DESIRE BEYOND BELIEF

Alan Hájek and Philip Pettit

Australasian Journal of Philosophy, Vol. 82, March 2004, 77-92.
Reprinted in *Lewisian Themes: the Philosophy of David K. Lewis,* eds. Frank Jackson and Graham Priest, Oxford University Press, 2004, 78-93.

Abstract

David Lewis [1988, 1996] canvases an anti-Humean thesis about mental states: that the rational agent desires something to the extent that he or she believes it to be good. Lewis offers and refutes a decision-theoretic formulation of it, the 'Desire-as-Belief Thesis'. Other authors have since added further negative results in the spirit of Lewis'. We explore ways of being anti-Humean that evade all these negative results. We begin by providing background on evidential decision theory, and on Lewis' negative results. We then introduce what we call the *indexicality loophole:* if the goodness of a proposition is indexical, partly a function of an agent's mental state, then the negative results have no purchase. Thus we propose a variant of Desire-as-Belief that exploits this loophole. We argue that a number of meta-ethical positions are committed to just such indexicality. Indeed, we show that with one central sort of evaluative belief — the belief that an option is right — the indexicality loophole can be exploited in various interesting ways. Moreover, on some accounts, 'good' is indexical in the same way. Thus, it seems that the anti-Humean can dodge the negative results.

David Hume's rejection of necessary connections between distinct existences was thoroughgoing. He was as wary of them among psychological states as he was of them among external events. In particular, he argued that there are no necessary connections between *beliefs* and *desires,* even those of a perfectly rational agent; thus, he maintained, there are no beliefs that rationally require corresponding desires, and there are no desires that rationally require corresponding beliefs. So one way of being an anti-Humean about mental states is to insist that rationality *does* place certain constraints on which beliefs and desires can be simultaneously held. For example, one sort of anti-Humean might insist that a state of believing (perhaps to a certain degree) something to be good requires a corresponding desire (or degree of desire) for that thing. Or, conversely, she might insist that every (degree of) desire requires a corresponding (degree of) belief in the goodness of the object of that desire.

David Hume set up the terms of the debate, but David Lewis gave formal expression to it. He turned to evidential decision theory, a widely endorsed theory of rational belief and desire. Decision theory represents the 'belief' component of a rational agent's state of mind with a

probability function, the 'desire' component with a value function. Lewis formulated a specific version of anti-Humeanism which he dubbed the 'Desire-as-Belief Thesis': roughly, a rational agent's degree of desire in a proposition A is always matched by her degree of belief in a related proposition, A^{o}. He then went on to refute the Desire-as-Belief Thesis, thus refuting a significant anti-Humean position. A number of other authors piled on further results against theses in the spirit of Desire-as-Belief, restricting further the anti-Humean's options. The collective upshot of these results is that rationality would be compromised by this sort of anti-Humean connection between belief and desire. Thus, in bearing no such connection to belief, it seems that desire goes 'beyond' belief, being apparently irreducible to it and, at least sometimes, unconstrained by it.

But philosophical positions do not die, they merely transmogrify. In this paper we explore ways of being anti-Humean that evade the Lewis-style negative results. The paper is in three sections. The first section provides background on evidential decision theory, and on Lewis' negative results. The second section introduces what we call the *indexicality loophole*: if the goodness or otherwise of a proposition is indexical, partly a function of an agent's mental state, then the negative results have no purchase. Thus we propose a variant of the Desire-as-Belief Thesis that exploits this loophole. This is not merely a technical point, for a number of meta-ethical positions are committed to just such indexicality. Indeed, the third section shows that with one central sort of evaluative belief — the belief that an option is right — the indexicality loophole can be exploited in various interesting ways. Moreover, on some accounts, 'good' is indexical in the same way. This should come as no surprise, given the parallel indexicality of 'rational' that many will acknowledge. Thus, it seems that the anti-Humean can dodge the negative results. Desire may not outrun corresponding belief so easily after all.

1. The anti-Desire-as-Belief results

1.1. Background: evidential decision theory

We follow Lewis in working within the framework of Bayesian decision theory, à la Jeffrey [1983]. Think of propositions as sets of possible worlds, including the empty proposition $\varnothing$. At any time, the mental state of a rational agent can be represented by a pair of functions <C, V>; C is the agent's subjective probability function ('C' evocative of '**c**redence function'), which assigns a number in the interval [0, 1] to each proposition. It conforms to the usual probability axioms: in particular, it is (at least) finitely additive: $C(A \cup B) = C(A) + C(B)$ if $A \cap B = \varnothing$. V is

the agent's **v**alue function, which assigns a real number to each proposition.[1] It obeys its own rule of additivity: if $\{A_i\}$ is a partition of A, then

$$V(A) = \sum_i V(A_i).\ C(A_i|A).$$

V(A) represents the desirability of A by the lights of the agent.

Evidential decision theory exhorts the rational agent to perform an action that maximises V; it is called 'evidential' because the conditional probability weights that figure in the sum can be regarded as giving information about the evidential relevance of A to A_i.[2]

Bayesianism also teaches a lesson about how an agent should update or revise belief in the light of new evidence. Suppose that the agent receives some evidence, the totality of which we denote by E, and on its basis updates to a new probability function C_{new}. The favored updating rule among Bayesians is *conditionalization:* C_{new} is related to C by:

$C_{new}(X) = C(X|E)$ (provided $C(E) > 0$).

