Cluelessness

Hilary Greaves, University of Oxford

Abstract. Decisions, whether moral or prudential, should be guided at least in part by considerations of the consequences that would result from the various available actions. For any given action, however, the majority of its consequences are unpredictable at the time of decision. Many have worried that this leaves us, in some important sense, *clueless.*

In this paper, I distinguish between 'simple' and 'complex' possible sources of cluelessness. In terms of this taxonomy, the majority of the existing literature on cluelessness focusses on the simple sources. I argue, contra James Lenman in particular, that *these* would-be sources of cluelessness are unproblematic, on the grounds that indifference-based reasoning is far less problematic than Lenman (along with many others) supposes.

However, there does seem to be a genuine phenomenon of cluelessness associated with the 'complex' sources; here, indifference-based reasoning is inapplicable by anyone's lights. This 'complex problem of cluelessness' is vivid and pressing, in particular, in the context of Effective Altruism. This motivates a more thorough examination of the precise nature of cluelessness, and the precise source of the associated phenomenology of discomfort in forced-choice situations. The latter parts of the paper make some initial explorations in those directions.

1. Cluelessness about objective betterness

The cluelessness worry. Assume determinism.[1] Then, for any given (sufficiently precisely described) act A, there is a fact of the matter about which possible world would be realised – what the future course of history would be – if I performed A. Some acts would lead to better consequences (that is, better future histories) than others. Given a pair of alternative actions A_1, A_2, let us say that

(OB: Criterion of objective c-betterness) A_1 is *objectively c-better* than A_2 iff the consequences of A_1 are better than those of A_2.

It is obvious that we can never be *absolutely certain*, for any given pair of acts A_1, A_2, of whether or not A_1 is objectively c-better than A_2. This in itself would be neither problematic nor surprising: there is very little in life, if anything, of which we can be absolutely certain. Some have argued, however, for the following further claim:

> **(CW_o: Cluelessness Worry regarding objective c-betterness)** We can never have even the faintest idea, for any given pair of acts (A_1, A_2), whether or not A_1 is objectively c-better than A_2.

This 'cluelessness worry' has at least some more claim to be troubling. (I return in section 2 to the question of whether it *is* troubling; and in section 6 to the question of what exactly "can't have the faintest idea" means.) This is most obvious in the case of consequentialism. For if (CW_o) is correct,

[1] Relaxing this assumption would complicate some parts of the discussion, but not in ways that are ultimately relevant to the issues in this paper.

and if in addition (as consequentialism holds) the moral status of an action is determined entirely by how it compares to alternative actions in terms of the goodness of its consequences, it seems to follow with particular clarity that we can never have even the faintest idea what the moral status of any given action is. But any plausible moral theory will agree that considerations of consequence-goodness are at least morally relevant – that they should be taken serious account of both in moral decision-making and in moral evaluation, as at least one important factor. And this too seems impossible in practice if (CW_o) is correct.[2]

The argument for (CW_o) stems from the observation that the relevant consequences include *all* consequences of the actions in question, *throughout all time*. In attempting actually to take consequences into account in practice, we usually focus on those effects – let us call them 'foreseeable' effects – that we take ourselves to be able to foresee with a reasonable degree of confidence. (These may or may not be any intuitive sense 'direct' effects, and may or may not be close to the point of action in time and/or space.) And while we are arguably correct in thinking that we are justified in being reasonably confident in our predictions of *these* effects, any choice of one act A_1 over another A_2 inevitably has countless additional consequences that our calculation takes no account of. A butterfly flapping its wings in Texas may cause a hurricane in Bangladesh; so too may my telling a white lie, refraining from telling that lie, moving or not moving my hand; a hurricane will certainly affect which other butterflies flap their wings or which other agents move their hands in which ways; and so the effects will ripple down the millennia. Any conclusion, on the basis of the calculations that we have carried out, that one act is indeed objectively better another is justified only insofar as we are justified in assuming

> **(NR_o (Non-reversal for objective c-betterness))** The net effect of taking into account all of these additional effects *would not reverse* the judgment that we reach based on the foreseeable effects alone.

But is (NR_o) true?

Here are two *bad* arguments for (NR_o).

The 'ripples on a pond' postulate. First, one might think that while there are indeed non-zero effects, traceable to even the most trivial of one's actions, stretching down through the millennia, still the *magnitude* of any individual such effect typically decays with time. Further – letting $\Delta V \coloneqq V(A_1)-V(A_2)$ be the amount by which the goodness of the consequences of A_1 exceeds that of A_2 – one might think that this decay is sufficiently fast that by far the largest contribution to ΔV comes from the foreseeable effects, most of which latter are in practice temporally close to the point of action. Call this the *'ripples on a pond' postulate*. It is suggested (though not strongly advocated[3]) by Moore:

[2] Here I am in agreement with Smart (1973, p.34), Kagan (1998, p.63) and Mason (2004), each of whom initially raises the issue of cluelessness in the context of consequentialism, but then notes that in fact the problem affects a much wider class of moral theories. In contrast, many others appear to regard the problem as peculiar to consequentialism (including: Norcross (1990), Lenman (2000), Cowen (2006), Feldman (2006), Dorsey (2012), Burch-Brown (2014)).

[3] Unlike Smart, Moore for the most part confines himself to asserting the *necessity* of defending (NR_o), by the means suggested or otherwise, "if any of our judgments of right and wrong are to pretend to probability" (Moore, ibid., section 93).

> "As we proceed further and further from the time at which alternative actions are open to us, the events of which either action would be part cause become increasingly dependent on those other circumstances, which are the same, whichever action we adopt. The effects of any individual action seem, after a sufficient space of time, to be found only in trifling modifications spread over a very wide area, whereas its immediate effects consist in some prominent modification of a comparatively narrow area. Since, however, most of the things which have any great importance for good or evil are things of this prominent kind, there may be a probability that after a certain time all the effects of any particular action become so nearly indifferent, that any difference between their value and that of the effects of another action, is very unlikely to outweigh an obvious difference in the value of the immediate effects." (1903, §93)

Similarly Smart (who apparently *does* advocate it):

> "[W]e do not normally in practice need to consider very remote consequences, as these in the end rapidly approximate to zero like the furthermost ripples on a pond after a stone has been dropped into it." (1973, p.33)

The 'ripples on a pond' postulate, though, is not plausible. To see this most vividly, note that even our most trivial actions are very likely to have unforeseen *identity-affecting effects* (although the same points could be made without appeal to identity-affectingness). Suppose, for example, that I pause on my way home from work, in order to help an old lady across the road. As a result, both she and I are in any given place – any given position on the pavement for the remainder of our respective journeys home, for instance – at different times, at least for the remainder of that day. As a result, we advance or delay the journeys of countless others, if only by a few seconds, relative to the situation in which I had not helped her across the road; both we and they affect which further parties enjoy chance meetings with whom; and so forth. At least some of these others were destined to conceive a child on the day in question, and if so, even our trivial influences on their day will affect, if not *whether* they conceive, then at least *which particular child* they conceive (since a delay in sexual intercourse of even a few seconds is overwhelmingly likely to affect which particular sperm fertilises the egg).[4] But once my trivial decision has affected *that*, it equally counts as causally responsible for *everything the child in question does during his/her life* (i.e., for the differences between the actions/effects of this child vs. those that the alternative, in fact unconceived, child would have performed/had) – and of all the causal consequences of all *those* things, stretching down as they do through the millenia. These consequences are clearly not negligible: many or most of the things that one child does and that the alternative child would not have done (or vice versa) amount to greater differences in goodness than the intrinsic value of one old lady's receiving help across the road on one occasion. Nor is it at all likely that the *number* of identities my action affects in generation *r* will decrease as *r* increases; on the contrary, it will increase.

