Maximal cluelessness

Andreas Mogensen

Global Priorities Institute | September 2019

GPI Working Paper No. 2-2019

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

Maximal cluelessness

ABSTRACT: I argue that many of the priority rankings that have been proposed by effective altruists seem to be in tension with apparently reasonable assumptions about the rational pursuit of our aims in the face of uncertainty. The particular issue on which I focus arises from recognition of the overwhelming importance and inscrutability of the indirect effects of our actions, conjoined with the plausibility of a permissive decision principle governing cases of deep uncertainty, known as the *maximality rule.* I conclude that we lack a compelling decision theory that is consistent with a long-termist perspective and does not downplay the depth of our uncertainty while supporting orthodox effective altruist conclusions about cause prioritization.

JOSH: You know, maybe Marky Mark wants to use his popularity for a good cause, make a contribution. In case you have never heard of that, a contribution is the giving of time, funds -

CHER: Excuse me, but I have donated many expensive Italian outfits to Lucy! And as soon as I get my licence I fully intend to brake for animals, and I have contributed many hours helping two lonely teachers find romance.

- *Clueless* (1995), written and directed by Amy Heckerling

1.

Suppose that I intend to donate some amount of money to charity. Here are two organizations to which I could donate:

- *Make-A-Wish Foundation* is a US non-profit that grants 'wish experiences' to children with critical illnesses. One such child was Miles Scott, who completed treatment for leukemia in 2013. Going by the name 'Batkid,' Miles got the chance to act as the side-kick to (an actor dressed as) Batman. Thousands of people lined the streets of San Francisco to cheer on the duo as they rescued a damsel in distress, foiled a bank robbery, and chased down a super-

villain who had kidnapped the San Francisco Giants' mascot. The event cost the City of San Francisco approximately $105,000, covered by a donation from the philanthropists John and Marcia Goldman. The average cost of granting a 'wish experience' is approximately $7,500.

- *Against Malaria Foundation* is a UK organization that funds the distribution of long-lasting insecticidal nets that protect people from malaria. These distributions typically involve partnering with local governmental agencies in sub-Saharan Africa, who take responsibility for logistics. It is estimated that the cost for AMF to purchase and distribute a long-lasting insecticidal net is around $4.50. GiveWell (2018) estimates that its average cost for saving the life of a child under 5 is $4,471 - 4,491. Since not all benefits derived from distributing nets involve saving young lives, GiveWell estimates the cost per outcome *as good as* averting the death of a child under five to be in the range $757 - 3,197.

Arguably, common sense grants me broad latitude in deciding where I donate, so long as the organizations to which I contribute are, in some sense, minimally decent. Therefore, I can permissibly give to either.

Some philosophers demur. Pummer (2016) and Horton (2017) argue that we are conditionally obligated to be *effective altruists*. Even when it is not obligatory to give $\$x$ to charity, they argue that a person choosing to give $\$x$ to some charity or other is obligated to give it to the organization that will most effectively help others, provided that there exist no adequate, countervailing agent-relative reasons for preferring a less effective organization. This need not rule out funding charities that are especially meaningful to us, but suboptimal considered impartially. In such cases, there may be adequate agent-relative reasons for donating less effectively. Nonetheless, this view seems to rule out the existence of broad latitude in choosing among minimally decent organizations. Of those who have no special ties to the charities among which they are deciding, Horton tells us that "these people ought to try to give to the charities that would use their gift to do the most good. As I have said, there is good data, readily available, to help them." (103-4)

A similar view is taken by Singer (2015). While conceding that stories like Miles Scott's are heart-warming, Singer raises a concern about the wisdom of spending so much money to grant 'wish experiences,' when lives are at stake. According to Singer, effective altruists "know that saving a life is better than making a wish come true. ... So they don't give to whatever cause tugs most strongly at their heartstrings. They give to the cause that will do the most good, given the abilities, time, and money they have available." (6-7)

In this paper, I set out reasons to believe that a concern to give to the cause that will do the most good does not imply what it is commonly thought to imply in comparing between organizations like Make-A-Wish Foundation and Against Malaria Foundation. Given apparently plausible assumptions, an agent whose sole concern is to maximize the good, impartially considered, need not prefer donating to Against Malaria Foundation over Make-A-Wish Foundation. Nor need she prefer donating to Make-A-Wish Foundation. More generally, I will argue that many of the priority rankings that have been proposed by effective altruists seem to be in tension with apparently reasonable assumptions about the rational pursuit of our aims in the face of uncertainty. My objection does not rest on doubts about the details of particular cost-effectiveness assessments. It derives instead from recognition of the overwhelming importance and inscrutability of the indirect effects of our actions, conjoined with the plausibility of a permissive decision principle governing cases of deep uncertainty: the so-called *maximality rule.*

Section 2 will outline the antecedents of the problem I discuss in the literature on cluelessness, explaining why we may think it plausible to respond to cluelessness with an imprecise doxastic state, represented by a set of probability functions. Section 3 makes the case that the maximality rule is a plausible decision rule for agents whose doxastic states are represented by this *set of functions model*. Section 4 then argues that, under plausible assumptions about the state of our evidence, a priority ranking on which impartially beneficent agents should prefer Against Malaria Foundation over Make A Wish Foundation is inconsistent with the conjunction of imprecision and maximality. Section 5 comments on the significance of this conclusion. Section 6 wraps up with a summary and conclusion.

I cannot make any claims to great originality for this paper. My conclusions are ultimately not very different from those of Lenman (2000). The arguments by which I arrive at these conclusions represent incremental extensions of ideas discussed by Greaves (2016). Nonetheless, incremental progress is progress, and I hope this paper will provide readers with renewed appreciation of the challenge posed to effective altruist cause prioritization by the overwhelming importance and inscrutability of the indirect effects of our actions. My aim is not to suggest that this challenge cannot be met, but to make sure that we face up to it.

2.