Jeffrey conditionalization generalises this to allow for less decisive learning experiences in which the agent's probabilities across a partition $\{E_1, E_2, ...\}$ change to $\{C_{new}(E_1), C_{new}(E_2), ...\}$, where none of these values need be 0 or 1:

$$C_{new}(X) = \sum_i C(X|E_i)C_{new}(E_i).$$

To summarise: the Jeffrey-style Bayesian claims that rationality requires one to assign propositions degrees of belief in conformity with the probability calculus, to update these degrees of belief according to certain rules, to assign propositions degrees of desirability subject to certain further constraints, and to act so as to maximise expected utility. This theory of rationality, as described so far, is perfectly compatible with Humean doctrine. Incompatible with it is a *further* constraint that codifies the way in which certain beliefs (or degrees thereof) rationally require certain desires (or degrees thereof).

---

[1] Strictly speaking there is no such thing as *the* value function of an agent: the value function is only unique up to fractional linear transformations.

[2] *Causal* decision theory replaces them with weights that measure the causal relevance of A to A_i. The lore has it that the two kinds of decision theory typically agree, only diverging on 'Newcomb problem' cases in which an action is evidence for some desired state of the world obtaining without in any way causing it. In such cases, a number of authors advocate the use of causal decision theory—e.g., David Lewis [1981; reprinted in Lewis 1986]. In the original Newcomb problem, the action is 'one-boxing' and the desired state of the world is 'the opaque box contains a million dollars'. Evidential decision theory apparently recommends one-boxing, while causal decision theory recommends two-boxing. For some skepticism about the lore, see Hajek and Hall [1994].

1.2 The Desire-as-Belief Thesis: an aerial view

Lewis [1988] canvases this anti-Humean constraint, which he calls the 'Desire-as-Belief Thesis'. The idea is that a rational agent desires something exactly to the extent that he or she believes it is good. Representing rational degrees of desire by a value function V and rational degrees of belief by a probability function C, Desire-as-Belief constrains which V's can co-exist with which C's. It requires that, corresponding to each proposition A, there is another proposition A^o such that

(DAB) $V(A) = C(A^o)$.

A natural interpretation of 'A^o' is that 'A is good', 'A is right', or something like that, although Desire-as-Belief itself is non-committal on this.[3]

Lewis shows that the Bayesian theory of rationality, combined with this Desire-as-Belief Thesis, *overconstrains* the agent. Anyone who adheres to Desire-as-Belief is unable to change their mind according to the Bayesian rules for rational revision or updating, and is thus epistemically paralysed. Something has to give. Lewis argues that rejecting the Bayesian theory of rationality is not an option; so it is Desire-as-Belief that must go. Collins [1988], Arló-Costa, Collins and Levi [1995], Byrne and Hájek [1996], and Lewis in his sequel paper [1996] provide further results against Desire-as-Belief that relax or modify various assumptions of Lewis' original refutation. It seems, then, that Desire-as-Belief is dead.

But killing off Desire-as-Belief is one thing, killing off anti-Humeanism another. Can we suitably modify the anti-Humean thesis so that the Bayesian agent is *not* overconstrained? There are two desiderata here: our modified thesis must be compatible with Bayesian decision theory, and it must capture a genuinely anti-Humean thesis. We will explore variants of Desire-as-Belief that can live peaceably with Bayesianism. However, before we can do that, we need to look at Desire-as-Belief, and the arguments against it, in more detail.

---

[3] Lewis himself is a causal decision theorist [1981]. Why, then, doesn't he formulate the Desire-as-Belief thesis in terms of his preferred decision theory? We want a measure of the *desirability* of states of affairs, not a measure of the choice-worthiness of the actions that may or may not bring them about. Lewis prefers causal decision theory as a theory of choice-worthiness of actions, and thus turns to it for guidance as to what one should *do*. However, he still thinks that evidential decision theory is an adequate theory of *desirability*, and that is what is at issue in this version of anti-Humeanism. Winning a million dollars is highly *desirable*, even if it's not something that an agent with two-boxing tendencies can bring about. And the anti-Humean thesis equates how *desirable* something is (as opposed to how choice-worthy it might be) with the probability that it is good. (Cf. [Lewis 1996: 304].)

1.3 The Desire-as-Belief Thesis: a view from the trenches

(DAB) has four unbound variables: V, A, C and A^o. In order to make a genuine statement, we must quantify over them. Here is how Lewis [1996: 308] states the Desire-as-Belief Thesis:

'there is a certain function (call it the *'halo' function*) that assigns to any proposition A a proposition A^o ('A-halo') such that, necessarily, for any credence distribution C,

(DAB) $V(A) = C(A^o)$.'

Quantifying over the halo function, or the haloed proposition that it assigns to each proposition, has the same effect. So we could state the thesis equivalently in terms of a quantification over propositions. Lewis calls both the equation and the thesis that quantifies over it 'DAB', but we find it useful to distinguish the two. He quantifies over the credence function C, but not over the value function V. However, since standard decision theory derives from an agent's set of preferences *both* a probability function *and* a value function, we should think of them as a pair, <C, V>, and we should quantify over that pair. Lewis' use of the word 'necessarily' appears to be redundant, for the subsequent quantification over *all* credence functions already suggests a necessary connection: if *any* credence function (be it actual or merely possible) conforms to (DAB), then it seems that the violation of (DAB) is impossible. We can capture the Desire-as-Belief Thesis, then, in the following quantified formula:

(Desire-as-Belief) $\forall A\ \exists A^o\ \forall \langle C, V\rangle\ V(A) = C(A^o)$.

With all the quantifiers in place, let us be clear about the domains of quantification. There are no restrictions explicitly stated, and none tacitly understood, so it is tempting to read Desire-as-Belief as:

For each proposition A, there is a proposition A^o, such that:
for each probability function/value function pair <C, V>,
$V(A) = C(A^o)$.