The cancellation postulate. Might one resurrect (NR_o) by arguing that although there are, for any choice of a given action A_1 over an alternative A_2, countless effects of significant size stretching arbitrarily far into the future, that nonetheless these unforeseeable effects are highly likely to cancel

[4] Lenman (2000) presses the point that at least many *morally important* actions, such as killings, abortions and procreative actions, are identity-affecting. Parfit has argued, as I do here, that the same is also true of less obviously identity-directed actions (1984, chapter 16).

one another out, and to do so to an arbitrarily high degree of precision as the time horizon stretches to infinity? If so, then their *combined* effect will be much smaller than the foreseeable effect, even if the effect of any *individual* unforeseeable consequence is comparable to that of the foreseen consequences. Call the postulate that these conditions do indeed obtain the *cancellation postulate*.

Unfortunately, the cancellation postulate is false. The theory of random walks tells us that while *some* degree of cancelling-out in such situations is all but certain, the combined effect of a large number *n* of probabilistically independent steps tends to grow with *n*, and in particular that it is highly unlikely to end up anywhere sufficiently close to zero.[5] This result is, on reflection, intuitively extremely plausible: the observation is that it is extremely unlikely, for instance, that the difference in net value between *everything this child does in his/her life* on the one hand and *everything the alternative child would have done in his/her life* on the other will just happen to be smaller than the intrinsic value of one old lady's receiving help across the road on one occasion, even if we pretend that each of a child's actions is probabilistically independent of each of the same child's other actions; and increasing the number of children involved will only exacerbate the problem.[6]

These arguments against possible defences of (NR_o) are equally reasons for thinking that (NR_o) in fact is true only in roughly 50% of cases. We are forced to conclude that (CW_o) is true, in the following sense: we can never be justified, for any given pair of acts, in having credence significantly greater than 50% that either is objectively c-better than the other.

2. Cluelessness about subjective c-betterness

The truth of (CW_o) would be troubling, however, only if it followed that there was *no* way for considerations of consequences to guide either decisions or evaluations; and (OB) is not the only possible route for that to happen. In fact, consequentialists in particular have long recognised both the availability and the indispensability of a second such possible route, viz. the appeal to a relation of *subjective* c-betterness among actions:

> **(SB: Criterion of subjective c-betterness)** Act A_1 is *subjectively c-better* than A_2 iff the expected value of the consequences of A_1 is higher than the expected value of the consequences of A_2 (where both expectation values are taken with respect to the agent's credences at the time of decision[7]).

[5] More precisely: A *one-dimensional simple symmetric random walk* is a series $\{S_n\}_{n=1,2,\ldots}$, where (1) for each n, $S_n = \sum_{j=1}^{n} Z_j$, and (2) the Z_j are independent random variables, each of which takes the value +1 or -1 with equal probability. It can be shown that in such a series, the expectation value of the magnitude $|S_n|$ is proportional to the square root of n. Thus, in particular, in the limit $n \rightarrow \infty$, this expectation value $E[|S_n|]$ tends to infinity, rather than to zero. Returning to our case of interest: If (simplifying) we assume that the 'effects' of each possible action can be parcelled into discrete components, each of which additively contributes an amount an amount of fixed magnitude but variable sign (thus: either ΔV or $-\Delta V$) to the overall goodness of the world in which it occurs, and that the signs of successive effects are probabilistically independent, then this theorem applies to the case of interest in the way suggested in the main text.

[6] Here I disagree both with Dorsey (2012), who claims that the cancellation postulate (in his terminology, `the balancing-out hypothesis') a priori `seems plausible' although `there is no evidence in its favour', and with Cowen (2006), who regards it as an adequate refutation of the cluelessness worry at least for 'big' actions.

[7] Or perhaps: the probabilities that are supported by the evidence that the agent possesses at the time of decision, i.e. the relevant 'evidential probabilities'. For majority of this paper, the distinction between

This will not help, however, if consideration of unforeseeable effects[8] similarly forces us to accept that

> **(CW$_s$: Cluelessness Worry regarding subjective c-betterness)** We can never have even the faintest idea, for any given pair of acts (A_1, A_2), whether or not A_1 is subjectively c-better than A_2.

Does consideration of unforeseeable effects force us to accept (CW$_s$)? Analogously to the above, it won't *if* we can defend the claim that

> **(NR$_s$ (Non-reversal for subjective c-betterness))** The net effect of taking into account unforeseeable effects would not reverse judgments of subjective c-betterness that we reach based on the foreseeable effects alone.

But, in contrast to the objective non-reversal condition (NR$_o$) discussed in section 1, we *can* defend its subjective analog (NR$_s$), at least for the sorts of 'unforeseeable effects' we have been considering thus far. For consider any possible but unforeseeable future effect[9] $E_1 \mapsto E_2$ that *might*, via the sorts of mechanisms we considered in section 1, result from my decision to perform act A_1 rather than A_2. For sure, it is *possible* that: if I did A_1 then E_1 would result and if I did A_2 then E_2 will result (in symbols: $A_1 \square\!\!\rightarrow E_1$ & $A_2 \square\!\!\rightarrow E_2$). Still, there is no particular reason to think that the correlations between my possible actions and these unforeseeable effects will be that way round, rather than the opposite ($A_1 \square\!\!\rightarrow E_2$ & $A_2 \square\!\!\rightarrow E_1$). It seems plausible, in that case, that given any credence function that it is rationally permissible for me to have at the time of decision, my credence in the second correlation hypothesis is exactly equal to my credence in the first correlation hypothesis. But if this is true for all unforeseeable possible effects $E_1 \mapsto E_2$, then the contribution of those unforeseeable effects to the difference in the *expected values* of A_1 and A_2 is precisely zero, and we have the following result:

> **(EVF)** The expected value of an action is determined entirely via its foreseeable effects.

But (EVF) entails (NR$_s$). Thus there can be no analogue of the cluelessness worry for *subjective* c-betterness.[10]

subjective credences and evidential probabilities will be of little import. It might become relevant in a more sophisticated discussion of the issues that I touch on in section 6.

[8] Once the focus is subjective rather than objective c-betterness, the appropriate definition of 'foreseeable' shifts slightly. For subjective purposes, we should include among 'foreseeable effects' not only 'effects that we can predict with a reasonable degree of confidence', but also any effects that we have clear overall reason to regard as *more* likely to follow on some courses of action than on other courses of action. That is, the 'foreseeable' effects need not exclude e.g. possible effects that are extremely unlikely either way, but whose probabilities are affected in definite ways by our choice of action.