In order to arrive at a statement of the problem on which I want to focus, I begin by outlining its antecedents in the literature on *cluelessness* (Lenman 2000; Mason 2004; Cowen 2006; Dorsey 2012; Burch-Brown 2014; Greaves 2016). In section 2.1, I set out the cluelessness problem as presented by Lenman (2000). In section 2.2, I reject what I take to be a naïve but tempting response. On this view, we needn't worry about our ignorance of the indirect effects of our actions, because ignorance of this kind is compatible with expected utility maximization. This underestimates the depth of the problem. Using the jargon proposed by Knight (1921), I suggest that we are dealing not with *risk*, but *uncertainty*: ignorance of a kind so deep that our evidence does not warrant assigning precise probabilities to all relevant contingencies. Under Knightian uncertainty, the doxastic attitudes of an agent who proportions her beliefs to the evidence cannot be represented by a unique probability function relative to which an expected utility for each act may be defined.

2.1

As developed by Lenman (2000), the cluelessness problem derives from the following apparently obvious suggestion: if we assess candidate actions in terms of the value of their consequences, we must take account of *all* of their consequences. After all, it's quite possible for the immediate consequences of some action to be good, but for the long-run consequences to be terrible.

As a dramatic illustration, consider a story reported by the priest Max Tremmel (Willis 2012). Tremmel was preceded as Kapellmeister of the cathedral in Passau by Johann Kühberger. Tremmel

was told by Kühberger that he had once rescued another boy from drowning in the Danube. That boy, he said, was Adolf Hitler. But for this single good deed, millions might have been spared from the horrors of the Nazi state.

We must keep in mind, then, that the consequences of our actions do not stop unfolding within the foreseeable future. Instead, they can be expected to go on and on and on.[1] Assuming that Kühberger did in fact save Hitler, the consequences of his act of rescue will presumably echo throughout all that remains of human history. But the same is true of all of those far less dramatic actions that made the difference between whether or not Kühberger rescued Hitler. This includes the decision of Kühberger's parents to conceive him and any decision made by anyone at any prior time in history that made a difference to whether or not Kühberger's parents would meet or happen to conceive a child at the particular time at which Johann Kühberger was conceived.

The fact that the total consequences of our actions have the potential to stretch so far into the future supports *ex post long-termism about value*, according to which the overwhelming determinant of the value-differences between the outcomes we can bring are good and bad events distributed across the long run.[2] [3] In conjunction with ex-post long-termism, the fact that the very long run is inscrutable seems to warrant the conclusion that any comparison between acts in respect of the value of their consequences is determined almost entirely by stretches of future history about which we are

[1] Moore (1903: §93) and Smart (1973: 33) argue that the morally significant effects of our actions drop off very quickly. Lenman (2000: 350-1) and Greaves (2016: 313-5) convincingly rebut this position.

[2] This presumes that we do not exhibit (a suitably high rate of) *pure time preference*, standardly assumed in economic assessments of long-term projects, in light of which later events are counted as less important simply by virtue of being later in time. The assumption of a pure rate of time preference may be considered as a concession to what economists understand the preferences of ordinary people to be. As Broome (1994: 131) notes: "Most philosophers are opposed to pure discounting, and I think most economists would be on their side; I do not think this is a major point of disagreement." For evidence that the general public does not in fact exhibit a positive rate of pure intergenerational time preference, see Frederick (2003).

[3] Cowen (2006) and Dorsey (2012) argue that the unforeseeable effects associated with any pair of actions are in fact highly likely to cancel out in any such comparison. For criticism, see Greaves (2016: 315-6).

utterly ignorant. Insofar as we aim to bring about good consequences through our actions, we then seem to be without meaningful direction. We are clueless.

2.2

Let us call the *Naïve Response* to the cluelessness problem the view that cluelessness is no impediment to deriving action-relevant guidance from the aim of bringing about good consequences once we keep in mind the standard principles of normative decision theory: in particular, the principle that rational decisions maximize *expected utility* (Arnauld and Nicole 1662; Bernoulli 1738; Ramsey 1926; von Neumann and Morgenstern 1947; Savage 1972).

According to the Naïve Response, cluelessness merely points to our inability to know for sure what outcome results from which action. That is no impediment to rational decision making. Given the most natural interpretation of their utility function as a positive linear transformation of moral value, the decision criterion for rational agents insofar are they are concerned to promote the best consequences is the maximization of *expected moral value*. Failure to know what will result from performance of a given action is entirely compatible with being able to assign probabilities to relevant states and cardinal values to relevant outcomes, allowing us to identify the action whose expected utility is at least as great as that of any other. Cluelessness is therefore nothing to worry about.

I call this 'the Naïve Response' because it is natural to object that it fails to take seriously the depth of our uncertainty. Not only do we not have evidence of a kind that allows us to know the total consequences of our actions, we seem often to lack evidence of a kind that warrants assigning precise probabilities to relevant states. Consider, for example, the various sources of uncertainty about the indirect effects of saving lives by distributing anti-malarial bed-nets noted by Greaves (2016). We have reason to expect that saving lives in this way will have various indirect effects related to population size. We have some reason to think that the effect will be to increase the future population, but also some reason to think that it will be to decrease the net population (Roodman 2014; Shelton 2014). It is not clear how to weigh up these reasons. It is even harder to compare the relative strength of the reasons for believing that increasing the population is desirable on balance against those that support believing that population decrease is desirable at the margin. That the distribution of bed-nets

is funded by private donors as opposed to the local public health institutions may also have indirect political consequences that are hard to assess via the tools favoured by the evidence-based policy movement (Clough 2015). To suppose that our uncertainty about the indirect effects of distributing anti-malarial bed-nets can be summarized in terms of a perfectly precise probability distribution over the relevant states seems to radically understate the depth of our uncertainty.

As Greaves (2016) observes, when our evidence is as incomplete, imprecise, or equivocal as it is in this case, we may believe that a rational agent will respond by adopting a belief state that is similarly incomplete, imprecise, or equivocal. Rather than a single probability function, a rational agent's response to evidence of this kind may be represented by a set of probability functions (a so-called *representor*) (Levi 1974; van Fraassen 1990; Joyce 2005, 2011; Schoenfield 2012). Denoting this set as R, the agent's confidence in any proposition p may be represented by the set valued function $Cr(p) = \{\Pr(p) : \Pr(\cdot) \in R\}$.[4]

Using this set of functions model to represent the agent's doxastic state, we can no longer appeal to the maximization of expected utility as the criterion of rational decision. Expected utility theory presumes a unique probability function capable of representing the agent's doxastic attitudes. When there is no such probability function, we need some alternative decision criterion.