This is clearly false as it stands. Value functions can take values outside [0, 1], whereas probability functions cannot; so there will be values V(A) that cannot be matched by any probability value. But there is some arbitrariness in how desirabilities are represented, and value functions can be rescaled to lie within the [0, 1] interval, provided they are bounded (cf. footnote 2). So the third quantifier in Desire-as-Belief should be understood to range just over <C, V> pairs for which V is so bounded.[4]

[4] In virtue of this tolerance of rescaling, there is some arbitrariness in the probability function as well as in the value function: there are various C's in the <C, V> pairs that represent a given agent, all equally well. But in that case the truth of Desire-as-Belief, if it is a truth, will be partly an artifact of the representational

Desire-as-Belief is the thesis that Lewis offers the anti-Humean and then refutes. Unlike some authors (e.g., Broome [1991]) we have no quarrel with Lewis in his representation of the thesis as anti-Humean. But we hasten to add that the Desire-as-Belief Thesis is only *one* anti-Humean position, and Lewis himself is careful to note that he 'shall uphold Humeanism against *one sort of opponent'* [1988: 323, our italics]. The order of quantifiers commits the thesis to the 'haloed' propositions being fixed once and for all. Thus, for a given A, we have a single A^o, etched in stone, irrespective of C and V—thus, irrespective of the agent who may be contemplating it. This suggests that a variant of Desire-as-Belief that upholds its anti-Humean spirit may be close at hand, as we shall soon see. We will soon explore whether it escapes Lewis's negative results. But first we need to get a sense of what those results are and how they are reached.

1.4 Lewis' Anti-Desire-as-Belief Results

Lewis [1988] presents the first result against Desire-as-Belief. Begin with a proposition A and a <C, V> pair for which (DAB) holds. He shows that apart from trivial cases, we can update by Jeffrey conditioning to a new pair <C', V'> for which (DAB) no longer holds. The proof is algebraic, and rather sophisticated: the probability of A^o responds to the Jeffrey shift according to a certain linear function, whereas the desirability of A responds according to a quotient of such functions. Thus, the shift breaks the equality between the desirability of A and the probability of A^o.

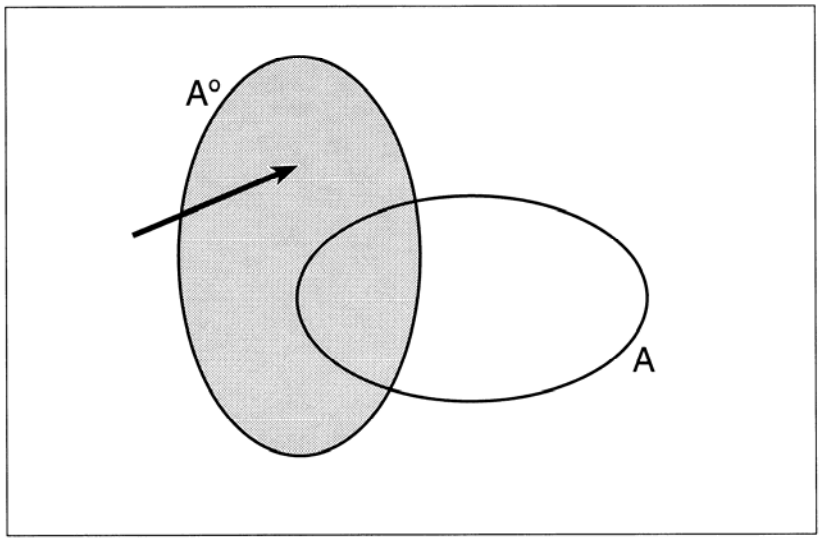
We do not dispute this result. But as Lewis himself says in his subsequent paper, it is 'needlessly complicated' [1996: 308]. Indeed, he offers there what he regards as a simpler refutation. He begins with this observation:

> DAB can be equivalently restated as a pair of equations: necessarily, for any A and C,
>
> (DACB) $V(A) = C(A^o|A)$
> (IND) $C(A^o|A) = C(A^o)$.
>
> To derive DACB, we recall that DAB is supposed to continue to hold under redistributions of credence, and we redistribute by conditionalizing on A [ibid. 308-9].

---

scheme (much as the fact that the freezing point of water is 0 degrees is an artifact of such a scheme). Lewis does intend Desire as Belief to capture a necessary connection between desires and beliefs, and says, '[i]nstead of speaking as I do of desires necessarily connected to beliefs, you might prefer to speak of beliefs that function as if they were desires; or of states that occupy a double role, being at once beliefs and desires. I take these descriptions to be equivalent' [1996: 308]. The necessity captured by Desire as Belief, then, apparently resides in its presumption of equality between V(A) and $C(A^o)$ on *one* way of making the arbitrary choice of <C, V> among the various possible choices. It should be remembered, however, that for a given agent there are equally valid representations for which the equality does not hold.

He then argues that (IND) is the culprit: given certain assumptions about C(A) and C(A^o|A), various redistributions of credence will make (IND) go from true to false. For example, moving probability from ¬A^o & ¬A to A^o & ¬A (as indicated by the arrow in the following figure) will increase C(A^o) while keeping C(A^o|A) constant.



Indeed, even weakened versions of Desire-as-Belief that replace the strict equality in (DAB) with something more anodyne—say, near-equality, or mere proportionality—will provide no refuge. For the sorts of cleavages that Lewis envisages between the two sides of (IND), and hence the two sides of (DAB), can be quite dramatic, and they can drive apart the two sides in both directions.

How worried should an anti-Humean be by all of this? Oddie believes that if Lewis' [1988] result is sound — he argues that it is not — 'it gets as close to being a reductio of realism about value as any argument could be' [1994: 452]. We do not agree. While we endorse Lewis' results, and more besides, we believe that there are important anti-Humean positions that can reasonably be called 'realist' and that are untroubled by them.[5] In the next section we investigate an important loophole in the results that anti-Humeans might try to exploit. The word 'loophole' should not suggest some mere technicality of no independent interest, the philosophical analogue

[5] The views identified later as immune to the anti-Desire-as-Belief proofs count as realist in the sense of representing evaluations as truth-conditional—indeed, so as to render *true* many evaluations.

