# DOING WHAT IS BEST

### By Christian Piller

Humeans and anti-Humeans disagree about what constitutes a reason for doing something. The Humean claims that all practical reasons depend on our desires, interests and concerns, on what we like and what we dislike. Using 'desire' in a broad sense that covers all such attitudes, the Humean doctrine is that there are no desire-independent practical reasons. One form of anti-Humeanism is the value theory of practical reason. According to it, we should not be guided by what we happen to want but by what is good. Believing something to be good would be a reason for wanting it to occur and, if possible, for bringing it about. In order to have a definite contrast with Humeanism, the value theorist will oppose any reduction of being good to what an agent wants. Beliefs about what is good would then be desire-independent practical reasons.

David Lewis has advanced an argument that is supposed to refute the value theory of practical reason.[1] This argument, if correct, should have altered moral philosophy. One reason why it has not done so might be its technical nature. I present Lewis' argument in very simple terms in §II. In §III, I discuss three objections that have been raised against it. Building on two of them, I propose in §IV my own analysis of where Lewis goes wrong. I am convinced that something must be wrong with Lewis' argument, because if it were correct it would prove too much. It would refute any plausible theory of practical reason. I shall argue that Humeans and anti-Humeans can agree that we should aim at doing what is best. Their differences concern what makes an option best. I explain in more detail in §I this philosophical reason for my search for a technical problem in Lewis' argument.

---

[1] See D.K. Lewis, 'Desire as Belief', *Mind*, 97 (1988), pp. 323–32; and its sequel 'Desire as Belief II', *Mind*, 105 (1996), pp. 303–13.

## I. THE SIMPLE THEORY OF RATIONALITY

Everyone has to make difficult decisions sometimes. For all these cases there is very simple advice: always do what is best. Would anything be wrong with this simple advice?

Certainly we expect more from a theory of practical rationality. As rational agents we want to do what is best, but doing what is best is an ideal, and a theory of rationality should show us a way to achieve this ideal. So we have to improve on this theory. Here is a suggestion: always do what you *believe* to be best. Your beliefs are accessible to you. This provides you with some handle on how to proceed. At least in principle, you are now in a position to use this theory. I shall call it the simple theory of practical rationality.

If doing what is best sets us the right ideal, then doing what one believes to be best is the right way of pursuing this ideal. (In this argument I use a notion of belief that makes it an all or nothing matter whether one believes something.) It is true that if you follow this method you might fail. If your belief about what is best is false, you will not achieve the ideal. But a good method need not be infallible. The fact that acting rationally does not guarantee success is not a legitimate complaint against any theory of rationality. And anyway, what should you do instead? It could hardly be right to do what you believe would not be best. Even if you achieved the ideal of doing what is best by doing something of which you wrongly believed that it was not best, you would not have achieved the ideal in a rational way. You would have achieved the ideal by pure luck, but not in virtue of acting rationally. So, I take it, our revision improves on the very simple theory. It gives us a method of how to pursue the ideal of doing what is best.

But even if it is an improvement, the simple theory will not satisfy us. We think that if we knew what was best, we would not need any further advice. Telling us to do what we believe to be best is trivial advice. It leaves the difficult questions open. First, what is the right standard by which to judge actions? Is an action best if it maximizes an agent's expected utility function? Is it best if its consequences are more valuable than those of the alternatives? And valuable for whom? Should you do what is best for yourself, or should you do what is best for everyone affected by your actions? And secondly, once you know what makes an option best, you will be confronted with the everyday problem of applying the relevant standard. How do you find out which of your options is the best one? The simple theory does not help with either of these tasks.

For the purposes of decision-making the simple theory is not detailed enough. But nevertheless it gives us something. We think that once we know which action is best, we know how to act. The simple theory thus provides the right framework for a more substantial account of practical rationality. In this respect the simple theory seems to be correct.

Having put the simple theory on the map, I now turn to Lewis' characterization of Humeanism and anti-Humeanism. 'A Humean thesis about motivation says that we are moved entirely by desires: we are disposed to do what will serve our desires according to our beliefs' ('Desire as Belief' p. 324). According to Lewis, 'we are within our rights to construe "desire" inclusively, to cover the entire range of states that move us'. Thus the anti-Humean who wants to claim that we can be moved by beliefs alone has to say that some desires, i.e., some states that move us, are beliefs. 'More cautiously he [the anti-Humean] might say that some beliefs are, at least, necessarily conjoined with corresponding desires' (pp. 323–4). Lewis calls this the *desire-as-belief thesis*. The beliefs the anti-Humean has in mind as being capable of motivating are, of course, evaluative beliefs. Taking this into account, the anti-Humean thesis is the following: 'It is impossible to have a belief about what would be good and lack the corresponding desire.... Or better, since we must acknowledge that desire and belief admit of degree, [an agent as construed by anti-Humeanism] desires things just to the extent that he believes they would be good' (pp. 324, 326).

Lewis talks about motivation, whereas in my characterization Humeanism is a claim about practical reasons. These two are connected in the following way: practical reasons are linked to rational motivation. Rational agents, I assume, will be motivated by the reasons they have. For Lewis, there is no real difference between theories of motivation and theories of rational motivation, because within his functionalist framework the rational relations in which attitudes stand to one another are constitutive of what attitudes they are. Thus my talk about reasons and rationality only makes something explicit that is already implicit in Lewis' account. The anti-Humean says that *for the rational agent* some beliefs are necessarily connected with corresponding desires, or, in other words, that holding some beliefs *rationally commits* one to having corresponding desires.