3.

This section will take up the question of which decision criterion should govern agents with imprecise credences represented by the set of functions model. My aim is not to establish that some particular decision criterion is uniquely correct, but merely to exhibit one such criterion as sufficiently plausible that it cannot be ruled out: namely, the maximality rule. I introduce the maximality rule in section 3.1. In section 3.2, I consider some of its drawbacks and note alternative decision criteria that avoid these

[4] Insofar as we are concerned that there may be no unique and determinate set of probability functions capable of representing the agent's confidence in *p* without reintroducing the false precision that seems to characterize the precise credences model, we may prefer to interpret *R* as a *fuzzy set*: i.e., a set membership in which is degree-valued, as opposed to binary (Lyon 2017).

drawbacks. I argue that these alternatives face other problems, which may lead us to prefer the maximality rule on balance.

3.1

In order to arrive at a statement of the maximality rule, I begin by describing a general framework for evaluating decision criteria for imprecise credences.

We may consider the task of constructing a criterion of rational decision making as involving specification of a strict preference relation, $\succ$, defined over the set of available acts, A, to which we associate an induced choice correspondence, $C_{\succ}(A)$, consisting of all acts that are not dispreferred to some alternative: $C_{\succ}(A) = \{a \in A |\ \ \forall a' \in A, a' \not\succ a\}$. Any act within the set defined by the choice correspondence is considered rationally permissible with respect to its alternatives. Any act outside the set is considered rationally impermissible with respect to its alternatives.

Assume that the norm for rational decision making with precise credences is the maximization of expected utility. Then a natural assumption within the set of functions model is that the agent's preference between any pair of acts $a, a' \in A$ supervenes on the expected utilities assigned to a and a' relative to the probability functions in R.[5] For example, letting $\mathrm{EU}_{\Pr(\,\cdot\,)}(a)$ denote the expected utility of a relative to some $\Pr(\,\cdot\,) \in R$, the following assumption seems compelling: if $\mathrm{EU}_{\Pr(\,\cdot\,)}(a) > \mathrm{EU}_{\Pr(\,\cdot\,)}(a')$ for all $\Pr(\,\cdot\,) \in R$, then $a \succ a'$. In other words, the agent prefers a to a' if every probability function assigns greater expected utility to a than to a'.

The maximality rule is defined by the assumption that this condition is both sufficient *and necessary* for strict preference between acts: $a \succ_{max} a'$ iff $\mathrm{EU}_{\Pr(\,\cdot\,)}(a) > \mathrm{EU}_{\Pr(\,\cdot\,)}(a')$ for all $\Pr(\,\cdot\,) \in R$. In other words, the agent prefers a to a' *just in case* every probability function assigns greater

[5] If we reject the assumption that rationality for agents with precise credences requires expected utility maximization, we can modify this framework as needed. For example, suppose we believe that rationality for agents with precise credences requires maximization of risk-weighted expected utility (Buchak 2013). Then we should think that for any $a, a' \in A$, the agent's preference between a and a' depends on the risk-weighted expected utilities assigned to a and a' relative to the probability functions in R.

expected utility to a than to a'. It follows that $a \in C_{\succ_{max}}(A)$ just in case there is no alternative $a' \in A$ such that $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) < \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$. In other words, an act is rationally permissible just in case no other act has greater expected utility according to every probability function in the agent's representor.

A natural extension of this framework defines the indifference relation in similar terms: $a \sim_{max} a'$ iff $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) = \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$. Thus, when it is neither the case that $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) > \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$, nor that $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) < \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$, nor that $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) = \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$, the agent neither judges a to be strictly better than a', nor a' to be strictly better than a, nor does she view a and a' as exactly equally preferable. The preference relation therefore violates *connectedness*: it is not the case that for all $a, a' \in A$ either $a \succ_{max} a'$, $a' \succ_{max} a$, or $a \sim_{max} a'$. In some cases, the agent's preferences are simply indeterminate.

So understood, the maximality rule resembles a kind of supervaluationist semantics (van Fraassen 1966; Fine 1975; Keefe 2000).[6] [7] According to supervaluationism, a proposition is true just in case it is true according to every admissible precisification and false just in case it is false according to every admissible precisification. If a proposition is not true according to every admissible precisification, nor false according to every admissible precisification, it indeterminate. Similarly, given the maximality rule, the agent has a determinate ranking of a and a' just in case every admissible precisification of her doxastic state ranks a and a' similarly in respect of expected utility. When there is no consensual ranking of a and a', the agent's preference with respect to these

[6] See also Rinard (2015) on supervaluationism and rational choice under imprecision.

[7] We could adopt an interpretation of the maximality rule that *just is* a supervaluationist semantics for a language expressing preference relations between acts, on which '$a \succ a'$' is true iff $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) > \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$, false iff $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) \not\succ \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$, and otherwise indeterminate (neither true nor false). Note that this is not the interpretation adopted here. On our construal, '$a \succ a'$' is false if it is not the case that $\mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a) > \mathrm{EU}_{\mathrm{Pr}(\,\cdot\,)}(a')$ for all $\mathrm{Pr}(\,\cdot\,) \in R$. That is why I say merely that this 'resembles a kind of supervaluationist semantics.'

options is indeterminate: she neither prefers a to a', nor a' to a, nor does she regard them as exactly equally good.

The foregoing discussion highlights a key attraction of the maximality rule. As Bradley and Steele (2015) put it, the maximality rule "does not contrive a preference between incommensurable options where there is none" (15). Every other decision rule defined in terms of a choice correspondence induced by a binary preference relation between acts determined by their expected utilities relative to the probability functions in R requires the agent to strictly prefer some act although that act is not strictly preferred relative to some admissible precisification of her doxastic state.

3.2

The maximality rule is not without its faults. First and foremost, note that when there are more than two options, some act, a, may be rationally permissible according to the maximality rule, although there is no $\Pr(\cdot) \in R$ relative to which a maximizes expected utility. It may be that every act that does maximize expected utility relative to some probability function is ranked below a according to some other. The maximality rule may therefore be said to count as permissible some acts that are not permissible according to any admissible precisification of the agent's credences.

