evaluative probability functions induces a unique overall probability function $P : \mathcal{P}(\Omega) \to [0,1]$. As an implication, the probability of a conjunction of a purely non-evaluative proposition and a purely evaluative one is the product of the probabilities of the two propositions. Formally,

$$\text{for any } p = p_\Phi \times \Psi \text{ and } q = \Phi \times q_\Psi \text{ where } p_\Phi \subseteq \Phi \text{ and } q_\Psi \subseteq \Psi,$$
$$P(p \cap q) = P(p) \times P(q) = P_\Phi(p_\Phi) \times P_\Psi(q_\Psi).$$

In consequence, if $p$ is purely non-evaluative and $q$ purely evaluative,

$$P_p(q) = P(q|p) = \frac{P(p) \times P(q)}{P(p)} = P(q), \text{ provided } P(p) > 0,$$

and thus purely evaluative beliefs are invariant under conditionalization on purely non-evaluative ones (and vice-versa).

To show the satisfaction of (a variant of) (1), (2) and (3), let us define a utility function $U$ over all purely non-evaluative propositions as follows: for any such $p$, $U(p) = P(p^\circ)$. Then (1) is satisfied, with the quantification restricted to purely non-evaluative propositions $p$. Further, because of the multiplicative decomposability of $P$, (3) is satisfied whenever $p$ is purely non-evaluative and $p^\circ$ purely evaluative. By implication, (2) is also satisfied in such cases.

If the 'halo' operator maps any purely non-evaluative proposition to a purely evaluative one, all of (1), (2) and (3) are therefore satisfied with the quantification restricted to purely non-evaluative propositions $p$. If the 'halo' operator maps some purely non-evaluative propositions to propositions that are not purely evaluative, then (2) and (3) are satisfied with the quantification restricted to those purely non-evaluative propositions $p$ for which $p^\circ$ is purely evaluative; such propositions $p$ presumably include maximally specific ones of the form $p = \{\phi\} \times \Psi$ where $\phi \in \Phi$.[3] As we have seen, the multiplicative decomposability of $P$ further guarantees that, whenever $p^\circ$ is purely evaluative, $P(p^\circ)$ is invariant under Bayesian updating on any purely non-evaluative propositions.

Under our construction, we can therefore conclude that, if we restrict the quantification as indicated, (1), (2) and (3) are consistent, both before and after updating on non-evaluative propositions.

[3] We think the least objectionable version of the desire-as-belief thesis is the one restricted to maximally specific non-evaluative propositions. On this thesis, we have $U(p) = P(p^\circ)$ for any $p = \{\phi\} \times \Psi$ with $\phi \in \Phi$, while $U(p)$ and $P(p^\circ)$ may come apart for some other propositions.

# 4 Concluding remarks

Lewis's basic argument is that the desire-as-belief thesis conflicts with the invariance requirement because their joint satisfaction requires the absence of any correlations between proposition $p$ and the corresponding proposition that $p$ is desirable – the independence requirement – but we cannot rule out such correlations. We have shown that if $p$ is purely non-evaluative and the proposition that $p$ is desirable purely evaluative, we *can* rule out such correlations, provided the relevant probability function is multiplicatively decomposable. Moreover, in this case, the desire-as-belief thesis and the invariance requirement remain satisfied even after Bayesian updating on non-evaluative propositions.

What went wrong with Lewis's argument? Under our construction, which involves a separation between non-evaluative and evaluative propositions, a multiplicatively decomposable probability function and quantification over propositions of the right kind, Lewis's argument is obviously not valid. It is valid only if violations of the independence requirement are inescapable, at least after certain instances of Bayesian updating.