[9] With slight abuse of terminology, where the context prevents any confusion from resulting, I use 'effect' both in the absolute sense (E_1 would be among the effects of choosing A_1) and in the comparative sense (the transition $E_1 \mapsto E_2$ would be an effect of choosing A_1 over A_2). It is, of course, the comparative sense that us ultimately important for the purposes of c-betterness.

[10] Feldman (2006) argues that in fact we are *more* clueless about subjective c-betterness than about its objective analog. But Feldman's argument assumes that in order to estimate which of two acts has the *higher* expected value (and by how much), we need to estimate what the expected value of each act *is*. This latter task is indeed massively more demanding, but (*pace* Feldman) is unnecessary.

3. Lenman's objection: The Principle of Indifference

The Principle of Indifference. Lenman (2000) objects to the reasoning in section 2 on the following grounds: this reasoning presupposes a Principle of Indifference, but (according to him) that principle is false.

To state the Principle of Indifference, we require a notion of *evidential symmetry* between mutually exclusive propositions. This notion is supposed to capture the idea of our having *no more evidence in favour of* one proposition than the other, or *no more reason to believe* one proposition than the other.[11] In particular, we suppose that two propositions are evidentially symmetric when we have *no* evidence that bears on the question of which of the two is true (say, on the assumption that one or the other is true). The Principle of Indifference can then be stated as follows:

> **(POI: Principle of Indifference)** Let $Q_1,\ldots,Q_n$ be any mutually exclusive propositions that are evidentially symmetric for S, and let Q be their disjunction. Let C be any credence function that is rationally permissible for S. Then for all i,j, $C(Q_i|Q)=C(Q_j|Q)$.

At first sight, this principle looks eminently reasonable. It also seems to tell the right story in at least some cases. For example, suppose you know that I am about to flip a coin, *and you know nothing else relevant to the question of whether it will land Heads or Tails*. (In particular, you have no information about whether or not the coin is fair, or about the mechanism by which I will flip it.) Plausibly, you are rationally required to have credence ½ that my coin will land Heads; any other credence seems unacceptably arbitrary.

Lenman is correct in claiming that the above defence of (NR_S) presupposes some form of Principle of Indifference. For that reasoning relies crucially on the claim that one's credence that *if the agent did A_1 then E_1 would result and if the agent did A_2 then E_2 would result* is rationally required to be equal to one's credence in the opposite act-effect correlation (i.e. that *if the agent did A_1 then E_2 would result and if the agent did A_2 then E_1 would result*). But the only reason given for thinking this is that we have *no more reason to believe* that the former correlation obtains than the latter, or vice versa. The required claim follows only if we assume something like (POI) for the present case. Otherwise we have no resources with which to criticise an agent who arbitrarily has credence (say) 0.9 that A_1 (resp. A_2) would lead to 'unforeseeable' effect E_1 (resp. E_2), while acknowledging that she has no *reason* for favouring this correlation over the other.

The 'problem of multiple partitions'. As is well known, however, an *unrestricted* POI (such as the one stated above) is inconsistent, at least unless the relation of evidential symmetry holds between far fewer proposition-pairs than we would naively have assumed[12]. The difficulty is the 'problem of multiple partitions'. It arises from the fact that for any partition $\{Q_1, \ldots, Q_n\}$ of Q, there are many

[11] I follow White (2010) in employing the terminology 'evidential symmetry' to be neutral between these and other formulations.

[12] White (*ibid.*) argues persuasively that the culprit in these paradoxes may indeed be a too-liberal interpretation of 'evidential symmetry', rather than POI itself. This is an important point for the general discussion of POI. Since such 'shifting of the bump in the carpet' would not in the end fundamentally change the state of the debate for current purposes, however, here I set it aside for simplicity of exposition, and assume that POI itself is shown to be at fault by the 'problem of multiple partitions' that I discuss in the main text.

other partitions of Q, none of which we are able to single out as privileged; and POI generally gives mutually inconsistent results when applied to distinct partitions.

This problem arises, in particular, when one partition is a 'selective fine-graining' of another: that is, the second partition involves further fine-grainings of some elements of the original partition but not others. Suppose, for instance, you know only that I am about to draw a book from my shelf, and that each book on my shelf has a single-coloured cover. Then POI seems to suggest that you are rationally required to have credence ½ that it will be red (Q_1=red, Q_2 = not-red; and you have no evidence bearing on whether or not the book is red), but also that you are rationally required to have credence $_1$/n that it will be red, where n is the 'number of possible colours' (Q_i = ith colour; and you have no evidence bearing on what colour the book is).)

This problem would be of merely theoretical interest if it was intuitively clear, in any given example, which partitions were 'natural' and which 'gerrymandered'. For in that case, we could restrict the Principle of Indifference to 'natural' partitions, and even without a precise criterion for naturalness, we would know when to apply vs not to apply POI in practice. And that is arguably an adequate response to the book-colour example: at least on reflection, it is clear that neither the partition {book is red, book is not red} nor the partition {book is red, book is blue, book is yellow…} is especially natural[13], so perhaps this is just a case in which POI clearly falls silent. Unfortunately, however, there are at least some wide classes of cases in which it is not even intuitively clear which partitions should be regarded as privileged for the purposes of POI. (In particular, this often happens in scenarios involving credences about some *continuous* quantity, as continuous quantities are apt to have multiple natural parametrisations that are non-linearly related to one another; for discussion, see e.g. Gillies (2000, p.38-42).)

Rejecting POI. In the light of this problem, a widespread consensus (e.g. Hacking 1965, Kyberg 1974, Van Fraassen 1989, Sober 2003, Schackel 2007, Norton 2008, North 2010) concludes that, despite any initial plausibility that it might naively seem to have had, the Principle of Indifference must simply be abandoned: that is, that there is no true constraint on rational credences even remotely *like* POI. (To find genuine rationality constraints on credences, these theorists often hold, we need to leave the domain of the *a priori* altogether: constraints, on this view, might well be found one information on frequencies and/or empirically obtained knowledge of mechanisms and state spaces is available, but not in any purely *a priori* manner.) Thus, in rejecting POI, Lenman is far from alone.

Lenman does not say very much about precisely what form of 'cluelessness' would result if his argument were accepted, or precisely why and for whom said cluelessness would be problematic. I will return to these issues in section 6. First, I will say why I think Lenman's treatment of these 'simple cluelessness' cases is too pessimistic (section 4), and then present a type of case that (however) I do think raises a genuine threat of cluelessness (section 5).

4. Defence of (NR_s) against Lenman's objection

[13] In particular: although it may initially be tempting to say that the partition into 'all possible colours' {red, blue, green, …} is natural and an appropriate candidate for applying POI, this is clearly implausible on reflection: there is nothing especially natural or unnatural, for instance, either about a partition that identifies turquoise as a colour distinct from blue and green, or about one that declines to do so.

Against this consensus, however: what the ‘problem of multiple partitions’ shows is only that a *fully unrestricted* Principle of Indifference is false (at least given a too-inclusive notion of evidential symmetry). It does not show that there are *no true restrictions* of POI. And simply rejecting all indifference-based reasoning wholesale, as Lenman apparently proposes to do, does seem to throw out too much baby with the bathwater: both in everyday reasoning, and in science. Quite independently of the issue of cluelessness, it seems clear that there are at least some cases in which *something very like* POI gives the right account.