In light of this, we may prefer the *liberal* rule, defined by the choice correspondence on which $a \in C_{\succ_{lib}}(A)$ just in case there is some $\Pr(\cdot) \in R$ such that for all $a' \in A$, $\mathrm{EU}_{\Pr(\cdot)}(a) \geq \mathrm{EU}_{\Pr(\cdot)}(a')$. By construction, the liberal rule never counts some act as rationally permissible unless it maximizes expected utility relative to some admissible precisification of the agent's credences. However, it has other drawbacks that may lead us to prefer the maximality rule on balance.

In the first instance, it rules out the rational permissibility of a certain kind of *ambiguity aversion*.[8] Consider the *Three Colour Urn Problem* noted by Ellsberg (1961). An urn contains 90 balls. 30 are red. The remainder are either black or yellow, with unknown proportion. Given your ignorance of the chances, we assume that $Cr(red) = \{1/3\}$, $Cr(black) = [0, 2/3]$, $Cr(yellow) = [0, 2/3]$, but $Cr(black\ or\ yellow) = \{2/3\}$. You are offered a choice over two pairs of bets. Bet f

[8] See Weirich (2004: 76-8).

yields \$100 if a red ball is drawn from the urn. Bet g yields \$100 if a black ball is drawn from the urn. Bet f', yields \$100 if a red or yellow ball is drawn from the urn the second time round, whereas g', yields \$100 if a black or yellow ball is drawn. Ambiguity averse agents prefer f to g and g' to f'.

Many people exhibit preferences of this kind (Trautmann and van der Kuilen 2015). However, a joint preference for f over g and g' over f' is inconsistent with the liberal rule: there is no probability function relative to which the expected utility of f is greater than that of g and the expected utility of g' is greater than that of f'. By contrast, this pattern of preferences is permissible according to the maximality rule. To the extent that ambiguity aversion appears rationally permissible,[9] this favours the maximality rule over the liberal rule.

In addition, the liberal rule cannot be represented as a choice correspondence induced by a binary preference relation between acts determined by their expected utilities relative to each $\Pr(\cdot) \in R$, subject to the constraint that the agent's preference between a and a' should be the same in $A = \{a, a'\}$ as in any superset of A. Call this constraint *binary independence*. To see that the liberal rule violates binary independence, suppose we have three acts, a, a', and a''. Suppose that there is some $\Pr_1(\cdot) \in R$ such that $\mathrm{EU}_{\Pr_1(\cdot)}a > \mathrm{EU}_{\Pr_1(\cdot)}a'$, whereas for any other $\Pr(\cdot) \in R, \mathrm{EU}_{\Pr(\cdot)}a < \mathrm{EU}_{\Pr(\cdot)}a'$. Similarly, suppose there is some $\Pr_2(\cdot) \in R$ such that $\mathrm{EU}_{\Pr_2(\cdot)}a > \mathrm{EU}_{\Pr_2(\cdot)}a''$, whereas for any other $\Pr(\cdot) \in R$, $\mathrm{EU}_{\Pr(\cdot)}a < \mathrm{EU}_{\Pr(\cdot)}a''$. Finally, assume that $\mathrm{EU}_{\Pr_1(\cdot)}a < \mathrm{EU}_{\Pr_1(\cdot)}a''$ and $\mathrm{EU}_{\Pr_2(\cdot)}a < \mathrm{EU}_{\Pr_2(\cdot)}a'$. Then $a \notin C_{\succ_{lib}}(\{a, a', a''\})$. When all three acts are available, $a \prec a'$ or $\prec a''$. However, $a \in C_{\succ_{lib}}(\{a, a'\})$ and $a \in C_{\succ_{lib}}(\{a, a''\})$. When only the two acts are available, $a \prec a'$ and $a \prec a''$. So binary independence is violated.[10]

Here is a different drawback of the maximality rule that may lead us to prefer some alternative decision criterion, albeit not the liberal rule. Consider the sequential decision problem noted by Elga (2010). Assume there is some proposition, p, such that $Cr(p) = [0.1, 0.8]$. You know

[9] On the rationality of ambiguity aversion, see Ellsberg (1961), Raiffa (1961), Al-Najjar and Weinstein (2009), Siniscalchi (2009), Bradley (2017: 175-177, 271-286).

[10] See Weatherson (1998) for further discussion of the significance of this observation. Thanks to [redacted] for helpful discussion of the argument in this paragraph.

that I will offer you two bets in quick succession. Bet A yields \$15 given $\neg p$, but results in - \$10 given p. Bet B yields \$15 given p, but results - \$10 given $\neg p$. Accepting both bets guarantees that you will gain \$5. Therefore, it seems plausible that no rational agent who knows the procedure in advance will decline both bets, as this amounts to passing up a sure \$5 gain. However, given suitable assumptions, the maximality rule entails that an agent may decline each bet without either declension being rationally impermissible.[11] If we wish to convict the agent of acting irrationally, we must therefore deny the principle that a sequence of acts is rational just in case each member of the sequence is rational.[12]

By contrast, the so-called *Γ-maxmin rule* requires the agent to accept each bet in the sequence just described, assuming that it is rational to reason by backward induction (Sahlin and Weirich 2014). According to the Γ-maxmin rule $a \succ_\Gamma a'$ iff $\min_{\Pr(\cdot)\in R} \mathrm{EU}_{\Pr(\cdot)}(a) > \min_{\Pr(\cdot)\in R} \mathrm{EU}_{\Pr(\cdot)}(a')$. In other words, a is strictly preferred to a' just in case the lowest expected utility assigned to a by some $\Pr(\cdot) \in R$ is higher than the lowest expected utility assigned to given to a' by some $\Pr(\cdot) \in R$. An agent who obeys the Γ-maxmin rule and reasons by backward induction will accept Bet A, knowing that this guarantees subsequent acceptance of Bet B and a sure \$5 reward, whereas declining Bet A guarantees subsequent rejection of Bet B and retaining her status quo wealth.

Nonetheless, Γ-maxmin has a number of significant drawbacks. Most obviously, it is extremely restrictive. The Γ-maxmin rule requires rational agents to behave as if the worst possible expected utility estimate associated with each act were correct. Rationality surely does not require such extreme pessimism, even if it may permit it.