Lewis' argument against anti-Humeanism exploits a structural difference in the principles that govern rational changes of beliefs and desires in the light of newly acquired evidence. Because rational beliefs and desires change differently, assuming a necessary connection between beliefs and desires would put too much constraint on the rational agent. Lewis' result is therefore very general: *no beliefs* rationally require certain desires. Consequently, whatever our understanding of evaluative beliefs might be, as long as they

are beliefs governed by the principles of rational belief change, their connection with desires cannot be a requirement of rationality.

Lewis' intended targets are internalist, i.e., reason-providing versions of meta-ethical cognitivism: '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' ('Desire as Belief II' p. 307). Mackie's rhetorical scepticism, according to which it is strange to assume the existence of properties that demand their realization of us, would have found a substantive argumentative foundation. Lewis, if correct, can prove that there is no goodness awareness of which would have to motivate rational agents.

The critical force of Lewis' argument, however, cannot be restricted to its intended domain. For example, if rational egoism is correct, then what I want to do should be determined by my beliefs about what is good for me. Lewis' argument would forbid such a principle of rationality, because, as we have seen, *no belief* can commit one to want something. The simple answer, 'So much the worse for egoism', is not a real option, because the critical force of Lewis' argument has by no means been exhausted yet. Take the notion of being good from one's own perspective. If you are concerned only about yourself, then this notion coincides with the notion of being good for you. But if you are concerned about other people as well, the two notions come apart. What is good from your perspective is determined by your concerns. Still, Lewis' argument has it that it cannot be the case that believing some action to be best from your perspective requires you to want to do it. Here is one further example. According to decision theory, rational agents evaluate various alternatives in the light of their desires and beliefs. Why should we not say that according to decision theory the best option is identified by the maximum of expected utility? If decision theory is correct, rational agents, who have beliefs about what is best, will want to do what they believe to be best in the sense given by decision theory. But again if Lewis is correct it could not be a demand of rationality that agents would have to want to do what they believe to be best in this sense.

Decision theory is the theory Lewis himself favours. Nevertheless my argument is more than just *ad hominem*. I want to suggest that as long as the criteria for bestness are left completely open, any plausible theory of rationality fits the framework of the simple theory. Some think that bestness should be understood as maximizing some sort of desire-independent value; others think bestness is at least partly determined by passing certain deontological tests; others again think that what a virtuous person would want to do determines what the best action is. Humeans have their own notion of bestness: what is best for someone depends on what this person wants and

cares about. Theories differ in regard to what according to them makes an action best, but they are all instances of the simple theory.[2] If I am right in the above, Lewis' argument would prove too much, because by refuting the simple theory it would refute any plausible theory of rationality.

There are at least two possible objections to what I have argued so far. The first challenges the correctness of the simple theory. Why, someone might ask, should we always do what we take to be best? Why not do what we think is good enough? An agent who always aims at what is best is obsessed with, let us say, quality. But moderation rather than obsession is a sign of true rationality.[3]

Although satisficing, aiming at what is good enough, looks like an alternative to the simple theory, it would not escape Lewis' argument. If the *belief* that something is best cannot be a reason, the *belief* that something is good enough could not be either. Furthermore, the freedom the simple theory gives us in interpreting bestness can even make room for satisficing. Suppose an action is best if and only if it has the weight of reasons on its side. Then it might be that our reasons are on the side of moderation, but moderation not in terms of a somehow limited reasonableness in our choice of actions, rather with respect to the value of the results we want to achieve. We shall still choose the action that we take to be best in terms of its rationality.

The second objection criticizes the very first step in my development of the simple theory: rational action is not, as I claimed, always doing what one believes to be best; sometimes it is simply doing what is best. One of my arguments, namely, that such a view would abolish the distinction between rational action and successful action, clearly rested on a restricted view of what bestness might be. Suppose bestness is understood as having the weight of reasons on one's side. Then the best action will always be the rational one, and success in the achievement of results is different from bestness in this sense. The main idea behind the move towards the simple theory is to make the demands of a theory of rationality accessible to the agent. Once we allow for an understanding of bestness in terms of an agent's psychological states, for example when bestness is determined by the maximum of an agent's expected utility function, this accessibility seems already to be satisfied by a theory that asks us to do what is best.

I am not convinced by this last objection. Whether it is money, some other good or expected utility that has to be maximized, in all these cases we

---

[2] The view expressed here makes all theories teleological: what one rationally ought to do is determined by what is good. Deontological theories would form a subgroup distinguished from the others by their account of what makes an option good. For example, an action whose maxim does not pass the categorical imperative test can never be good. For further discussion of this view see J. Broome, *Weighing Goods* (Oxford: Blackwell, 1991), pp. 1–16.