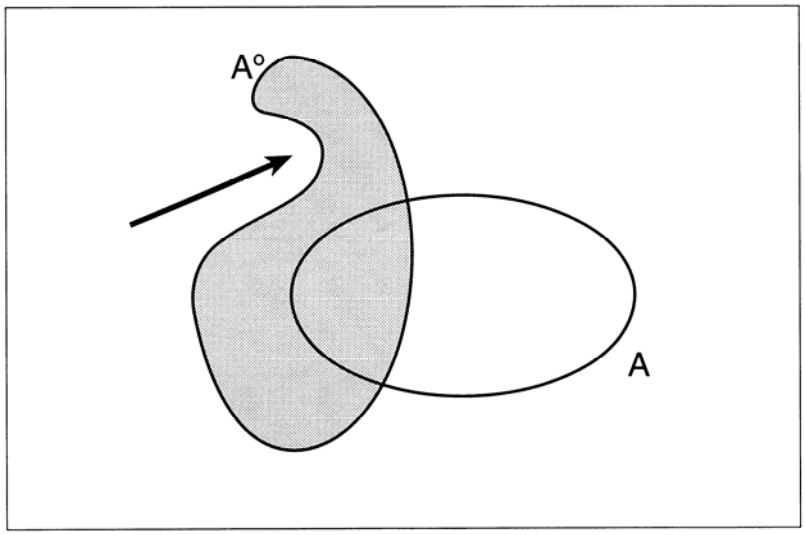
of an unintended gap in the taxation code. For as we will see, this loophole can exploited without ad hocery by philosophical positions with distinguished pedigrees. In the final section we look at the way in which indexicalist theories of 'good' and consequentialist and non-consequentialist theories of 'right' do exploit the loophole that we identify.

2. A loophole in the anti-Desire-as-Belief results

Focus on the 'E A' order of the quantifiers in the formulation of the Desire-as-Belief Thesis: 'there exists a halo function, such that for each <C, V> …'. As we have observed, this fixes the halo function, and thus the haloed-propositions, once and for all, irrespective of the agent who is contemplating them. The trick, then, is to indexicalise the halo function, and thus the haloed-propositions, to the agent. Specifically, we allow A^o to be a function *both* of A *and* of the <C, V> pair in whose scope it appears. We can make that salient by indexing A^o by the pair: $A_{<C,V>}{}^o$. Thus, we are led to a variant of Desire-as-Belief that reverses its quantifiers: for each <C, V>, there exists a halo function… Equivalently, making explicit the haloed propositions that the various halo functions assign:

(Indexical Desire-as-Belief) $\forall <C, V> \forall A \exists A^o \; V(A) = C(A^o)$.

Indexical Desire-as-Belief evades all the negative results that we have discussed or mentioned, since they assumed that A^o remained fixed throughout redistributions of credence. Lewis' first result holds A^o fixed through an instance of Jeffrey conditionalization that cleaves apart the two sides of (DAB). His second result holds A^o fixed through a redistribution of probability that cleaves apart $C(A^o|A)$ and $C(A^o)$, and hence the two sides of (DAB). But if instead we allow the identity of A^o to change as the distribution of probability changes, we have no guarantee that the required cleavages will take place. For example, in the second result, for all we know the haloed proposition assigned to A will shift in just the right way to compensate for the shift in probability, maintaining (DAB) throughout. Diagrammatically, as probability is shifted, threatening to increase $C(A^o)$, the boundary of A^o could move to avoid it, keeping both $C(A^o)$ and $C(A^o|A)$ constant after all:



To be sure, the probability of the *former* A° cannot be identified with the new V(A), but on the indexical view of the halo, that is no longer the proposition at issue. Similar points can be made about the negative results due to other authors. Each of them refutes a version of Desire-as-Belief (quantitative in some cases, as Lewis' version is, qualitative in others) in which the quantification over A° *precedes* the quantification over the doxastic state (be it a probability function, or a qualitative counterpart). So A° is already fixed before we get to the doxastic state, and so remains the same throughout changes in that state. However, if A° is a moving target, the results are blocked.

So we have found a loophole in the results. But can the anti-Humean exploit it?

Lewis, as usual, sees it coming, and labels our indexicality thesis 'inconstancy'. He writes:

> if … we only require that for any given credence and value functions *C* and *V*, there exists a halo function chosen ad hoc to satisfy the desired conditions with respect to that particular pair of *C* and *V*, then our task is almost trivial. We need only require that *C* and *V* have the right ranges of values: for any *A* there exists some *B* such that *V(A) = C(B)* [1996: 312].[6]

[6] Presumably the requirement that Lewis has in mind is that the range of V be a subset of the range of C. But we suggest, on the contrary, that our task might be quite *non*-trivial, for this requirement on the ranges of C and V might be non-trivial. For example, if C has a finite range (which plausibly is the case for any human agent), then some values of V might 'fall between the cracks', finding no match among the values of C. Indeed, unless the range of C is the entire [0, 1] interval, there is some danger that there will be a degree of desire that finds no partner among the degrees of belief. In that case, desire goes beyond belief once again.

But non-trivial though the task may be, that does not mean that it is impossible, or even unreasonably difficult. Or to *make* it almost trivial, we might adopt a hybrid strategy: first restrict the scope of Desire-as-Belief to just those <C, V> pairs for which C's range is [0, 1], and then indexicalise the halo-function.

Yet he is unfazed: 'It's too easy, and no anti-Humean should celebrate such an easy victory'. Why not? Anti-Humeanism understood in this way may be a trivial truth, but turning Lewis' own words against him, 'a trivial truth is still a truth' [1988: 323]. And an easy victory is still a victory.