In addition, Weatherson (1998) shows that Γ-maxmin violates *Restricted Conglomerability*, according to which for any $a, a' \in A$, if a will be strictly preferred to a' upon learning p and a will be strictly preferred to a' upon learning $\neg p$, then a must be strictly preferred to a'. Bradley and Steele (2015) show that in sequential decision problems that include the option to gather information, Γ-

[11] The same is true of the liberal rule.

[12] See Weatherson (1998), Elga (2010), and Rinard (2015) for discussion.

maxmin may require the agent to pay to avoid free information in cases where the maximality rule permits but does not require information aversion of this kind.

In sum, there exist competing decision criteria that are superior to the maximality rule in some respect or other, but I do not know of any alternative decision criterion that is all-things-considered preferable. I submit that we cannot rule out the maximality rule. As a result, we ought to avoid drawing any conclusions that are inconsistent with it.

4.

In this section, I argue that an agent whose utility function is a positive linear transform of impartial good will not prefer donating to Against Malaria Foundation over Make-A-Wish Foundation if she responds to cluelessness with imprecision and satisfies the maximality rule, provided that she shares our evidence. Section 4.1 emphasizes the depth of our uncertainty concerning the indirect effects of donating to Against Malaria Foundation. Section 4.2 reflects on the lessons to be drawn in applying the maximality rule to a choice between these organizations.

4.1

In comparing Make-A-Wish Foundation unfavourably to Against Malaria Foundation, Singer (2015) observes that "saving a life is better than making a wish come true." (6) Arguably, there is a qualifier missing from this statement: 'all else being equal.' Saving a child's life need not be better than fulfilling a child's wish if the indirect effects of saving the child's life are worse than those of fulfilling the wish. We have already touched on some of the potential negative indirect effects associated with the mass distribution of insecticide-treated anti-malarial bed-nets in section 2.2, but they are worth revisiting in order to make clear the depth of our uncertainty.

Firstly, there are potential effects on population. When people survive childhood in greater numbers, it is natural to expect the population to grow. The explosion in global population observed since the 17th century is arguably attributable principally to declining mortality (McKeown 1976). However, we must also account for the impact of reduced childhood mortality on family planning. When childhood mortality declines, parents in developing countries need not have as many children in

order to ensure that they can be supported in old age. As a result, averting child deaths may cause the rate of population growth to decline (Heer and Smith 1968). It is the position of the Gates Foundation that averting child deaths at the current margin will reduce population size (Gates and Gates 2014). Many studies confirm that the effect of reduced childhood mortality on population size is offset by reduced fertility (Schultz 1997; Conley, McCord, and Sachs 2007; Lorentzen, McMillan, and Wacziarg 2008; Murtin 2013). Others find that the reduction in births is less than one-to-one with respect to averted child deaths (Bhalotra and van Soest 2008; Herzer, Strulik, and Vollmer 2012; Bhalotra, Hollywood, and Venkataramani 2012). Unfortunately, the studies just noted are of different kinds (cross-country comparisons, panel studies, quasi-experiments, large-sample micro-studies), with different strengths and weaknesses, making it difficult to draw firm conclusions.[13]

In any case, there are potential negative effects associated with both an increase and a decrease in population size. We are most familiar with the adverse impacts projected from increased population. The fears voiced by doomsayers in the 1960s and 70s (Ehrlich 1968; Meadows et al. 1972) have largely failed to materialize. Nonetheless, concern about overpopulation has renewed in recent years. In 2007, the UK's All Party Parliamentary Group on Population, Development, and Reproductive Health warned that the Millennium Development Goals "are difficult or impossible to achieve with current levels of population growth in the least developed countries and regions." (4) A growing population also puts significant strain on the natural world, with attendant declines in wild populations, species loss, and reductions in biodiversity (Crist, Mora, and Engelman 2017).

There are nonetheless potential benefits associated with increased population size. Evidence drawn from long-run historical trends indicates that larger populations are associated with increased technical innovation (Kremer 1993). Nor is this especially surprising (Ord 2014). Assuming that the number of technical innovations per person is independent of population size, we should expect more innovations in larger populations merely in virtue of their size. Furthermore, individuals living in larger populations may be more innovative on average due to network effects, and overpopulation itself can spur technological change. Looking specifically to the history of food production in Europe,

[13] See Roodman (2014) for discussion.

Boserup (1976) argues that temporary overshoot of a population's carrying capacity drives technical innovations that increase the population's long-term carrying capacity.

We should also note that there are powerful arguments supporting the view that the existence of additional people with lives worth living is intrinsically valuable (Broome 2005; Huemer 2008). Debates concerning overpopulation often seem to assume that different populations are to be ranked as better or worse in terms of their average well-being. However, this entails the *Sadistic Conclusion*, according to which it can be better to add some number of lives that contain only suffering as opposed to a much larger number of lives that are well-worth living (Arrhenius 2000). If we grant that a larger population with a lower average but a greater total of welfare can be preferable, the question of whether there are currently too many people becomes even harder to assess.

All of the indirect effects we have so far considered unfold over relatively brief timescales. There are also the potential long-run consequences to consider. We have already noted that the impact of a reduced malarial disease burden on population size may contribute to environmental degradation and resource strain, as well as potentially affecting the pace of technological innovation. There is also evidence to suggest that it will accelerate economic growth through beneficial impact on child development (Sachs 2003; Bloom et al. 2004), though the evidence is far from univocal (Acemoglu and Johnson 2006). What are the possible downstream impacts of these effects?

For example, what is their impact on the risk of human extinction? Further degradation of the environment and changes in the pace of technological innovation due to a lowered malarial disease burden would presumably make some difference to the risk of extinction, if only a small difference. However, even small changes in extinction risk may be considered of enormous significance. Assuming a total utilitarian axiology, Bostrom (2013: 18-19) argues that a conservative projection of the total future population yields an estimate of the expected moral value of reducing extinction risk by one millionth of one percentage point that is at least the value of a hundred million human lives. Giving a mere one percent credence to less conservative estimates that take into account the potential for (post-) humanity to spread to the stars and for future minds to be implemented in computational hardware, Bostrom calculates the expected value of reducing the risk of extinction by as little as one

billionth of one billionth of one percentage point to be one hundred billion times the value of a billion human lives.