We gave one everyday example above (in which you know only that I am about to flip a coin, and that the two sides of the coin are labelled Heads and Tails). For scientific examples: medical trials, for instance, ultimately aim to guide (posterior) credences about which medications have which effects, or about which patients have which conditions.[14] But, as a matter of statistical reasoning, one can arrive at such restrictions on rational posterior credences in response to evidence only given claims about what prior credences ought to have been. The standard procedure here is to assume that the rational agent begins, prior to gathering evidence, with a `flat’ or `uninformative’ prior – but that is just another term for a prior that satisfies some suitable form of indifference principle. An epistemic agent who was genuinely willing to make any such assumption would be left with no grounds for following his doctor’s advice regarding the post-test probability that he has any given medical condition, or regarding which treatments are likely to help. What this shows is that, the quantity of ink that has been spilt against indifference reasoning notwithstanding, none of us is in fact this type of epistemic agent.

The (warranted) more optimistic view of the situation vis-à-vis cluelessness is as follows. For sure, the Problem of Multiple Partitions shows, as we conceded above, that the conjunction of an overly general Principle of Indifference with an insufficiently critical application of the notion of evidential symmetry leads to paradox. Suppose we concede for the sake of argument that the culprit is the fully general Principle of Indifference (i.e., rather than the notion of evidential symmetry). Then this *fully general* Principle of Indifference is false. But since just about any principle, true or otherwise, is a special case of some natural generalisation that is false, this establishes little. It shows only that the true principles in this vicinity have to be *restrictions* of the Principle of Indifference, rather than that original principle itself.

We must further concede that we do not (yet?) know how to formulate the appropriate restrictions, at least for many or most of the cases of interest. This is an unfortunate situation, in theoretical terms: we are in a situation of impoverished understanding, and we would prefer to understand more. But situations of impoverished understanding should not in themselves surprise us: no-one thinks that the business of epistemology has been completed, whether or not it is completable.

There are, then, some ‘good cases’: cases in which some form of indifference reasoning generates rational constraints on credences, *and we are in a position to recognise these cases as such*, notwithstanding the fact that we do not (yet?) know precisely what form of indifference reasoning it is that does the generating. It is equally clear – intuitively – that the case in hand is just such a ‘good

[14] Here I concur with the ‘Bayesians’ over the ‘classicists’ regarding the cognitive aim of experimentation. For an accessible survey of this controversy in the foundations of statistics, see Sober (2008, chapter 1). On the general point that a wholesale rejection of indifference reasoning goes too far, I am in agreement with White (2010).

case'. While there are countless possible causal stories about how helping an old lady across the road *might* lead to (for instance) the existence of an additional murderous dictator in the 22nd century, any such story will have a precise counterpart, precisely as plausible as the original, according to which *refraining* from helping the old lady turns out to have the consequence in question; and it is intuitively clear that one ought to have equal credences in such precise-counterpart possible stories. And the failure (and paradoxical nature) of a *completely general* Principle of Indifference provides no grounds for doubting this intuitive verdict.

5. **Complex cluelessness**

There are, however, cases that threaten cluelessness in a structurally very different way, and that fall outside the scope of any even remotely plausible form of POI. I will refer to the existence of these cases, and the problem that they arguably pose for anyone who seeks to guide their actions even partially by considerations of goodness of consequences, as the 'Complex Problem of Cluelessness'. The remainder of the paper is much more tentative than sections 1-4; its purpose is more to raise than to resolve a problem.

The cases in question have the following structure: For some pair of actions of interest A_1, A_2,

> (CC_1) We have some reasons to think that the unforeseeable consequences of A_1 would systematically tend to be substantially better than those of A_2;
> (CC_2) We have some reasons to think that the unforeseeable consequences of A_2 would systematically tend to be substantially better than those of A_1;
> (CC_3) It is unclear how to weigh up these reasons against one another.

This talk of 'having some reasons' and 'systematic tendencies' is not as precise as one would like; but some examples should convey the idea. The most vivid[15] examples of this phenomenon occur in the context of 'Effective Altruism' (as outlined by e.g. MacAskill (2015), Singer (2015)). In this context, the agent is considering devoting a significant portion of her resources, in terms of time and/or money, with the express purpose of causing as much good as possible for a fixed amount of input resource. Since the actions in question here involve at least moderate and optional sacrifice on the part of the agent, and since in addition the whole *point* of the actions under consideration would be to maximise good, any cluelessness about which actions have that property feels particularly galling – hence (perhaps) the special vividness.

Here is just one such example. Effective altruists place a lot of weight on the recommendations of independent charity evaluators, whose aim is to rank charities, as far as possible, in terms of overall cost-effectiveness: 'amount of good done per dollar donated'. One charity that consistently comes out top in these rankings, at the time of writing, is the Against Malaria Foundation (AMF), a charity that distributes free insecticide-treated bednets in malarial reasons. To justify this verdict, the charity evaluators clearly need (*inter alia*) estimates of the *consequences* of distributing bednets, per extra net distributed (and hence per dollar donated). Equally clearly, however, these charity evaluators, just like everyone else, cannot possibly include estimates of *all* the consequences of distributing bednets, from now until the end of time. In practice, their calculations are restricted to

[15] 'EA' is not, of course, the *only* source of examples with this structure. In fact, cases with the structure given in (CC_1)-(CC_3) are ubiquitous. I return to this point, and its significance, in section 7.

what are intuitively the ‘direct’ (‘foreseeable’?) consequences of bednet-distribution: estimates of the number and severity of cases of malaria that are averted by bednet-distribution, for which there is reasonably robust empirical data. In fact, the standard calculation[16] focusses exclusively at the effectiveness of bednet-distribution in averting *deaths from malaria of children under the age of 5*, and (using standard techniques for evaluating death-aversions) concludes that those benefits alone suffice for ranking AMF’s cost-effectiveness above that of most other charities. It is only if our condition (NR_s) holds *when these effects alone are treated as the ‘foreseeable’ ones* that the charity evaluators’ calculations can have the intended significance.

Averting the death of a child, however, has knock-on effects that have not been included in this calculation. What the calculation counts is the estimated value *to the child* of getting to live for an additional (say) 60 years. But the intervention in question also has *systematic* effects on others, which latter (1) have not been counted, (2) in aggregate may well be far larger than the effect of prolonging the child’s life on the child himself, and (3) are of unknown net valence. The most obvious such effects proceed via considerations of population size.[17] In the first instance, averting a child death directly increases the size of the population, for the following (say) 60 years, by one. Secondly, averting child deaths has longer-run effects on population size: both because the children in question will (statistically) themselves go on to have children, and because a reduction in the child mortality rate has systematic, although difficult to estimate, effects on the near-future fertility rate.[18] Assuming for the sake of argument that the net effect of averting child deaths is to increase population size, the arguments concerning whether this is a positive, neutral or a negative thing are complex. But, callous as it may sound, the hypothesis that (overpopulation is a sufficiently real and serious problem that) the knock-on effects of averting child deaths are *negative* and *larger in magnitude than the direct (positive) effects* cannot be entirely discounted. Nor (on the other hand) can we be confident that this hypothesis is *true*. And, in contrast to the ‘simple problem of cluelessness’, this is not for the bare reason that it is *possible* both that the hypothesis in question is true, and that it is false; rather, it is because there are complex and reasonable arguments on both sides, and it is radically unclear how these arguments should in the end be weighed against one another.