Lewis is unfazed because he would not regard it as anti-Humeanism. He says: 'there is nothing that should make us want to say that A^o is the proposition that *A* is objectively good' [1996: 312]. Note that we have moved some distance from the original terms of the debate: whether beliefs are necessarily connected to desires. Originally, no mention was made of *objective goodness*. In fact, Lewis moves even further, suggesting in a footnote [ibid. 312] that the real anti-Humean goal is to deliver *objective ethics*. But recall that Lewis claimed at the outset only to codify and to refute the position of 'one sort of opponent' to Humeanism. Now he seems to be giving the anti-Humean rather grander ambitions. In other words, he appears to be strengthening the position of his 1988 paper: Desire-as-Belief may only be *an* anti-Humean position, but there is a suggestion that it is the only serious contender. Another contender, Indexical Desire-as-Belief, may be tenable (trivially so, perhaps), but he does not regard it as *serious* because it falls short of the grander ambitions.

There are two issues here that we should distinguish: whether or not there is a *non-trivial* version of Indexical Desire-as-Belief, and whether or not it delivers *objective ethics*. Showing the truth of Indexical Desire-as-Belief, with no constraint on the halo-function, is an almost trivial task according to Lewis. Showing the truth of Desire-as-Belief, with a halo-function so constrained as to be fixed for all agents, is as non-trivial as tasks get: downright impossible. But it remains a live possibility that there is a tenable version of Indexical Desire-as-Belief that tells us something substantive about the halo-function. Indeed, it might even tell us something substantive about a halo-function that will underwrite an ethical theory—perhaps even an *objectivist* ethical theory, as we will see. Certainly, none of the negative results that we have seen rules out this possibility. For an ethical theory, even an objectivist ethical theory, can be indexicalist in the way that we have identified.

3. Exploiting the indexicality loophole

There are ethical theories — more generally, accounts of evaluation — under which something like a halo-operator, or a halo-predicate, is given an anti-Humean, desire-related role and yet cast in a manner that would exploit the indexicality loophole. We will give a quick

overview of some of them in this final section. We do not argue for any of the views canvassed, being content to identify ways of thinking about evaluation and ethics that would escape the anti-Desire-as-Belief results and yet that have considerable currency in contemporary moral thought.

If there is to be a plausible candidate for a predicate that plays the halo role, then it must direct us to a property such that to believe that a certain scenario or a certain action has that property is to have a corresponding degree of desire for the scenario or action in question. There is a well-known family of views in the ethical literature — 'internalist' views, as they are often called (see [Darwall et al. 1992])— that in one version does posit this sort of internal connection between evaluative belief and desire. They may say that it is logically impossible to have the belief without the desire, or that the rational agent who has the belief will have the desire, or that if the desire is held with full understanding, or in canonical mode, or the like, then it will be attended by the desire. They may not strictly equate the degree of the desire with the degree of the belief, as Desire-as-Belief does as it stands; they may replace strict equality with near-equality or proportionality. They will still run into apparent conflict, however, with Lewis' second anti-Desire-as-Belief proof since, as we saw in §1.4, that proof will work against such weakened versions of Desire-as-Belief too. In any case we shall show that even if certain internalist views are not weakened in that way, they still have a means of escape from the negative results via the indexicality loophole.

The question before us, then, is whether any extant, internalist theories of evaluation construe an evaluative predicate in a manner that exploits the indexicality loophole. Do any of them construe an evaluative predicate in a manner that allows the semantic value of the predicate to vary with a change in the agent's credence or value distribution: with a variation in the <C,V> pair by which the agent at a given time is characterised? We look at two broad possibilities, one associated with the predicate 'good', the other with the predicate 'right'.

3.1 Indexicalist theories of 'good'

The most obvious, ordinary predicate that might be thought to play the halo role is 'good'. So is there any way of thinking about goodness that might connect with the loophole?

One way is the theory of moral evaluation that G. E. Moore [1911] called *subjectivism*. It holds that when someone says that a prospect is good, then that utterance expresses the belief that the speaker has an attitude of approval towards the prospect. Thus the content of the sentence 'It is good that prospect P obtains' is 'I have an attitude of approval towards P'. A recent relative of

such a subjectivist theory is the speaker-relativism defended by James Dreier [1990]. He argues that when one makes an evaluation of a prospect P, then the content is best represented as 'P accords with the relevant standards', where the relevant standards are fixed indexically and as we shall assume here may just be the standards that happen to be espoused by the particular speaker.

Both of these theories are indexicalist theories of goodness. A given sentence of the form 'P is good' will have a different truth-condition, according to these theories, in the mouths of different speakers. It will be true in John's mouth if and only if John has certain attitudes, it will be true in Mary's if and only if Mary has certain attitudes, and so on. The sentence will have the same *character* for both speakers, in David Kaplan's terminology, but it will have a different content in the mouth of each; the content will be finally fixed in each case by the nature of the speaker's attitudes.

While explicit subjectivism and speaker-relativism about 'good' may not be commonly espoused, the indexicalism that they exemplify is entailed, arguably, by the range of fashionable theories that are often called 'expressivist'. Those theories hold that to describe something as good is voluntarily and conventionally to express an attitude of approval towards it. But plausibly, therefore, the theories are committed to maintaining that speakers describe something as good when they register the presence of the attitude to be expressed — when they believe that they approve — and that conditions are right for communicating the existence of the attitude. (See [Jackson and Pettit 1998; 2003].) And that is indistinguishably close to being committed to the indexicalist view that to say something is good is to say — and no doubt also to show — that one approves of it.