[3] Such a view is developed in M. Slote, *Beyond Optimizing* (Harvard UP, 1989).

do not want to call an action rational if the agent, even after thinking about what to do, chooses it without having any idea about whether it is the one that has the relevant maximizing property. If awareness of monetary gains is a necessary condition for the rational pursuit of money, we also think that agents must have an idea about what the expected utilities for various options are, or less formally, an idea about how actions would get them what they want, if they are to act rationally by the lights of decision theory. In general, only reference to beliefs about what is best satisfies the condition that a theory of rationality has to make accessible demands.

However, in order to see that Lewis' argument would prove too much, we do not really need to decide the question whether we should do what is best or what we believe to be best. A reflective agent deliberates about what to do. His practical deliberations end with the judgement that a certain option is best, and this judgement tells him what to do. A reflective agent will always want to act on his belief about what is best, otherwise there would be no point to practical deliberation. A reflective agent might be wrong and do what is not best. But if he deliberates correctly, he will do what is best. In all these cases he will *want to do* what is best because he correctly *believes* it to be best. A belief and a desire are necessarily conjoined. Lewis' argument, if correct, would show that there cannot be a reflective agent who deliberates correctly. A theory of rationality will not demand reflectivity in all cases, but any theory has to *allow* that someone can reflectively follow it. Something must be wrong with Lewis' argument, because it would entail that no rational agent could reflectively use the correct theory of practical rationality, whatever it might be, in a correct way. In short, with the idea of a reflective agent who deliberates correctly, we simply exclude cases in which doing what is best and doing what one believes to be best can come apart.

I can develop the point from a different angle. Confronted with a value theory of practical reason, Humeans can always try to defend their position by arguing for an externalist conception of evaluative beliefs, i.e., they can deny that believing something to be good has to motivate a rational agent. But, *pace* Lewis, a Humean cannot reasonably claim that *no beliefs* can be such that they have to motivate the rational agent. Take beliefs about *what one has reason to do*. Any minimal understanding of what it is to be a rational agent will contain the idea that if rational persons believe that some option is favoured by their reasons, then they must be motivated accordingly, otherwise they just would not be rational. Thus even the Humean has to allow that some beliefs, namely beliefs about one's reasons, have motivational influence in so far as one is rational.

Of course, the Humean will accept only beliefs about *Humean* reasons as motivating the rational agent. The notion which defines Humeanism,

'desire-dependence', will then have to be developed in such a way that *beliefs about Humean reasons*, e.g., beliefs about one's own desires, count as practical reasons. Again, one could object that not beliefs about reasons, but only the reasons themselves have to motivate a rational agent. And again my minimal answer is that rational motivation and true beliefs about what one has reason to do must be necessarily connected.


## II. LEWIS' ARGUMENT

Lewis' general strategy is to show that the idea that our desires should be guided by our beliefs about what is good is incompatible with two other principles of rationality. Because these other principles are unproblematic, it is the desire-as-belief thesis that has to be rejected.

What are these other two unproblematic principles of rationality? First, there is an epistemological principle. Epistemic attitudes come in degrees. Some beliefs are stronger than others. A system of beliefs can only be rational if the degrees of the beliefs are representable by a probability function C. The following principle (P1) is a theorem of probability theory. The propositions $E_i$ in this principle form a partition, i.e., they are a set of mutually exclusive and jointly exhaustive propositions. For example, E and not-E form a partition.

P1.   $C(A) = \Sigma_i C(A \mid E_i)C(E_i)$.

Whereas $C(A)$ means the rational degree of belief of a person in respect to proposition A, $C(A|E)$ means the rational degree of belief of a person in respect to A's being the case given that E is the case. $C(A|E)$ is a conditional probability and is defined as follows: $C(A|E) = C(A \& E)/C(E)$. (P1) is a plausible condition on rational beliefs. Some people might be opposed to it because they oppose a quantitative notion of belief in general. But whatever the merits of such a notion, the conflict between the desire-as-belief thesis and other principles of rationality can hardly be explained by the fact that one of these principles assumes degrees of belief. Thus, at least in the context of our discussion, (P1) should be regarded as an innocent principle.

What is Lewis' second unproblematic principle? It shows us how in decision theory desires and beliefs determine the degree to which we should want other things. It says that a system of desires has to be such that their degrees can be represented by an expected utility function V. Again the E-propositions form a partition:

P2.   $V(A) = \Sigma_i V(A \& E_i)C(E_i|A)$.

Why is (P2) supposed to be unproblematic and acceptable to both Humeans and anti-Humeans? I have argued that everyone agrees that we should always do what we believe to be best. (P2) tells us how to figure out which of our options is best. In a decision problem you have a choice between two actions. One action, A1, will give you something good for sure, let us say £100. (This example is meant to be neutral in regard to Humeanism and anti-Humeanism. The Humean will understand 'It is good to get £100' in terms of what an agent wants.) The other action, A2, will give you £120 or nothing; and suppose you know that there is only a very small chance that A2 will give you £120. What should you do? (P2) confirms our intuitive judgement: it is rational to do A1. Although V(£120) is higher than V(£100), V(A1) is higher than V(A2) because C(£120|A2) is very low. Why should we do A1? Because A1 is better than A2. A1 is better than A2 because it promises more. To get £100 for certain is a better prospect than to have a tiny chance of getting £120. A1 has a higher *expectation of goodness* than A2. Our ordinary concept of the best option is one that is determined by the expectation of goodness as defined in (P2).