To get a Principle of Indifference to be of any help here, we would have to regard conditions (CC_1)-(CC_3) above – conditions under which there are competing reasons of quite different characters, and no obviously canonical way of weighing those reasons against one another – as conditions of “evidential symmetry” for the purposes of POI. To be sure, *at the level of description in the previous sentence*, the evidential situation is ‘symmetric’ between the two propositions in question. However, in this case – unlike the ‘simple problem cases’ – this appearance of symmetry disappears as soon as we probe to a deeper level. There is an obvious and natural symmetry between the thoughts that (i)

[16] http://www.givewell.org/international/top-charities/amf#Whatdoyougetforyourdollar

[17] A different sort of concern that (however) would equally be grist to the ‘new cluelessness’ mill has been pressed by Emily Clough (2015): that some effective-altruist-funded interventions might have large and negative longer-run consequences via their political effects. In particular, Clough worries that direct funding of front-line health services by outsiders might diminish the tendency of governments of low-income countries to provide high-quality healthcare services themselves (and of the citizens of the countries in question to demand such things from their governments).

[18] For attempts to determine the latter, see, for example, Roodman (2014) and references therein, and Shelton (2014).

it's *possible* that moving my hand to the left might disturb air molecules in a way that sets off a chain reaction leading to an additional hurricane in Bangladesh, which in turn renders many people homeless, which in turn sparks a political uprising, which in turn leads to widespread and beneficial democratic reforms… and (ii) it's *possible* that refraining from moving my hand to the left has all those effects. But there is no such natural symmetry between, for instance, the arguments for the claim that the world is overpopulated and those for the claim that it's underpopulated, or between the arguments for and against the claim that the direct health benefits of effective altruists' interventions in the end outweigh any disadvantages that accrue via diminished political activity on the part of citizens in recipient countries. And, in contrast to the above relatively optimistic verdict on the Principle of Indifference, clearly there is no remotely plausible epistemic principle mandating equal credences in p and not-p whenever arguments for vs. against p are inconclusive.

Relatedly: unlike the 'simple problem of cluelessness', which strikes many people as sophistical from the start (at least once a notion of subjective betterness is admitted), this 'complex problem of cluelessness' feels real and important – at least to many of us, in some circumstances. Many who would otherwise be drawn to Effective Altruism nonetheless refrain from donating any significant portion of their earnings, not because of any positive belief that refraining from donating will have better consequences[19], but from a sense that they would require more confidence that their donations really would be doing some significant amount of good – less cluelessness – before they are willing to take the bold-feeling step of donating a significant proportion of their income. And, among those who do donate, many donate significantly less than they would if they had no such cluelessness-based worries; they commit *partially* to the EA ethos and in consequence 'hedge their bets', donating some significant amount (in case doing this really does do a lot of good), but far less than they might (in case their sacrifices are all just wasted, or, worse, actually harmful in the long run). Furthermore, even those 'hard effective altruists' who have somehow overcome these worries for practical purposes will, I think, admit that they still feel the pull of these concerns. There is a deep sense of 'decision discomfort' attending the predicament of being forced to make decisions in situations of the character we are now discussing.

6. The nature of cluelessness

Three questions about cluelessness. I have argued that although the cases normally focussed on in the cluelessness literature ('simple problem' cases) generate no genuine threat of cluelessness, nonetheless there does exist a different class of cases (the 'new problem' cases) that do generate such a genuine threat. In light of this, the following three questions become salient.

First: What is the right theoretical description of cluelessness? That is, what exactly is our predicament, in terms of both epistemic and practical normativity, when we face a situation of this type? [20]

[19] *Some* people do refrain from donating for this other reason – some people think, for example, that thye should not child mortality reduction because "there are too many people anyway". Those people are (or take themselves to be) in a simpler epistemic situation, and are not my focus here.

[20] Those unconvinced by the arguments of section 4 can take this discussion to apply equally to the 'simple problem'; there has been surprisingly little said about the precise nature of cluelessness in the 'simple problem' literature.

Given an answer to this first question, we could use it to tackle the second and third questions. Second: to what extent is it actually true, in cluelessness cases, that consideration of consequences cannot guide moral/practical decision-making or evaluation? And third: What is the source of the phenomenology of deep `decision discomfort' that seems to attend (genuine) cluelessness cases, for agents who are at least approximately rational?

A sceptic might respond to these questions as follows. (1) Just as orthodox subjective Bayesianism holds, here as elsewhere, rationality requires that an agent have well-defined credences. Thus, insofar as we are rational, each of us will simply settle, by whatever means, on her own credence function for the relevant possibilities. And once we have done that, subjective c-betterness is simply a matter of expected value with respect to whatever those credences happen to be. In this model, the subjective c-betterness facts may well vary from one agent to another (even in the absence of any differences in the evidence held by the agents in question), but there is nothing else distinctive of 'cluelessness' cases; in particular, (2) there is no obstacle to consequences guiding actions, and (3) there is no rational basis for decision discomfort.

Imprecise credences. This sceptical response *may* in the end be the correct one. But since it at least *appears* that something deeper is going on in cases like the one discussed in section 5, it is worth exploring alternatives to the sceptical response. The alternative line I will explore here begins from the suggestion that in the situations we are considering, instead of having some single and completely precise (real-valued) credence function, agents are rationally required to have *imprecise credences*: that is, to be in a credal state that is represented by a many-membered *set* of probability functions (call this set the agent's 'representor').[21] Intuitively, the idea here is that when the evidence fails conclusively to recommend any particular credence function above certain others, agents are rationally required to remain neutral between the credence functions in question: to include all such equally-recommended credence functions in their representor.

Above (in section 2), we defined subjective c-betterness in terms of expected values. But subjective expected values are, as they stand, defined only for agents who have precise credences. There is thus an open question about how the notion of 'subjective c-betterness' should be extended to the case of *imprecise* credences. Relatedly: we have not yet said anything about how subjective c-betterness relates to normative questions of what one *ought to do.* But we have noted that on any plausible normative theory, there will be some important connection. There will, in that case, similarly be an open question about how to extend the normative theory to the case of imprecise credences.

Three criteria of permissibility under imprecise credences. We will have forged a connection to (some sort of) normativity if we state a principle linking imprecise credences to (some sort of) permissibility. Consider, then, the following three rival principles of permissibility for the imprecise-

[21] Since it deals in credence functions, this approach is broadly Bayesian. The more orthodox Bayesian alternative holds that agents are always rationally required to have some particular precise credence function, but that, especially in situations like the ones we are considering here, it is either the case that many credence functions are rationally permissible (that is, the 'uniqueness thesis' fails), or (if uniqueness does hold) that agents are not in any position to know which credence function is rationally required. Several of the issues I discuss below for the imprecise-credence case also have natural counterparts in the precise-credence framework; in the main text, I focus exclusively on the imprecise-credence case only for reasons of brevity.

credence case[22]; each is a generalisation of the ‘maximise expected value’ principle in the precise-credence context.