More generally, we may think that very great significance attaches to even very small changes in the chance of some outcome whose impact on the welfare of sentient beings will be distributed over time-scales measured in thousands, millions, or billions of years, whether it concerns the long-run survival of our species or the long-term trajectory of Earth-originating civilization. We noted earlier that it is reasonable to believe that the overwhelming determinant of the value-differences between the outcomes we can bring about are due to good and bad events distributed across centuries, millennia, and beyond. An analogous principle - *ex ante long-termism with respect to value* - plausibly applies to differences in expected value. Relative to any reasonable probability function, differences in expected value between the acts available to us are almost entirely attributable to possible good and bad outcomes distributed across the long term.

4.2

For the reasons just noted, a sensible comparison between Make-A-Wish Foundation and Against Malaria Foundation in respect of promoting the impartial good cannot rest on the observation that saving a child's life is better than fulfilling a child's wish. Relative to any reasonable probability function, very little of the difference in expected value between these acts turns on effects of this kind. It is determined principally by possible long-term impacts. These long-term impacts are very hard to probabilify with even moderate precision, whereas even small differences in the probability of persistent, large-scale events such as human extinction will decisively tip the balance when comparing the expected moral value of these alternatives.

The aim of the foregoing section was not to argue for donating to Make-A-Wish Foundation, of course. It was to highlight the depth of our uncertainty. It was intended to emphasize that our evidence concerning the total impact of our choice between these organizations is incomplete, imprecise, and equivocal. Moreover, it was intended to render plausible the view that the evidence is sufficiently ambiguous that the probability values assigned by the functions in the representor of a rational agent to the various hypotheses that impact on the long-run impact of her donations ought to

be sufficiently spread out that some probability function in her representor assigns greater expected moral value to donating to the Make-A-Wish Foundation. Therefore, an agent whose utility function is a linear transform of moral value but who responds to cluelessness with imprecision and obeys the maximality rule will not strictly prefer donating to Against Malaria Foundation.

Admittedly, I cannot prove that a rational agent's credences must be so spread out. There are a number of barriers to constructing a proof of that kind. Firstly, we do not as yet have a plausible, tractable epistemology for imprecise credences. White (2010) puts forward the *Chance Grounding Thesis*, according to which your "spread of credence should cover the range of chance hypotheses left open by your evidence" (174). However, Joyce (2011) notes the following counterexample. Imagine a coin with a black side and a grey side of whose bias you are totally ignorant, except that you know it was chosen randomly from an urn containing a coin of bias $1 - \beta$ for every coin of bias β that it contains. Then the range of chance hypotheses concerning the result of the toss that are left open by your evidence is maximally spread out, but your credence that the coin will land with its grey side up will be sharp: $C(\text{red}) = \{1/2\}$. So the Chance Grounding Thesis is false. Joyce does not offer any concrete suggestion to replace it, describing it instead as "merely the most extreme of a range of possible positions." (289)

That is just one obstacle. Anti-luminosity is another. Even if we knew how imprecise credences should relate to our evidence, membership of a given probability function in the representor of an agent who respects her evidence might be impossible for us to determine because we are not always in a position to know what our evidence is (Williamson 2000; Srinivasan 2015).

Proving definitively that your credences must be so spread out as to include a probability function of a certain kind is therefore not in general within our powers. Accordingly, my argument rests in large part on an appeal to intuition. But the intuition to which I am appealing strikes me as sufficiently forceful and sufficiently widely shared that we should consider the burden of proof to fall on those who deny it.

5.

This section will consider to what extent the conclusion reached in section 4 generalizes. I consider two dimensions along which we might wish to generalize. Firstly, we might want to include other charitable organizations and altruistic initiatives in our comparison. I take up this issue in section 5.1. Obviously, I cannot compare across all possible organizations. My remarks will be selective and sometimes merely gestural. We may also wonder to what extent the suggestion that a rational agent who responds to cluelessness with imprecision need not donate to Against Malaria Foundation when donating to Make-A-Wish Foundation is possible depends specifically on the assumption of the maximality rule. I take up this issue in section 5.2, examining whether a similar conclusion could be reached given the liberal rule.

5.1

Suppose I am right that in a binary choice between donating to Against Malaria Foundation and Make-A-Wish Foundation, an agent whose utility function is linear in moral value and responds to cluelessness with imprecision is rationally permitted to donate to Make-A-Wish Foundation if the maximality rule is correct. It's then natural to wonder what happens if we include other organizations in the comparison. After all, in reality we are not limited to choosing between Make-A-Wish Foundation and Against Malaria Foundation.

It is also worth noting that the choice correspondence defined by the maximality rule does not obey *expansion consistency* (also known as *Sen's Condition* β). It is not always the case that when $a, a' \in A \subset A^*$, $a \in C_{\succ_{max}}(A)$, and $a' \in C_{\succ_{max}}(A^*)$, then $a \in C_{\succ_{max}}(A^*)$. It may happen that the expected utility of a is not higher than that of a' according to every $\Pr(\cdot) \in R$ and that the expected utility of a' is not higher than that of a according to every $\Pr(\cdot) \in R$, but nonetheless there is some third alternative a'', such that the expected utility of a'' is higher than that of a according to every

$\Pr(\cdot) \in R$ but not higher than that of a' according to every $\Pr(\cdot) \in R$.[14] Therefore, both a and a' are permissible given the choice set $A = \{a, a'\}$, whereas a' but not a is permissible in the choice set $A^* = \{a, a', a''\}$. Hence, it is possible that although the maximality rule permits donating to Make-A-Wish Foundation in a binary comparison with Against Malaria Foundation, it permits donating to Against Malaria Foundation but not Make-A-Wish Foundation when these options are embedded in a larger set of alternatives.