Humeans and anti-Humeans agree on (P2); they differ in regard to whether something has to be added to (P1) and (P2) in order to have a set of sufficiently strong conditions of rationality. Whereas the Humean will allow any system of desires, the degrees of which are representable by V, to be reasons, the anti-Humean will demand a further condition. Desires have to reflect our perceptions of value: only then can they be legitimate practical reasons. The postulate that needs to be added, Lewis thinks, is the desire-as-belief thesis.

The idea we want to express is that rational desires should match our beliefs concerning the value of states of affairs. Suppose for every proposition A there is a proposition Å which says that it is good that A. If someone believes that Å, i.e., believes that A's being the case is a good thing, then he should want A to be the case.[4]

P3.  $C(Å) = V(A)$.

Now we have all the ingredients for Lewis' argument. Lewis attempts to show that these three principles restrict in implausible ways the values the two functions C and V can take. There are situations in which a rational agent will not have beliefs and desires that obey all three of the principles.

---

[4] I have simplified a bit. (P3) does not allow for degrees of goodness – either something is good or it is not. It allows for degrees of beliefs in respect to goodness, but these cannot capture degrees of goodness. There is a difference between being totally convinced that my phone only works half the time and thinking that there is a 50–50 chance that my phone works perfectly. Lewis' proof is not affected by removing this simplification. See Lewis, 'Desire as Belief' pp. 330–1; and §IV below.

1. *The first problem: Lewis 1988*

Suppose an agent's desires and beliefs satisfy all the three principles at a certain time. But over time the agent learns certain things. Suppose an agent learns that E is the case. How should his other beliefs and desires change?

To take beliefs first, before the agent learned E, his belief in A was C(A). The agent's epistemic system also contained the conditional probability C(A|E). It is generally assumed that a rational agent's new degree of belief in A will equal his old conditional probability of C(A|E). This conditional probability expresses the epistemic commitment of how to react to an experience of E. When one actually experiences E, it is rational to believe what one has committed oneself to believe. In the case of learning E we have $C_{new}(A) = C_{old}(A|E)$. Rational beliefs evolve by conditionalization.

How does the same evidence affect the desires of the agent, i.e., how does V(A) change if the agent learns that E is the case? (P2) gives the answer. The conditional probabilities, C(E|A) and C(not-E|A), which are the weights for V(A & E) and V(A & not-E), become 1 and 0 respectively. Why? Because $C_{new}(E|A) = C_{old}(E|A \& E)$, and a belief in something given that it is the case should always be 1. Thus one element of the sum that determines V(A) in (P2) drops out, and the other gets all the weight: $V_{new}(A) = V_{old}(A \& E)$.

The desire-as-belief thesis (P3) tells us that some beliefs, beliefs about what is good, are necessarily conjoined with desires. Lewis exploits the difference between how beliefs and desires change in the light of new evidence to show that there will be situations in which not all three principles (P1)–(P3) will be satisfied. In particular, he shows the following:

Problem L1.   There is no proposition A such that learning that some evidence E obtains could change both the degree of belief in A and the degree to which A is desired.

An agent whose beliefs and desires satisfy (P1)–(P3) 'cannot simultaneously change both his opinion about whether A and his desire about whether A'. But there are situations, Lewis claims ('Desire as Belief' p. 328), in which there is a proposition that has such an effect:

> 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 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 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.

The proof Lewis gives for (L1) is a bit complicated. An easier proof will emerge as we go on.

## 2. *The second problem: Price*

There are consequences we can derive from our three principles. The epistemological principle (P1) is a principle for all beliefs. So it should also hold in respect to beliefs concerning the goodness of states of affairs:

T1.  $C(\text{Å}) = \Sigma_i C(\text{Å} | E_i) C(E_i)$    [(P1) applied to °-propositions].

(P2) determines the value of the V-function for any proposition A. (P3) tells us that the value of the V-function for A equals the value of the C-function for Å. We can put these two things together and get

T2.  $C(\text{Å}) = \Sigma_i V(A \& E_i) C(E_i | A)$   [from (P2) and (P3)].

According to (P3), we can replace the value of the V-function for any proposition by the value of the C-function for the corresponding °-proposition. Doing this for the values of the V-function in (T2), we get

T3.  $C(\text{Å}) = \Sigma_i C([A \& E_i]°) C(E_i | A)$   [from (T2) and (P3)].

   Now let us compare (T1) and (T3). Both of them determine the value of C for °-propositions. In situations that satisfy all the three conditions (i)–(iii) given below they would give us different credence values for °-propositions, and then (P1)–(P3) could not all be satisfied:

(i)     For all $E_i$, $C(\text{Å} | E_i) = C([A \& E_i]°)$
(ii)    For all $E_i$, $C(E_i) \neq C(E_i | A)$
(iii)   For all $E_i$ and $E_j$ ($i \neq j$), $C([A \& E_i]°) \neq C([A \& E_j]°)$.

The formal idea of a situation that satisfies (i)–(iii) is the following: the first elements of the sums of (T1) and (T3) are the same (condition i) but the second elements of these sums are different (condition ii). So the same elements get different weights. This makes a difference if the items that are weighted are themselves of different value (condition iii). Huw Price gives the following example in which it seems reasonable to assume that (i)–(iii) are satisfied.