> **(LP) Liberal criterion of permissibility**: Act A is permissible in circumstances C iff no other act that is available in C has higher expected value with respect to all elements of the representor.

> **(RP) Restrictive criterion of permissibility:** Act A is permissible in circumstances C iff no other act that is available in C has higher expected value with respect to any element of the representor.

> **(SP) Supervaluational criterion of permissibility**: It is determinately true that A is permissible in C iff there is no other action available in C that has higher expected value with respect to all elements of the representor. It is determinately false that A is permitted in C iff with respect to each element of the representor, some other act available in C has higher expected value. Otherwise it is indeterminate whether or not A is permitted in C.

In the imprecise-credence model, situations of cluelessness seem to be ones of ‘intra-representor disagreement’, in the following sense:

> **(ID: Intra-representor disagreement)** In a situation of cluelessness, the elements of one’s representor disagree with one another on the question of which act(s) maximise(s) expected value.

I will return shortly to the question of what each of the criteria (LP), (RP), (SP) implies (given (ID)) for our questions (2) and (3). First, a short digression.

Degree of neutrality among normative theories. Up to this point in the discussion, I have been at pains to theorise in a way that is neutral among rival candidate accounts of the connection between considerations of consequences and normative principles: I have discussed only (various notions of) *betterness*, assuming only that there is some important connection between betterness and normativity (in particular: I have not assumed consequentialism). But our key questions, in the present section, concern the implications of cluelessness *for decision-guidance and decision discomfort*. Clearly, nothing can be said about these implications without taking on *some* commitments about the nature of that connection. From here, it therefore becomes less straightforward to maintain complete neutrality.

At first sight, a discussion focussed on the above criteria (LP), (RP), (SP) might in fact seem to be of interest only in the context of maximising consequentialism. Clearly, versions of these criteria *are* potentially of interest in that context. For *one* way in which such a criterion could arise begins from a corresponding criterion for subjective c-betterness in the case of imprecise credences, and adds to that a (maximising-consequentialist) principle according to which an act is morally permitted iff no other available act is subjectively c-better.

[22] These criteria (or close cousins thereof), and others, are discussed in more detail in e.g. Elga (2010), Williams (2014), Rinard (2015), Weatherson (MS).

Do non-consequentialists, therefore, have to get off the boat at this point in my discussion, and each conduct an entirely separate discussion of the normative issues, in the context of their own particular non-consequentialist account of the precise connection between betterness and normativity? In fact, while *complete* neutrality among normative theories may be impossible in this part of the discussion, we can be substantially more optimistic than that regarding the prospects for continuing theory-neutrality. With suitable flexibility over the interpretation of the notions of 'value' and 'permissibility' in the criteria (LP), (RP), (SP), it is not clear that anything in those principles need immediately alienate any of quite a wide variety of non-consequentialists.

To see this, first suppose, for example, that one favours the theory developed by Scheffler (1982), combining impartial-consequentialist considerations with agent-centred prerogatives. In that theory, there is considerable flexibility (within the bounds of morality) for the agent to decide the relative weightings of impartial considerations, on the one hand, vs. considerations that are specially important only from the agent's own 'personal point of view' on the other. But still, once the agent has settled this question, (1) he will be equipped with a value function, and (2) *rational* permissibility (given the values that the morally acceptable agent has thereby settled on) presumably requires, in the precise-credence case, maximising the expectation value of that value function.

More generally, it is widely recognised that, if one is willing to countenance agent-relativity of the value function, then just about any normative theory can be represented via a value function (in arguably misleading terminology, any normative theory can then be 'consequentialised'): one merely needs to construct some function that accurately represents the verdicts of the theory in question on questions of overall (moral or rational) comparative *choiceworthiness*. (See, e.g., Portmore (2009) and references therein.) And once that is done, at least one very natural account of permissibility under uncertainty involves a criterion of maximising expected choiceworthiness; the criteria (LP), (RP) and (SP) can then be regarded as extensions of the criterion of 'choiceworthiness' in this (not necessarily consequentialist) sense to the case of imprecise credences.

This is of course not to say that the structure of *every* normative theory will necessarily be such as to make principles anything like (LP), (RP), (SP) either (1) applicable even in principle, or (2) the most natural or illuminating accounts of normativity under imprecise credences. It is also not entirely clear, given only that consideration of consequences *in the ordinary sense* leads to 'intra-representor disagreement' in cases like those discussed in section 5, that this will remain true even when the value function is interpreted as capturing this potentially far broader class of normative considerations. The identification of normative theories for which the discussion as I carry it out here is thus unsuited, and the development of analogous lines of thought for those theories, unfortunately lie beyond the scope of this paper. For the remainder of the paper, I will simply assume that the correct theory of normativity is amenable to the general ideas of maximising subjectively expected choiceworthiness and that (ID) supplies the correct characterisation of cluelessness situations within an imprecise-credence approach, and investigate the prospects for developing an adequate account of cluelessness on that assumption.

Cluelessness via (LP). Assuming (LP), a situation of cluelessness is one in which *each* of the actions among which one is clueless is permitted. In this sense, theory (based on consideration of 'consequences') indeed issues no guidance in the agent's choice among these options; yet, for practical purposes, the agent still has to choose. He must, therefore, choose *arbitrarily*. Might cluelessness, therefore, amount simply to the predicament of being forced by circumstance to make an arbitrary choice?

Something in the ballpark of 'arbitrariness' certainly seems key to the phenomenon of cluelessness. But, it is important to recognise, forced arbitrary choice cannot on its own suffice for cluelessness. To see this, consider Buridan's Ass. The ass's predicament is that (there are no relevant imprecise credences but) two actions tie for first place, either in terms of known actual value or in terms of subjective expectation value. That is, the ass knows that the two options in question are (in objective or subjective terms) *equally good*. In this type of predicament, too, one is forced to choose arbitrarily. But here, *unlike* a situation of genuine cluelessness, there is no call for decision discomfort or paralysis. To be sure, Buridan's Ass itself (the story has it) failed to recognise this, succumbed to decision paralysis, and died of starvation as a result. But most of us, I take it, have progressed beyond this irrationality. We are perfectly happy *in such cases of known equal goodness* simply to choose arbitrarily; we feel no deep decision discomfort in *those* cases. What this shows is that the phenomenon of cluelessness, insofar as it is real, cannot be *merely* a matter of forced arbitrariness. There must be something deeper going on.

(LP), however, seems unable to capture any deeper sense of cluelessness. For it is a theory according to which the actions in question *all have the same normative status as one another* (viz., that of being permitted). It thus seems committed to the view that vis-à-vis considerations of cluelessness, decision situations like those we considered in section 5 really are relevantly just like that of Buridan's Ass. Insofar as the phenomenon of cluelessness in fact does involve some *rational* deeper sense of decision discomfort, this counts against (LP).