Internalists about the predication of goodness who hold that that predication is explicitly or implicitly indexical will be able to exploit the indexicality loophole. Let the sentence 'A is good' serve for 'A^{o}'. Then an agent's degree of belief in the proposition expressed will correspond to his or her degree of desire for A. None of the anti-Desire-as-Belief proofs will get a purchase, for A^{o} is a shifting target. As we imagine moving probability around, we imagine changes in an agent's attitudes; these changes, in turn, may change the content or truth conditions of 'A^{o}'. Probability moves between different worlds, but this does not necessarily break the putative agreement between a degree of desire and a corresponding degree of belief.

Some anti-Humeans may not take much solace, however, from these observations. The indexicalist proposal has to face a difficulty of the kind raised by Lewis in another connection: the alleged agreement between probabilities of conditionals and conditional probabilities. Having

presented his first round of famous 'triviality results' against such agreement [1976; reprinted in 1986], Lewis considers and rejects an indexical account of conditionals according to which the content of a spoken conditional varies according to the probability function of its speaker.[7] He argues that conditionals must have a fixed interpretation across individuals, so that what one says in endorsing the claim 'If A, then B' is what another may deny or what one may later reject. 'Else how are disagreements about a conditional possible, or changes of mind?' [1986: 138]. That sort of difficulty arises, notoriously, for indexicalist theories of evaluation too. In saying that something is good I will be reporting my approval, and when you deny that it is good you will not be rejecting what I say but reporting your own disapproval; strictly there will be no disagreement, at least no disagreement on a matter of fact, between us.

We do not think that the problem raised against indexicalism about 'good' is decisive. Suppose that we assume — as do all relevant speakers — that we are isomorphically minded, and that if we differ in the attitudes we hold, then something is going wrong on one or both sides: our attitudes are not getting to be formed under the right inputs, or according to the right processes, or whatever. In that case there will be clear utility in conversing, even conversing in indexicalised language, about what is good. And so equally there will be a clear sense in which we disagree if I stick to the view that A is good and you to the view that it is not.

But this is not the place to defend indexicalism about 'good'. Our purpose is only to notice that here is a reasonably well-established position in which anti-Humeans may take refuge. The anti-Desire-as-Belief proofs are impressive but they are not capable of dislodging indexicalist-cum-internalist views of goodness, or indeed of any value predicate.

3.2 Indexicalist theories of 'right'

One of the central predicates in the evaluative lexicon is 'right', where rightness is predicated of ways things may be that agents are in a position to bring about. These ways are a proper subclass of what we have been calling 'propositions'. We will call these particular propositions 'options'.

No agent can deem an option the right one to take and yet fail to take it without owing us an explanation as to what went amiss. For that reason, there is often assumed to be an intimate

[7] Van Fraassen [1976] suggests such an account in defense of the alleged agreement against Lewis' triviality results. This debate between Lewis and van Fraassen was an important source of inspiration for our considering the analogous indexicality idea here.

connection between predicating rightness of an option one faces — and thereby, presumptively, expressing the belief that it is right — and forming a corresponding desire. And that is the sort of connection that gives support to a broadly internalist, anti-Humean picture. The question, then, is whether room is made within any standard theories of rightness for exploiting the indexicality loophole in the anti-Desire-as-Belief proofs.

Consequentialism

There are two main sorts of theories of rightness, consequentialist and non-consequentialist. The consequentialist theory holds that whether an option is right is determined wholly by whether it promises to promote the good: in what many authors take to be the most plausible version, whether the option maximises expected goodness. Goodness is taken by consequentialism in a neutral, non-indexical sense, so that the loophole that we explored earlier does not remain open. This non-indexically conceived goodness may be taken to be a universal property such as sentient happiness, or something much more parochial, like the prosperity of a particular country or culture or species or individual. The typical consequentialist may require that expected happiness be maximised — this is the utilitarian version of the doctrine — but someone who is egomaniacal enough to think that *their* expected happiness should be maximised — maximised by everyone, not just by them alone — will count as a consequentialist too.

Is there any way in which a consequentialist theory of rightness might make room for the anti-Humean, offering a means of exploiting the indexicality loophole? It turns out, on a little reflection, that there are two ways in which it may do this. What it is right for someone to do — what maximises expected goodness among the options available to that person — may be taken to be what maximises expected goodness-according-to-the-agent. Or it may be taken to be what maximises expected-according-to-the-agent goodness. These steps may be taken separately or together: the goodness and the expectation may be relativised independently or at the same time. In the first step, the consequentialist says that the right option for any agent is that which maximises the expectation of subjective goodness; in the second that the right option is that which maximises the subjective expectation of goodness: or, more strictly, subjectively expected goodness.

Under either of these moves the content of a sentence like ‘A-ing is the right option for agent x’ will vary with the identity of x. More particularly, the content of the sentence will vary depending on the credence function C_x of x: depending, intuitively, on the person’s beliefs about what things are good, or about how probable it is that good outcomes will follow on given

choices. 'Right' will be more perspicuously written as 'right_{Cx}'; its semantic value will depend on the particular credences that x holds.

This means, then, that the indexicality loophole is going to be accessible for anti-Humean purposes. Consequentialists can say that x's degree of belief in the rightness of an option is linked with x's degree of desire for that option, and they need not worry about the anti-Desire-as-Belief proofs. For the movements of x's probabilities that would supposedly break that linkage may alter the content of sentences about what it is right for x to do. Those sentences will be perspicuously represented as claims about what is right_{Cx} and as we imagine shifts in the credences of x, the semantic value of 'right_{Cx}' may move at the same time.

How plausible is it that the consequentialist theory of right should relativise the predicate to agents in this way? It may not be plausible to think that consequentialists link what is right with the maximisation of subjective goodness only; by most accounts, agents may make a mistake about what is good and so about what is right. But there are arguments in the literature for why the consequentialist theory of right should associate rightness with the maximisation of subjectively expected goodness rather than with goodness that is in some sense objectively expected: say, expected according to perfectly informed subjects. These suggest that for any agent x, the right option is to maximise expected-by-x goodness, not expected-by-perfectly-informed-subjects goodness. See [Jackson 1991] and, for an opposing viewpoint, [Menzies and Oddie 1992].