But is Lewis's argument also sound in such cases? If the independence requirement is violated, either before or after a certain instance of Bayesian updating, it is simply not clear whether we still have any reason to insist on the invariance requirement. If there is a correlation between $p$ and $p^\circ$, then it is no surprise that our evaluation of $p$ may change after learning that $p$. Similarly, if Bayesian updating leads us to redistribute our beliefs not only over non-evaluative propositions but also over evaluative ones, then it is no surprise – indeed, it is expected – that some of our evaluations may change. The kinds of propositions that feature in Lewis's examples of Bayesian updating, such as $\neg(p \wedge p^\circ)$, have both non-evaluative and evaluative implications, namely for the conjunction of $p$ and $p^\circ$, and for this reason it is only natural that learning them affects our assessment of each of the two conjuncts conditional on the other.

Why does Lewis think the invariance requirement is justified? He notes that it will be violated only if it is violated for at least one world (i.e., a maximally specific proposition). But this is impossible, he argues, because worlds are maximally specific with regard to all relevant characteristics, including those evaluative characteristics determining what is good or desirable: '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.' (Lewis 1988, p. 332) But Lewis's argument trades on an ambiguity about the objects of evaluation. It is certainly plausible that our evaluations of the *non-evaluative* characteristics of a world should be fixed by a specification of all its relevant evaluative characteristics. However, it is less obvious that our evaluation of the *evaluative* characteristics of a world should be fixed by the specification of its non-evaluative ones or even of both its evaluative and non-evaluative ones. For whenever there is a probabilistic correlation between what is true and what is desirable, agents may plausibly prefer those worlds with the evaluative characteristics such as to render desirable what is likely to be true.

An example may help to clarify this point. Suppose we seek to evaluate

athletes (or athleticism) on the basis of performance in a decathlon. The relevant non-evaluative facts are those relating to each athlete's achievement in each of the events making up the decathlon: running a certain time in the 100 metres, jumping a certain height in the high-jump and so on. The relevant evaluative characteristics are those determining the relation between achievements in each of the events and how good an athlete someone is. Now it is surely the case that, given a full specification of the latter, the evaluation of someone's athletic prowess is fixed by their performance in each of the decathlon events. On the other hand, someone's evaluation of the relation between performances and athleticism may well depend on how likely they regard it that they will achieve certain performances. If we are poor at the 100 metres but good at the high-jump, for instance, then we will prefer worlds where the mapping from performance to athleticism favours high-jumpers over sprinters. But, given that (contrary to expectation) we will do well in the 100 metres, we might prefer the opposite. So our evaluations of worlds will not be invariant under changes in our beliefs about our likely achievements.

In short, the conditions under which Lewis's argument is valid are ones in which the invariance requirement – Lewis's key premise in arguing against the desire-as-belief thesis – is implausible. Thus Lewis's argument is valid only in cases in which it is unsound. To be sure, we agree with Lewis that the following three claims are mutually inconsistent:

- The desire for proposition $p$ is always given by the belief that $p$ is desirable.
- One's desire for a proposition never changes upon learning that the proposition is true.
- There are sometimes correlations between proposition $p$ and the proposition that $p$ is desirable, either before or after admissible instances of Bayesian updating.

But we do not think the inconsistency should worry us; nobody would plausibly assert all three claims at once. Humeans would uphold the second and the third claims, but drop the first. Desire-as-belief theorists endorsing our present construction would uphold the first and the second claims, but drop the third. And desire-as-belief theorists who recognize the possibility of correlations between $p$ and $p^\circ$ would keep the first and the third claims, but bite the bullet by giving up the second. Of course, the desire-as-belief thesis may give rise to other problems, but we think the particular collapse into contradiction suggested by Lewis is not one of them.[4]

Departments of Philosophy and Government
London School of Economics
London WC2A 2AE, U.K.

[4] We are grateful to Franz Dietrich and Wlodek Rabinowicz for very helpful comments and discussions.

# 5 References

Lewis, D. 1988. Desire as Belief. *Mind* 97: 323-32.

Lewis, D. 1996. Desire as Belief II. *Mind* 105: 303-13.