Cluelessness via (RP). Assuming (RP), a situation of cluelessness is one in which *no* act is permitted: if the probability functions in the agent's representor mutually disagree about the permissibility of each available option, then, given (RP), all available options are impermissible. On this account, therefore, situations of cluelessness are thus ones of (rational or moral) dilemma.

It is worth noting, however, that the resulting 'dilemmas' would be significantly different in character from those of a more familiar character. Dilemmas have been most extensively discussed in the moral case, where they are normally thought to arise (if at all) in cases in which the agent faces a set of jointly exhaustive options, each option being in some significant way abhorrent. (This class includes 'lesser evil' cases; a typical example is Williams' case of Jim and the Indians (Williams (1973), pp.98-9).) Given a deontological theory[23] that issues absolute prohibitions, for example, it could easily happen that every available option violates at least one of the theory's prohibitions, and is therefore wrong according to the deontological theory. On this approach and given this kind of dilemma, it is easy to understand how moral dilemmas could give rise, if not to cluelessness, then at least to deep discomfort: the agent is forced to make a choice, but (if morally conscientious) has a strong moral aversion to some particular feature of every available option. In the present case, in

[23] Moral dilemmas are normally thought not to occur on a *consequentialist* approach; indeed, depending on their intuitions as to the plausibility of moral dilemmas in general, many theorists take this to be either a significant advantage or a significant disadvantage of consequentialism. It is therefore worth noting that given imprecise credences and the criterion (RP) for moral permissibility, this link between moral dilemmas and *subjective* consequentialism would fail.

contrast, there need not be anything *abhorrent* about any of the options: it need not be, for instance, that any of the options involves killing, or letting disaster occur, or cruelty, or any such thing. It is only if the agent has a strong aversion to moral wrongness *per se* that there will necessarily be any such sense of abhorrence attending moral dilemmas of the kind under consideration.

And in any case – more pertinently for the purposes of our present discussion – there is no reason to think that moral dilemmas of either type should lead specifically to a sense of *cluelessness,* rather than to some other form of discomfort. Even those deontologists who think that lesser-evil cases constitute moral dilemmas generally acknowledge that in some sense the *appropriate thing to do* is to choose the lesser evil, rather than to be paralysed by the observation that all available acts are wrong. And in the absence of any notable imbalance among the options in terms of goodness, amount of 'evil' and so forth, a perfectly acceptable way to respond to a situation in which all options are impermissible *and that is all that can be said* is simply to pick arbitrarily. For this too would be a case in which theory positively tells us that there is nothing to choose, morally, among the options in question. (RP), therefore, like (LP), also seems to furnish only a very shallow sense of cluelessness, little deeper (if at all) than that facing Buridan's Ass. Again, insofar as the phenomenon of cluelessness seems deeper, this counts against (RP).

Cluelessness via (SP). Assuming (SP), for any given option in a situation of cluelessness, it is *indeterminate* whether or not that option is permitted. This verdict seems to offer more promise for capturing the intuitive sense of cluelessness: the agent seeks to choose his actions in response to the permissibility facts, but his actions must be determinate, while those permissibility facts remain stubbornly indeterminate. Nor is it obvious that we can say here, as we did in the cases of (LP) and (RP): "But theory tells us that all actions have the same moral status (viz., here, that of *indeterminate permissibility*), so we are free just to choose among them arbitrarily". For the relevant candidates for normative status arguably do not include 'indeterminate permissibility': rather, they include only the first-order evaluations *permitted, required, forbidden*. And in these *first-order* terms, the criterion (SP) does *not* tell us that the available actions all have the same normative status: depending on the details of the case, it either (1) tells us either that it's indeterminate whether they do or not, or (2) tells us that (it's determinately true that) the available options do *not* have the same moral status, but that it's indeterminate which particular options are required/permitted/forbidden. While these remarks amount to only a very preliminary exploration of the possibilities, therefore, the imprecise-credence model together with (SP) is the most promising route I am aware of for capturing the phenomenology of deep decision discomfort.[24]

7. **Mundane cluelessness**

Effective Altruism makes the new problem of cluelessness particularly *vivid*, and is therefore a good context (for us) to focus on in considering 'new cluelessness'. Clearly, though, insofar as what is distinctive of those cases is (as I have suggested) the satisfaction of conditions (CC_1)-(CC_3) in section

[24] The phenomenon I am calling 'decision discomfort' is discussed at greater length by Williams (2016) (who calls it 'angst'). Williams argues that mere knowledge that one's action is indeterminately permissible does not suffice for the phenomenon in question, but goes on to offer an alternative account of the features of the situations in question that he thinks do account for it. I am not convinced of the details of Williams' proposal, but I lack the space to explore this further here.

5, the basic phenomenon is far from specific to that context: cases with the structure in question also occur in myriad other decision contexts, at both large and small scales. For example: (1) A government's decision-making predicament for any large-scale policy decision – for instance, concerning whether or not to go to war, or whether or not to raise taxes to finance additional spending on education or healthcare. (2) An individual's decision as to which degree course to sign up for, which job to accept, whether or not to have children, how much to spend on clothes, whether or not to give up caffeine.[25] In these cases, no less than the effective-altruist examples discussed above, (a) there are good consequence-based reasons/arguments for favouring each of two alternative actions and also (b) there is no obviously canonical way of weighing up those reasons or arguments against one another.

It follows that insofar as the source of cluelessness is the satisfaction of conditions (CC_1)-(CC_3), one should feel clueless in these everyday cases no less than in the effective-altruist cases.

To what extent *do* we feel clueless in everyday cases? Some people, for sure, suffer from decision paralysis every time the arguments for and against rival possible actions are inconclusive. But here, the appropriate line to take seems more dismissive of a 'clueless' reaction than was arguably the case in section 5. *This* sort of ubiquitous decision paralysis seems more to be a pathology, bordering on a mental illness, rather than the norm. Most of us simply learn to live with the need to resolve decisions with some arbitrariness and with incomplete guidance from data and theory, and, while we may feel more secure in non-arbitrary cases, neither are we especially bothered by the need for arbitrariness, or "sensible judgment", when that need does arise. This suggests that either an excessive deference to the sense of arbitrariness in the effective-altruist cases is also a pathology, or conditions (CC_1)-(CC_3) do not after all strike to the root of the phenomenology of cluelessness in effective-altruist cases. In the latter case, the correct account of cluelessness might lie altogether elsewhere than in the imprecise-credence accounts explored in section 6.