We mentioned above that indexicalist positions about 'good' are often faulted for not representing people in ethical debate as disagreeing about any matter of fact. In differing on whether something is good, under an indexicalist view of goodness, people report quite consistent facts: A reports that *she* approves, B reports that *he* doesn't approve. Does a similar difficulty beset the consequentialist view of rightness? We don't think so. We said that there was a possible way around the earlier objection: specifically, that if speakers share an assumption that they are isomorphically minded, then they may treat differences in attitudes of approval as signaling that one or both of them are forming their attitudes inappropriately and that there is room for discussion. This response is particularly plausible in the present case. Suppose that I report that by my lights a certain course of action is right, maximising subjectively expected goodness, and you think that it is not right (for someone in my situation) because you have different credences, even though you value relevant things the same way. Then by ordinary criteria there is still room to argue and debate. For even if we think that differences in credences do not involve either of us in denying what the other holds, we do assume in common that such differences may signal that

one or both of us are lacking information of the kind that we treat as relevant in forming credences. Thus we cannot be indifferent to the fact that we have different credences in some proposition; there is a sense in which we disagree and may learn from one another.

Non-consequentialism

The non-consequentialist theory of rightness can be formulated in a number of ways. One that marks the distinction from consequentialism very elegantly holds that whether an option is right is determined in part by whether it instantiates the neutral, non-indexically represented good: whether it itself — or perhaps a relationship it establishes for the agent — instantiates the properties that are thought to make something good. (See [Pettit 1997, 2000].)

Suppose that in a given case the only property relevant to goodness is non-violence. Where consequentialists would say that the right option in any choice is that which maximises expected non-violence, non-consequentialists would deny this if the option that maximises non-violence overall itself involves a violent act – the war that promises to end all wars, say. They will require that a right option go some distance towards instantiating goodness. And what they say of non-violence in this simplified example, they will equally say about other properties that they think of as good-makers. Suppose that promise-keeping or honesty or kindness is good, for example, and now consider a case where only one such property is relevant; this is a simplification that makes the doctrine easier to state and defend. Non-consequentialists will say that the right option for all agents in such a situation is to ensure in some measure — at the extreme, to ensure only — that the property is instantiated by them in their actions or relationships. On some accounts it must be instantiated by them at the time in question, on others it must be instantiated by them over the course of their lives; the first approach enjoins instantiation by a certain time-slice of the person, the second by the person as he or she endures through time.

The non-consequentialist theory of rightness offers the same opportunity as the consequentialist theory for anti-Humeans to resist the anti-Desire-as-Belief proofs. No matter how non-consequentialism is formulated, it is bound to engage credences in the same manner as consequentialism. This is obvious insofar as the goodness invoked in the theory of the right is taken to be goodness-according-to-the-agent. But it turns out that non-consequentialism also involves credences in the distinctive manner in which consequentialism involves them.

Consider the extreme form of non-consequentialism that says that for any good-making property, the important thing for agents to do is to instantiate that property themselves, whatever the consequences for the overall realisation of the property: the important thing is to keep their

own hands clean in regard to the property, as an unfriendly comment might have it. Even this form of non-consequentialism is co-extensive with what we might describe as an egocentric consequentialism. Plausibly, it will equate what is right for agents in any choice with what maximises their own expected instantiation of the property: their expected instantiation of the property at the time of action or over the course of their lifetime.[8]

But once we see that the non-consequentialist theory of the right can be cast in this form, it should be obvious that the same arguments that make consequentialism safe for anti-Humeans — safe from the threat of the anti-Desire-as-Belief proofs — can make non-consequentialism safe too. Anti-Humeans are free to embrace non-consequentialism and to argue that what it means to say that an option is right for an agent x is that the option maximises the expected-by-x instantiation by x of the good, whether at the time of action or over x's lifetime. And once they say that, they will have turned the required trick; they will have relativised the predicate 'right' in such a way that it can be plausibly rewritten as 'right_{Cx}'.

Non-consequentialists may say that because they are concerned with the instantiation of a property, not with its overall realisation, they do not have to invoke expectations in the same manner as consequentialists. But this is simply false. If the theory is that the agent should maximise the instantiation of the property through time, then there may be just as much uncertainty about what will do this as there is about what will maximise the overall realisation of the property. And even if the theory bears only on what the time-slice of the agent should instantiate, uncertainty and expectation will still be relevant. Suppose that non-consequentialists enjoin not harming others, for example, at whatever loss in good consequences overall. That injunction is still going to leave open the question of how likely it is that any option chosen in accord with this principle really is going to leave others unharmed.

Many a slip twixt cup and lip; and equally many a slip between what an agent sets out to instantiate at a time and what may actually transpire. And in any case there are going to be many properties for which any adherent of time-relativised non-consequentialism will have to acknowledge a place for uncertainty. Take the property of promise-keeping, for example. The theory that says that an agent's time-slice ought to keep its promises will have to acknowledge

---

[8] The approach is not likely to identify the right in such a way that it is only actual consequences — actual instantiation effects — that count. Like an actual-consequences version of consequentialism, this would have the counter-intuitive result that an action may count as right, though done out of culpable recklessness or malice. And so the approach will plausibly equate what is right for agents in any choice with what maximises their expected instantiation of the relevant good-making property at the time of action, or over their lifetime.

that what are to be kept, plausibly, are the promises-according-to-current-memory of the agent; and any such reference to memory immediately introduces credence.