   My dear Aunt Agatha may die in 1989. Let A be the proposition that she does so. My interest in the truth of A is entirely constrained by the facts that I am Aunt Agatha's sole heir, that she is periodically wealthy, and that money is my sole joy. Thus I think that it would be good if Aunt Agatha dies in 1989 if and only if in that case I inherit a

fortune. To all intents and purposes, therefore, Å is the proposition 'I inherit a fortune in the case that A'. Both Agatha's prospects and mine depend on the state of the economy. There may or may not be a recession in 1989: let $E_1$ be the proposition that there is and $E_2$ ($\equiv$ not-$E_1$) the proposition that there is not. This affects my prospects in this way: I am a lot less likely to inherit if Agatha dies in a recession than if she dies otherwise (since obviously in a recession there is less likely to be a fortune to be inherited). To put some figures on it, suppose $C(Å|E_1) = 0$ and $C(Å|E_2) = 1$.... As for Aunt Agatha's prospects, she has always said that she would hate to die in a bear market, and so she will try to hang on if the economy is down. Let us say then that $C(A|E_1) = 1/6$ and $C(A|E_2) = 1/3$. Suppose finally that $C(E_1) = C(E_2) = 1/2$.[5]

This example satisfies the conditions (i)–(iii). With regard to (i), we know already that $C(Å|E_1) = 0$ and $C(Å|E_2) = 1$, and we need to argue that $C([A \& E_1]°) = 0$ and $C([A \& E_2]°) = 1$. Huw will not get any money if Agatha dies while there is a recession, but he will if there is no recession, so (i) is satisfied. (Of course, (i) will not always be satisfied, as in a case in which A stands for something you are indifferent about and $E_1$ stands for something very important to you.) What about condition (ii)? It is easy to compute that $C(E_1|A) = 1/3$, whereas $C(E_1)$ equals $1/2$. Finally, we also know that it makes a difference for the goodness of Agatha's death whether there is a recession or not. Thus $C([A \& E_1]°)$ is different from $C([A \& E_2]°)$, and condition (iii) is satisfied as well.

If we compute $C(Å)$ according to (T1), we get $C(Å) = C(Å|E_1)C(E_1) + C(Å|E_2)C(E_2) = 1/2$. If we use (T3) instead, $C(Å) = C([A \& E_1]°)C(E_1|A) + C([A \& E_2]°)C(E_2|A) = 2/3$. This shows that the satisfaction of (P1)–(P3) is not compatible with the reasonable assignments of personal probabilities in Price's example. The problematic implication of the desire-as-belief thesis (P3) is thus the following:

Problem L2. No situation can be such that the desires and beliefs of a rational agent satisfy conditions (i)–(iii) above.

## 3. *The third problem: Lewis 1996*

We have already seen how V(A) changes in the light of newly acquired evidence. Suppose now that the evidence is A itself: $V_{new}(A) = V_{old}(A \& A) = V_{old}(A)$. Thus acquiring the information that A is the case will not change our degree of desire for A. $C_{old}(Å)$, however, will change to $C_{new}(Å)$, which is determined by $C_{old}(Å|A)$. Because V(A) does not change when we learn that A, $C(Å)$ cannot change either, if (P3) is true. Surely (P3) should still hold after we learn that A. Thus $C(Å)$ must always equal $C(Å|A)$. This is the third worrying implication of the desire-as-belief thesis.

---

[5] Huw Price, 'Defending Desire-as-Belief', *Mind*, 98 (1989), pp. 119–27.

Problem L3.  Nothing can be evidence for its own goodness: i.e., for all A,
$C(Å) = C(Å|A)$.

Lewis rightly claims that this is just implausible. Obviously the occurrence of
a thing can be evidence for its own goodness. Suppose John only joins the
party if he is in a good mood, and if he is, the party will be a success. Then
John's arrival is evidence for its own goodness.

4. *A fourth problem*

One of the conditions of Price's counter-example was the following: $C(Å|E)$
$= C([A \& E]°)$. I have said that this condition is by no means trivial. In all
cases in which A is indifferent and E is good (and the indifferent A does not
subtract from E's value), this condition will not be satisfied. But we can
easily derive this condition from the desire-as-belief thesis. Suppose we
assume that $C(Å) = V(A)$ still holds after we learn that E is the case. $V(A)$ will
change to $V(A \& E)$ and $C(Å)$ to $C(Å|E)$; thus after we learn that E,
$V(A \& E)$ will equal $C(Å|E)$. Applying the desire-as-belief thesis to $V(A \& E)$,
we get

Problem L4.  $C([A \& E]°) = C(Å|E)$.

There will be cases in which this does not hold. You are pretty confident
that the conjunctive state of affairs consisting of your being happy and the
mathematical truth that two plus two equals four is good. This conjunctive
state is good because it is good that you are happy. But it is neither good nor
bad that two plus two equals four, and this is so independently of whether
you are happy or not.

5. *How the problems are connected*

The second problem is an instance of the third problem. In Price's example
Agatha's death is evidence for its own goodness. Agatha's death is evidence
for there being no recession which is a sufficient condition for her death's
being good.