However, I conjecture that if we were to make a similar comparison between Make-A-Wish Foundation and others organization that are currently ranked highest by effective altruist evaluators like GiveWell,[15] we would reach similar conclusions. The comparison between Make-A-Wish Foundation and Against Malaria Foundation made in section 4 may be considered a case study, exemplifying a pattern that should be expected to recur for similar comparisons between similar projects. We simply do not know enough about the long-run impact of acts whose short-term effects involve improvements in the health and well-being of people living in poverty in developing countries. For example, we do not know the sign and magnitude of any such action on the chance of human extinction. We may expect that any change, whether positive or negative, will be extremely small, but even extremely small shifts in the chance of extinction - changes as small as one billionth of one billionth of one percentage point - may end up dominating expected value calculations. If our evidence cannot rule out that the chance of extinction is ever so slightly higher given a choice to donate to one of GiveWell's top charities as opposed to an organization like Make-A-Wish Foundation, an impartially benevolent agent who responds to cluelessness with imprecision and obeys the maximality rule arguably need not prefer the former.

---

[14] For example, suppose R contains exactly three probability functions: $\Pr_1(\cdot)$, $\Pr_2(\cdot)$, and $\Pr_3(\cdot)$ such that $\mathrm{EU}_{\Pr_1(\cdot)}(a) = 2, \mathrm{EU}_{\Pr_1(\cdot)}(a') = 1, \mathrm{EU}_{\Pr_1(\cdot)}(a'') = 3$, $\mathrm{EU}_{\Pr_2(\cdot)}(a) = 1, \mathrm{EU}_{\Pr_2(\cdot)}(a') = 2, \mathrm{EU}_{\Pr_1(\cdot)}(a'') = 3$, and $\mathrm{EU}_{\Pr_3(\cdot)}(a) = 1, \mathrm{EU}_{\Pr_3(\cdot)}(a') = 3, \mathrm{EU}_{\Pr_3(\cdot)}(a'') = 2$.

[15] As of the time of writing, GiveWell's top charities are Malaria Consortium, Deworm the World Initiative, Hellen Keller International's vitamin supplementation program, Against Malaria Foundation, Schistosomiasis Control Initiative, Sightsavers' deworming program, the END fund's deworming program, and GiveDirectly.

Of course, there are further options to consider besides interventions that improve the health and well-being of people living in extreme poverty. We might be especially interested in assessing acts that are directly aimed at improving the long-run future of Earth-originating civilization. In this paper, I have emphasized the case for both *ex post* and *ex ante* long-termism about the value. Anyone who aims to maximize moral value and becomes convinced of these claims might be expected to shift their attention to projects that are specifically focused on the long-term future of humanity (Beckstead 2013; Bostrom 2003, 2013). These might include efforts to reduce the risk of near-term extinction for our species: for example, by spreading awareness about dangers posed by synthetic biology (Nouri and Chyba 2008) or artificial intelligence (Bostrom 2014).

The problem is that we do not have good evidence of the efficacy of such interventions in achieving their ultimate aims. Nor is such evidence in the offing. The idea that the future state of human civilization could be deliberately shaped for the better arguably did not take hold before the work of Enlightenment thinkers like Condorcet (1822) and Godwin (1793). Unfolding over time-scales that defy our ability to make observations, efforts to alter the long-run trajectory of Earth-originating civilization therefore resist evidence-based assessment, forcing us to fall back on intuitive conjectures whose track record in domains that *are* amenable to evidence-based assessment is demonstrably poor (Hurford 2013). This is not a case where it can be reasonably claimed that there is good evidence, readily available, to constrain our decision making.

Further study of the implications of imprecision and maximality in comparing far-future focused interventions with competing alternatives is required, but I believe we should be antecedently sceptical that such interventions will come out as rationally obligatory for impartially beneficent agents who satisfy imprecision and maximality, given the paucity of evidence for their efficacy.

5.2

I noted earlier that given the choice set, $A = \{a, a'\}$, the liberal rule and the maximality rule must agree on whether a is preferred to a' or vice versa. Therefore, in a binary comparison between Make-A-Wish Foundation and Against Malaria Foundation, the two rules coincide. Why, then, have I focused on the maximality rule in my discussion?

The two rules behave differently when the set of alternatives has more than two elements. In general, the liberal rule is more restrictive than the maximality rule. Any act that is permissible according to the former is permissible according to the latter, whereas the converse does not hold. Note also that the liberal rule fails to satisfy expansion consistency.[16] Hence, it's possible that although both the liberal rule and the maximality rule permit donating to Make-A-Wish Foundation in a binary comparison with Against Malaria Foundation, the liberal rule permits donating to Against Malaria Foundation but not Make-A-Wish Foundation when these options are embedded in a larger set of alternatives relative to which the maximality rule permits donating to either.

Nor need this be considered a bare possibility.[17] Suppose there is some probability function relative to which donating to Make-A-Wish Foundation has greater expected moral value than donating to Against Malaria Foundation. Let it be a probability function assigning suitably high probability to the conjunctive hypothesis that distributing anti-malarial bed-nets leads to population increase and population increase leads to suitably undesirable downstream effects. In that case, supporting family planning initiatives that reduce the rate of population growth in developing countries will presumably have even greater expected moral value. We may expect this observation to generalize. If a probability function in the representor of an impartially beneficent agent assigns a suitably high probability to certain negative downstream effects associated with reducing the malarial disease burden, some alternative intervention that directly targets the relevant problem will have greater expected moral value than something completely irrelevant like helping to make a child's wish to be superhero for the day come true. Because the potential negative effects associated with reducing the malarial disease burden are diverse in character, the same alternative need not consistently rank above Make-A-Wish Foundation. Nonetheless, if there is always some varying alternative with greater expected moral value relative to any probability function in the agent's representor, the liberal rule forbids donating to Make-A-Wish Foundation. By contrast, when there is no common alternative

[16] The case noted in footnote 14 is one example in which this failure is manifest.

[17] I'm grateful to [redacted] and [redacted] for pointing this out to me.

that ranks higher according to every probability function, the maximality rule permits donating to Make-A-Wish Foundation.

For this reason, the choice between the maximality rule and the liberal rule need not be idle. However, the issue requires further study. In particular, it is not obvious that suitable alternatives actually exist. Consider again a probability function assigning suitably high probability to the conjunctive hypothesis that distributing anti-malarial bed-nets leads to population increase and population increase leads to suitably undesirable downstream effects. We might agree that if we were able to do something about population growth without additional adverse consequences, doing so would have greater expected moral value relative to any such probability function. We may nonetheless wonder whether the antecedent is satisfied. No doubt, there are many family planning initiatives that seem sensible. However, many interventions that have seemed sensible have had no effect or have made things worse (MacAskill 2015: 1- 10). Is there actually some intervention to which I can lend my support whose efficacy in reducing population growth is supported by sufficiently robust evidence, and whose potential negative side-effects are sufficiently robustly counterevidenced? Sadly, providing a convincing answer to this question also exceeds the scope of this paper.