'Right' and 'rational'

The foregoing arguments identify reasons why 'right' under a consequentialist or non-consequentialist theory should be construed as 'right$_{Cx}$'. And they point thereby to a possible connection between belief and desire that is immune to the anti-Desire-as-Belief proofs: a connection between an agent's degree of belief that an option is right$_{Cx}$ and the degree of the agent's desire for the option.

But it is worth noting in conclusion that if those arguments are sound then equally, and perhaps less controversially, they point us towards a connection that some theorists may wish to defend between a rational agent's credence that an option is *rational* in the Bayesian sense — assuming that agents can have such credences — and the agent's strength of desire for that option. Agents will believe that an option is rational in this sense if and only if they believe that it maximises their subjectively expected utility. And that means that for any agent x the rational option can be perspicuously represented as the rational$_{Cx}$ option, in which case immunity to the anti-Desire-as-Belief proofs is immediately available.

The analogy between 'right' and 'rational' is worth noting, because it may serve to reassure readers that there is nothing strange going on in the preceding discussion of how a belief in the rightness of an option may be tied to a degree of desire for that option. There is nothing strange in the idea that there might be such a connection in the case of rationality, since the belief that an option is 'rational' — at least 'rational' in the Bayesian sense — is so clearly indexical in character and so clearly not the sort of predicate that Lewis had in his sights. We hope that once the analogy between 'right' and 'rational' is noted, the line of argument run with rightness will not seem strange or surprising. Recall that it was Lewis [1996] who moved the Desire-as-Belief debate into the realm of ethics, when the issue of whether or not desires and beliefs are necessarily connected could equally have been cast as one about rationality. Without espousing here any of the indexicalising theories surveyed, it should be obvious that for anyone intent on resisting the anti-Desire-as-Belief results, they are there for the taking.

The analogy between 'right' and 'rational' is worth noting for another reason too. The fact that the mode of resistance to the anti-Desire-as-Belief results available in the cases of 'right', or 'good', parallels a mode of resistance equally available in the case of 'rational' — in a subjective, 'Humean' sense — indicates that the resistance may not offend the spirit of Hume. When he

insisted on the distinctness of beliefs and desires, even their distinctness in the rational agent, he surely had in mind the distinctness of desires from beliefs in features of *the desire-independent world*, not from beliefs in matters related to *the formation of those desires themselves*. The anti-Desire-as-Belief results that Lewis articulated and inspired can be seen as supportive of that core Humean position. But as Lewis himself said in a different context [preface to 1983: x], 'Philosophical theories are never refuted conclusively'. And so it is with anti-Humeanism.[9]

California Institute of Technology

Princeton University

References

Arló Costa, Horacio, John Collins and Isaac Levi 1995. Desire-as-Belief Implies Opinionation or Indifference, *Analysis* 55: 2-5.

Broome, John 1991. Desire, Belief and Expectation, *Mind* 100: 265-267.

Byrne, Alex and Alan Hájek 1996. David Hume and Decision Theory, *Mind*, 106:411-428.

Collins, John 1988. Belief, Desire, and Revision, *Mind* 97: 333-342.

Darwall, S., A. Gibbard, et al. 1992. Towards Fin de Siècle Ethics: Some Trends, *Philosophical Review* 101: 115-89.

Dreier, James 1990. Internalism and Speaker Relativism, *Ethics* 101:6-26.

Hájek, Alan and Ned Hall 1994. The Hypothesis of the Conditional Construal of Conditional Probability, in *Probability and Conditionals*, eds. Ellery Eells and Brian Skyrms, Cambridge: Cambridge University Press: 75-111.

[9] We thank especially Alex Byrne, John Collins, Ned Hall, Frank Jackson, David Lewis, Michael Smith, Peter Vranas, and an anonymous referee for the *Australasian Journal of Philosophy* for helpful comments.

Jackson, F. 1991. Decision-Theoretic Consequentialism and the Nearest and Dearest Objection, *Ethics* 95: 461-82.

Jackson, F. and P. Pettit 1998. A Question for Expressivists, *Analysis* 58: 239-51.

Jackson, F. and P. Pettit 2003. Locke, Expressivism, Conditionals, *Analysis* 63: 86-92.

Jeffrey, Richard 1983. *The Logic of Decision*, Chicago: Chicago University Press, 2nd ed.

Jeffrey, Richard 1992. *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.

Lewis, David 1976. Probabilities of Conditionals and Conditional Probabilities, *Philosophical Review* 85: 297-315. Reprinted in Lewis 1986.

Lewis, David 1981. Causal Decision Theory, *Australasian Journal of Philosophy* 59: 5-30. Reprinted in Lewis 1986.

Lewis, David 1983. *Philosophical Papers* Vol. I, Oxford: Oxford University Press.

Lewis, David 1986. *Philosophical Papers* Vol. II, Oxford: Oxford University Press.

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

Lewis, David 1989. Dispositional Theories of Value, Part II of a Symposium in *Proceedings of the Aristotelian Society* Supp. Vol.:113-137.

Lewis, David 1996. Desire as Belief II, *Mind* 105: 303-313.

Menzies, P. and G. Oddie 1992. An Objectivist's Guide to Subjective Value, *Ethics* 102 (3): 512-34.

Moore, G. E. 1911. *Ethics,* Oxford: Oxford University Press.

Oddie, Graham 1994. Harmony, Purity, Truth, *Mind* 103:451-72.

Pettit, Philip 1997. The Consequentialist Perspective, in *Three Theories of Ethics: A Debate,* M.Baron, P.Pettit, M.Slote, Oxford: Blackwell.

Pettit, Philip 2000. Non-consequentialism and Universalisability, *Philosophical Quarterly*, 50:175-90.

Van Fraassen, Bas 1976. Probabilities of Conditionals, in *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I, eds. W. L. Harper and C. Hooker, Dordrecht: Reidel: 261-301.