The first problem was that the desire-as-belief thesis implies that there is
no E that could change both $C(A)$ and $V(A)$. Suppose that A is evidence for
its own goodness. Given that $C(A) \neq 1$, then there would be evidence,
namely, A, that on the one hand changes $C(A)$ and, because A is evidence
for its own goodness, on the other hand also changes $V(A)$. Thus, if (L3) is
false, i.e., if something can be evidence for its own goodness, then (L1) would
be false as well.

The fourth problem does not seem to have much to do with the question
whether something can be evidence for its own goodness. But it can be used

to prove (L1) in an easy way. In constructing his counter-example, Price used (T1) and (T3), which are consequences of (P1)–(P3).

T1.  $C(\AA) = \Sigma_i C(\AA \mid E_i) C(E_i)$
T3.  $C(\AA) = \Sigma_i C([A \ \& \ E_i]°) C(E_i \mid A)$.

(L4) says that the value of the first part of each of these sums always has to be the same.

L4.  For all $E_i$, $C([A \ \& \ E_i]°) = C(\AA \mid E_i)$.

Thus (T1) and (T3) will only give us the same sums if either the probability weights will also be the same, i.e., if for all $E_i$, $C(E_i \mid A) = C(E_i)$, or if what is weighted is the same, i.e., $C([A \ \& \ E_i]°)$ is the same for all $E_i$. If A is not evidence for or against E, then E is not evidence for or against A. If all the $C([A \ \& \ E_i]°)$ are the same, then no E makes an evaluative difference. Therefore (T1) and (T3) will only give us the same sums if for all E, either E is not evidence for A or E does not make an evaluative difference. This just is (L1).

### III. LEWIS' CRITICS

Huw Price, John Broome and Graham Oddie have raised three different objections to Lewis' argument.[6] I want to comment briefly on them.

To explain Price's idea, I shall focus on the third problem. The desire-as-belief thesis has the implausible consequence that nothing can be evidence for its own goodness. Price suggests that we should therefore reformulate the desire-as-belief thesis. Instead of saying that $C(\AA)$ equals V(A), Lewis' anti-Humean should make the claim that $C(\AA \mid A) = V(A)$. Price (p. 122) thinks that this thesis is also more plausible:

> It is that if we can see that the value we ascribe to A would be liable to change, were we to discover that A, then the appropriate value to use in deliberation is the value A would have for us in those circumstances. The guiding principle is that whenever it makes a difference, we should assess a possible outcome under the hypothesis that it is the actual outcome.

Price's suggestion cannot deal with all the problems. In particular, it cannot handle (L4). But there are independent reasons why Lewis' anti-Humean should not take up Price's suggestion. Suppose some people believe that our world is the best possible world, in the sense that our world contains

only good things and all the good things. Whenever these people are unsure about whether a specific event will occur, they think that if it occurs it will be good. If they believe with a degree of 1/2 that A will occur, then it seems reasonable that the degree of goodness they assign to A should also be 1/2. But under Price's hypothesis they will assign a value of 1 to A. After all, given that A occurs it will be good. The same holds for not-A. An inconsistency arises because, on the one hand, these people believe that our world contains all good and only good, and on the other they believe that both A and its negation are good. So we can put Price's suggestion aside.

John Broome takes issue with Lewis' understanding of the conflict between Humeans and anti-Humeans. Broome distinguishes between the desire-as-belief thesis and the desire-as-expectation thesis. In general, an agent's expectation of any quantity is a probability-weighted average of the values the agent thinks it can assume. (P2) tells us that it is a rationality requirement on degrees of desires that they are expectations in this sense. According to Broome (p. 266), Humeans and anti-Humeans agree on the desire-as-expectation thesis, but neither of them should hold the desire-as-belief thesis:

> Lewis thinks this conclusion [that the desire-as-belief thesis is false] damages the anti-Humean position, because he associates this position with the desire-as-belief thesis. But actually an anti-Humean is no more committed to the desire-as-belief thesis than a Humean. Any teleologist (we are supposing) is committed to the desire-as-expectation thesis. But an expectation is not a belief.

If Broome is right, neither the desire-as-expectation thesis nor the desire-as-belief thesis can separate Humeans from anti-Humeans. Both agree on the first thesis and both deny the second. What, then, is the difference? Anti-Humeans, Broome goes on to say, hold the following position:

> Sometimes 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 result from beliefs alone, need enter into it.

The anti-Humean says that we should act in accordance with our beliefs about what is good. Desires only come into the picture as the rational consequences of our evaluative beliefs. Evaluative beliefs are the foundation of practical reasons. I am in complete agreement with Broome in regard to this characterization of anti-Humeanism. But how should we capture this position if not by something like the desire-as-belief thesis?

Whereas Broome thinks no one should accept the desire-as-belief thesis, I think everyone should. After all, everyone should want to do what he believes to be best. The issue that really divides Humeans and anti-Humeans is what makes an option best. The Humean will look at an agent's

desires to determine bestness, the anti-Humean will explain bestness in terms of desire-independent values. Both Humeans and anti-Humeans will endorse (P3); the difference between them lies in their different interpretations of Lewis' °-function.