8. Conclusions

Let A_1, A_2 be available actions, and let $V(A_1)$, $V(A_2)$ the overall goodness of the worlds that would ensue if I performed acts A_1, A_2 respectively. 'Simple cluelessness' was supposed to arise merely from the likelihood that the largest contribution to the *objective* value-difference $V(A_1)$-$V(A_2)$ is due to unforeseeable effects of these actions, while (however) that contribution is of unknown sign. I have argued (*contra* Lenman) that while that is indeed very likely, it poses no problem for a 'subjective' criterion of c-betterness, framed in terms of the *expectation* values of $V(A_1)$ and $V(A_2)$. That was because while the unforeseeable effects almost certainly dominate the *objective* value-difference $V(A_1)$-$V(A_2)$, they make zero contribution to the *expected* value-difference $E[V(A_1)-V(A_2)]$. To be sure, the case for that last claim relies on some restriction of the Principle of Indifference. But, I have argued, despite the fact that a fully general Principle of Indifference is paradoxical, it is

[25] Of course, in private decision-making in particular, the 'theory of the good' that the decision-maker seeks to employ for the evaluation of consequences is unlikely to be an impartial one. But just as considerations of cluelessness might (I argued above) look structurally just the same for a wide class of rival moral theories (including non-consequentialist ones), so considerations of cluelessness look structurally just the same in the domain of rationality as they do in the theory of morality. If, for instance, the private individual cares only about his own family, then the relevant value function for that context is one concerning well-being of his family alone, but the remainder of the discussion is largely unaffected.

overwhelmingly plausible that some suitable restricted Principle of Indifference is true. The simple problem of cluelessness is no problem, for consequentialists or for anyone else.

Matters look somewhat different, however, in a different type of case. In 'simple problem' cases, the unforeseeable effects under consideration were ones that, while they *could* result from (say) some particular act A_1, they could *equally easily, and in precisely analogous ways*, result from any of the relevant alternative acts. It is this precise analogy between the possibility that (say) choosing A_1 over A_2 would lead to effect E_1 rather than E_2, and the 'opposite' possibility that choosing A_1 over A_2 would lead to E_2 rather than E_1, that renders plausible the indifference reasoning that is so intuitive in those cases. In contrast, in 'complex problem' cases (I stipulated), one has more specific reasons for suspecting *particular, systematic* correlations between acts and 'indirect' effects, but too many such reasons: non-isomorphic reasons that point in different directions, and for which there is no canonical weighing-up operation. In those cases, no form of indifference principle is at all plausible, and the threat of cluelessness is more genuine.

It is not at all obvious on reflection, however, what the phenomenon of cluelessness really amounts to. In particular, it (at least at first sight) seems difficult to capture within an orthodox Bayesian model, according to which any given rational agent simply settles on some particular precise credence function, and the subjective betterness facts follow. Here, I have explored various possibilities within an 'imprecise-credence' model. Of these, the most promising account – *on the assumption that* the phenomenon of cluelessness really is a genuine and deep one – involved a 'supervaluational' account of the connection between imprecise credences and permissibility.

It is also not at all obvious, however, how deep or important the phenomenon of cluelessness really is. In the context of effective altruism, it strikes many as compelling and as deeply problematic. However, mundane, everyday cases that have a similar structure in all respects I have considered are also ubiquitous, and few regard any resulting sense of cluelessness as deeply problematic in the latter cases. It may therefore be that the diagnosis of would-be effective altruists' sense of cluelessness, in terms of psychology and/or the theory of rationality, lies quite elsewhere.

Acknowledgements

For valuable discussions and correspondence, I am grateful to John Cusbert, Teru Thomas, Robbie Williams, to the audience at the Aristotelian Society, and to the participants in a workshop on 'Ethics and uncertainty' held at Stockholm University.

References

Burch-Brown, J. M. (2014). 'Clues for consequentialists.' *Utilitas*, *26*(01), 105-119.

Clough, E. (2015). 'Effective altruism's political blindspot.' *Boston Review*, July 14 2015. Available online at https://bostonreview.net/world/emily-clough-effective-altruism-ngos. Accessed 1 July 2016.

Cowen, T. (2006). 'The epistemic problem does not refute consequentialism.' *Utilitas*, *18*(04), 383-399.

Dorsey, D. (2012). ‘Consequentialism, metaphysical realism and the argument from cluelessness.’ *The Philosophical Quarterly*, *62*(246), 48-70.

Elga, E. (2010). ‘Subjective probabilities should be sharp.’ *Philosophers’ Imprint* 10(5).

Feldman, F. (2006). ‘Actual utility, the objection from impracticality, and the move to expected utility.’ *Philosophical Studies*, *129*(1), 49-79.

Gillies, D. (2000). *Philosophical theories of probability*. Routledge.

Kagan, S. (1998) *Normative Ethics*. Westview Press.

Hacking, I. (1965). *The logic of statistical inference*. Cambridge University Press.

Kyberg, H. (1974). *The logical foundations of statistical inference*. Dordrecht: Reidel.

Lenman, J. (2000). ‘Consequentialism and cluelessness.’ *Philosophy & public affairs*, *29*(4), 342-370.

MacAskill, W. (2015). *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. Guardian Faber Publishing.

Mason, E. (2004). ‘Consequentialism and the Principle of Indifference.’ *Utilitas*, *16*(03), 316-321.

Moore, G.E. (1903). *Principa Ethica*. Cambridge University Press.

Norcross, A. (1990). ‘Consequentialism and the unforeseeable future.’ *Analysis*, *50*(4), 253-256.

North, J. (2010). An empirical approach to symmetry and probability. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, *41*(1), 27-40.

Norton, J. D. (2008). Ignorance and indifference. *Philosophy of Science*, *75*(1), 45-68.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Portmore, D. W. (2009). Consequentializing. *Philosophy Compass*, *4*(2), 329-347.

Rinard, S. (2015). ‘A decision theory for imprecise probabilities.’ *Philosophers’ Imprint* 15(7).

Roodman, D. (2014). ‘The impact of life-saving interventions on fertility.’ Available online from http://davidroodman.com/blog/2014/04/16/the-mortality-fertility-link/. Accessed 12 July 2016.

Scheffler, S. (1982). *The rejection of consequentialism*. Oxford: Clarendon Press.

Shackel, N. (2007). Bertrand’s Paradox and the Principle of Indifference. *Philosophy of Science*, *74*(2), 150-175.

Shelton, J. ‘Taking exception. Reduced mortality leads to population growth: An inconvenient truth.’ Global health: Science and practice, 2(2), May 2014, pp.135-8.

Singer, P. (2015). *The most good you can do: How effective altruism is changing ideas about living ethically*. Text Publishing.

Smart, J. J. C. (1973) ‘An outline of a system of utilitarian ethics.’ In Smart and Williams (eds.), 1973.

Smart, J. J. C. and Williams, B. (eds.) (1973). *Utiltarianism: For and against.* Cambridge University Press.

Sober, E. (2003). An Empirical Critique of Two Versions of the Doomsday Argument–Gott's Line and Leslie's Wedge. *Synthese*, *135*(3), 415-430.

Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge University Press.

Van Fraassen, B. (1989). *Laws and symmetry*. Oxford University Press.

Weatherson, B. (MS). 'Decision making with imprecise probabilities.'

White, R. (2010). 7. Evidential Symmetry and Mushy Credence. *Oxford studies in epistemology*, *3*, 161.

Williams, B. (1973). 'A critique of utilitarianism.' In Smart and Williams (eds.), 1973.

Williams, J. R. G. (2014). 'Decision-making under indeterminacy.' *Philosophers' Imprint* 14(4).

Williams, J. R. G. (2016) 'Indeterminacy, angst and conflicting values.' Forthcoming in *Ratio*.