Lewis' argument rests on the idea that beliefs change by conditionalization. Graham Oddie denies that they do. The following example illustrates his point. Thinking back to Price's Aunt Agatha, we know that, calculated in accordance with (P2), the degree of goodness that an agent will assign to her death is 2/3. Assuming that this is the correct calculation, the agent should be completely confident that her death is good to a degree of 2/3. Suppose it turns out that Agatha dies in a recession. How should the agent change his belief with respect to the degree of goodness of Agatha's death? We think that now he should invest complete confidence in the hypothesis that the degree of goodness of Agatha's death is 0. But how is this possible? It is a well known feature of conditionalization that once you invest complete confidence in a hypothesis no evidence could ever change your belief.

Oddie concludes that beliefs about degrees of goodness do not evolve by conditionalization. His explanation is that evaluative propositions are not timeless propositions but are, as he calls them, 'changing' propositions. Such propositions, Oddie argues (p. 466), never change by conditionalization:

> Consider any changing proposition, like 'It is raining' (R). On Monday it is not raining, you are standing outside totally dry (D), and so your credence in R (in the light of D) is 0. But on Tuesday along comes the downpour, and your evidence for this is overwhelming because you are now standing outside soaking wet (~D), and your credence in R, in the light of ~D, rockets to 1.... It is clear, then, that credences involving changing propositions cannot, and should not, be revised by conditionalization.

Oddie is on the right track. We have to focus on how evaluative beliefs change in the light of new evidence. I do not think, however, that conditionalization is to blame.


## IV. THE SOLUTION OF THE PROBLEMS

As a first step towards a solution we have to remove the simplifying assumption that something is either simply good or not good. Lewis has shown ('Desire as Belief' p. 330) how his argument works for propositions about degrees of goodness. Let $\mathring{A}_j$ be the proposition that A is good to degree $g_j$. We take all hypotheses concerning A's degree of goodness and weight them with $g_j$. This gives us (P3*), a generalized version of the desire-as-belief thesis:

P3*.  $\Sigma_j C(\mathring{A}_j)g_j = V(A)$.

If an agent is completely confident that A's degree of goodness is $g_j$, then according to (P3*) the agent should desire A to degree j. I shall mainly refer to this special case of (P3*): if $C(\mathring{A}_j) = 1$, then $V(A) = j$.

The second step of the solution focuses on what degrees of goodness are. I repeat that a Humean need not object to such a notion, because nothing has ruled out a desire-based account of goodness. When I argued that everyone ought to accept (P2), I said that our ordinary comparative judgements about which option is better rest on judgements of goodness that obey (P2). Getting £100 for sure is better than having a small chance of getting £120. The degree of goodness of the first alternative is, under any understanding of goodness, higher than that of the second. It is thus natural to suggest that, formally speaking, degrees of goodness simply are expectations of goodness.

In my informal discussion of Lewis' result, I said that if Lewis were right, even reflective Humeanism would be refuted. Now we are in a position to explain this fact within the framework of his argument. A reflective Humean orders options on a scale of goodness in regard to how he perceives their expected utilities. If degrees of goodness are expectations of goodness, i.e., if $\mathring{A}_j$ if and only if $V(A) = j$, then any condition like (P3) or (P3*), which demands that beliefs concerning $\mathring{A}_j$ should be in accordance with the value we attribute to $V(A)$, just introduces a condition of reflectivity. A rational agent who is certain that he wants A to a degree $g_j$ will also want it to exactly that degree.

The claim crucial to my solution is that degrees of goodness behave, formally speaking, like expectations of goodness. Rational belief change regarding beliefs about degrees of goodness should thus be seen as belief change about expectations of goodness. It is this fact that will show us where Lewis' argument has gone wrong. In order to see how beliefs about expectations change in the light of new evidence, I shall first look at a related phenomenon. The belief-involving nature of expectations makes the change of second-order beliefs a useful model.

Suppose you receive some evidence that changes your first-order belief in regard to $p$ from, let us say, 1/3 to 1/2. How does your belief about your degree of belief in respect to $p$ change? Assuming that the evidence does not affect your second-order ability of having accurate beliefs about the beliefs you have, you will change from being convinced that your degree of belief regarding $p$ was 1/3 to being convinced that now it is 1/2. We encounter what above I claimed was a puzzling phenomenon. Even if you were certain that your degree of belief was 1/3, now you are certain it is something else. Does this show that conditionalization cannot capture such a change? Not

at all. Applying conditionalization both at the first-order and at the second-order level shows us that the object of your second-order belief has changed. Whereas before the evidence was received the object of your second-order belief with regard to your degree of belief concerning $p$ was your believing to a degree of $1/3$ that $p$, it is now your believing to a degree of $1/2$ that $p$. Introducing degrees of beliefs makes the indexical nature of the question 'What is your second-order belief regarding your belief that $p$?' apparent. The evidence has changed your degree of belief concerning $p$. Thus it has changed what your second-order belief is about.

Treating degrees of goodness as belief-involving expectations, we can directly transfer the result of the above to the case of beliefs about degrees of goodness. Learning that E is the case can affect the belief $C(V(A) = j)$ at both levels. As in the case of beliefs about beliefs, the crucial change takes place on the level of the object of the second-order attitude. An evaluatively relevant piece of information E changes $V(A)$ to $V(A \& E)$. This means that the object of the belief in the degree of goodness of A has changed as well. The value of your degree of belief in A's expectation of goodness can change from 1 to 0, because the evaluatively relevant information E has changed what this belief is about. Whereas, before the information, you were certain that A is good to some degree, now you are certain that A is good to some other degree.

Everything is now in place to criticize the reasoning that leads to the problems (L1)–(L4). I start with (L4), which I derived thus: learning that E changes the left-hand side of $V(A) = C(\mathring{A})$ to $V(A \& E)$, and applying (P3) to $V(A \& E)$, we get $C([A \& E]^\circ)$. The right-hand side changes from $C(\mathring{A})$ to $C(\mathring{A}|E)$, which gives us the implausible consequence (L4), $C([A \& E]^\circ) = C(\mathring{A}|E)$. Introducing degrees of goodness and taking into account that degrees of goodness are expectations of goodness, we see that the right-hand side transformation was simply mistaken. The information E, if evaluatively relevant, changes the object of $C(\mathring{A}_j)$. Whereas before learning E the agent was certain that A's degree of goodness was $g_j$, now he is certain that A's degree of goodness is not $g_j$ any more; it is, let us say, $g_k$. The only consequence we could now derive along such lines would be $C([A \& E]^\circ_k) = C(V(A \& E) = k)$. But this would not be implausible any more.[7]

As for problem (L3), in order to solve it we have to show how something can be evidence for its own goodness. Because $V(A) = V(A \& A)$, learning that A does not change $V(A)$, and therefore also the object of $C(\mathring{A}_j)$ does not change by learning that A obtains. Thus, given (P3), $C(\mathring{A})$ really does equal $C(\mathring{A}|A)$. But how can this be, if A is evidence for its own goodness? In

---

[7] The same criticism applies to the three lemmas derived in Lewis, 'Desire as Belief II' pp. 310–11, where Lewis criticizes Price's suggestion mentioned above.

Price's example in which A, Agatha's death, is evidence for its own goodness because it is evidence for not-E, i.e., no recession, which is a necessary and sufficient condition for A's being good, how good is it that Agatha dies? (P2) gives the answer $V(A) = V(A \& E)C(E|A) + V(A \& \text{not-}E)C(\text{not-}E|A)$. The fact that A is evidence for its own goodness *has already been reflected* in the above calculation of V(A). $C(\text{not-}E|A)$ is bigger than $C(E|A)$; thus the preferred outcome A & not-E gets the bigger weight. In accordance with (P3*), which degree of goodness we should assign to A, i.e., for which degree of goodness $g_j$ $C(\mathring{A}_j)$ should be 1, will also already have been determined by the value of V(A). Thus it is true that $C(\mathring{A}_j)$ equals $C(\mathring{A}_j|A)$. Nevertheless A can be evidence for its own goodness. Whether it is or not influences the value of V(A), and thus also determines the degree of goodness $g_j$ assigned to A in $C(\mathring{A}_j)$.[8] (L1) and (L2) do not throw up independent problems. (L2) is an instance of (L3); and I have already argued that refuting (L3) also refutes problem (L1).

I can highlight the main point of my solution by considering the following objection. All you have shown, it might be said, is that a Humean can escape Lewis' argument. Only a Humean can accept that degrees of goodness are expectations of goodness, i.e., that $\mathring{A}_j$ if and only if V(A) = j. V is a function that tells us the degree to which some proposition is *desired*, but for the anti-Humean the proposition that A is good to some degree has nothing to do with desires. Lewis leaves the interpretation of the °-function open. Thus you are right that one *can* take it to mean the same as the V-function. But if one does so, one has smuggled in a Humean analysis of goodness.

I answer that the real problem is to find the right way to express the anti-Humean position. Let us distinguish between a function VAL(A) assigning to propositions desire-independent degrees of being valuable, and a function DES(A) assigning rational degrees of desire to propositions. The anti-Humean wants to say that our beliefs about how valuable something is should inform the degree to which we desire it, i.e., it is a condition of rationality that if $C(VAL(A) = j) = 1$, then DES(A) = j. What is essential to my solution, and this also shows its close connection to Broome's point, is simply that not only DES but also VAL obeys a principle like (P2). They are both expectations in this sense. All I need to claim is that even an objective notion of value will be probabilistic. (We can even drop the assumption that it is belief-involving: an objective notion of probability would fit our purposes

[8] A further problem remains. I have shown that something can be *indirect* evidence for its own goodness, i.e., something can be evidence for something else that is a condition of its being good, but should it not also be possible that something can be *direct* evidence for its own goodness? Suppose A is good if and only if A occurs. In such a case a partition consisting of all possible assignments to C(A) plays the role of the partition E/not-E in the case of its being indirect evidence for its own goodness.

equally well.) That any notion of value will be probabilistic, and will thus conform to (P2), seems highly plausible. Under any plausible understanding of value a lottery ticket, for example, will become more valuable once it has been declared the winning ticket. If asked what the difference is between my functions VAL and DES on the one hand and Lewis' *intended* interpretation of the °-function and V on the other hand, I would have to answer 'Not much'. The only difference is to make it apparent that degrees of goodness are expectations. The trick of Lewis' argument is to exploit the difference between how expectations and beliefs change in the light of new evidence. But if we compare expectations and *beliefs about expectations*, the relevant difference has vanished.[9]

*University of York*