6.

I have argued that we know much less about what it would mean to rationally promote the impartial good than we think we do. In particular, I have argued that a rational agent who is impartially beneficent need not prefer donating to Against Malaria Foundation rather than Make-A-Wish Foundation if she obeys the maximality rule. I have offered reasons to expect that this conclusion will generalize to many similar cause comparisons.

I do not insist that the maximality rule is correct. I merely claim that it is sufficiently plausible that we cannot rule it out. For all we know, orthodox effective altruist conclusions about cause prioritization are all true. In fact, I am inclined to believe they are. The problem is that I do not know how to set out and argue for a decision theory that is consistent with a long-termist perspective and supports these conclusions without downplaying the depth of our uncertainty. Then again, as a

philosopher, I know that I am inclined to believe a great many things for which I lack an adequate response to certain apparently compelling sceptical challenges. Some may share my conviction that this is just one of those cases. But those who are already sceptical of effective altruist conclusions undoubtedly will not.

Bibliography

Acemoglu, Daron and Johnson, Simon (2006) Disease and development: the effect of life expectancy on economic growth. NBER Working Paper No. 12269.

All Party Parliamentary Group on Population, Development, and Reproductive Health (2007) Return of the population growth factor: its impact upon the Millennium Development Goals. <http://www.appg-popdevrh.org.uk/Return%20of%20the%20Population%20Growth%20Factor.pdf> Last accessed 13/03/2019.

Al-Najjar, Nabil, and Weinstein, Jonathan (2009) The ambiguity aversion literature: a critical assessment. *Economics and Philosophy* 25, 249-284.

Arnauld, Antoine and Nicole, Pierre (1662) *La logique, ou l'art de penser*. Paris: Jean Guignart, Charles Savreux, & Jean de Lavnay.

Arrhenius, Gustaf (2000) An impossibility theorem for welfarist axiology. *Economics and Philosophy* 16, 247–266.

Beckstead, Nick (2013) On the overwhelming importance of shaping the far future. PhD thesis: Department of Philosophy, Rutgers University.

Bernoulli, Daniel (1738) Specimen theoriae novae de mensura sortis. *Papers of the Imperial Academy of Sciences in Petersburg* 5, 175-192

Bhalotra, Sonia and van Soest, Arthur (2008) Birth-spacing, fertility, and neonatal mortality in India: dynamics, frailty, and fecundity. *Journal of Econometrics* 143, 274-290.

Bhalotra, Sonia, David Hollywood, and Atheendar Venkataramani (2012) Ferility responses to infant maternal mortality: quasi-experimental evidence from 20th century America. <https://www.aeaweb.org/aea/2013conference/program/retrieve.php?pdfid=363> Last accessed 13/03/2019.

Bloom, David, David Canning, and Dean Jamison (2004) Health, wealth, and welfare. *Finance and Development* 41, 10-15.

Boserup, Ester (1976) Environment, population, and technology in primitive societies. *Population and Development Review* 2, 21–36

Bostrom, Nick (2003) Astronomical waste: the opportunity cost of delayed technological development. *Utilitas* 15, 308-14.

(2013) Existential risk prevention as a global priority. *Global Policy* 4, 15-31.

(2014) *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press.

Bradley, Richard (2017) *Decision theory with a human face*. Cambridge: Cambridge University Press.

Bradley, Seamus and Steele, Katie (2015) Can free evidence be bad? Value of information for the imprecise probabilist. *Philosophy of Science* 83, 1-28.

Broome, John (1994) Discounting the future. *Philosophy and Public Affairs* 23, 128-156.

(2005) Should we value population? *Journal of Political Philosophy* 13, 399-413.

Burch-Brown, Joanna (2014) Clues for consequentialists. *Utilitas* 26, 105-119.

Clough, Emily (2015) Effective altruism's political blind spot. *Boston Review* <https://bostonreview.net/world/emily-clough-effective-altruism-ngos> Last accessed 13/03/2019.

Condorcet, Marquis de (1822) *Esquisse d'un tableau historique des progrès de l'esprit humain*. Paris: Masson et Fils.

Conly, Dalton, Gordon McCord, and Jeffrey Sachs (2007) Africa's lagging demographic transition: evidence from exogenous impacts of malaria ecology and agricultural technology. NBER working paper No. 12892.

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

Crist, Eileen, Camilo Mora, and Robert Engelman (2017) The interaction of human population, food production, and biodiversity protection. *Science* 356, 260-264.

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

Ehrlich, Paul (1968) *The population bomb.* New York, NY: Ballantine Books.

Ellsberg, Daniel (1961) Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics* 75, 643-669.

Elga, Adam (2010) Subjective probabilities should be sharp. *Philosophers' Imprint* 10, 1-11.

Fine, Kit (1975) Vagueness, truth, and logic. *Synthese* 30, 265–300

Frederick, Shane (2003) Measuring intergenerational time preference: are future lives valued less? *Journal of Risk and Uncertainty* 26, 39-53.

Gates, Bill and Gates, Melinda (2014) 2014 Gates annual letter. <https://www.gatesfoundation.org/Who-We-Are/Resources-and-Media/Annual-Letters-List/Annual-Letter-2014> Last accessed 13/03/2019.

GiveWell (2017) Sayana Press. < https://www.givewell.org/international/technical/programs/sayana-press> > Last accessed 13/03/2019.

(2018) 2018 GiveWell cost-effectiveness analysis - version 3. <https://docs.google.com/spreadsheets/d/1IhtZJcWWUQRndEHFHgty5OLdLL4FkNbFMTsKcwGz03g/edit#gid=1537947274 > Last accessed 25/07/2018.

Godwin, William (1793) *Enquiry concerning political justice and its influence on morals and happiness*. London: G. G. J. and J. Robinson.

Greaves, Hilary (2016) Cluelessness. *Proceedings of the Aristotelian Society* 116, 311-339.

Heer, David and Smith, Dean (1968) Mortality level, desired family size, and population increase. *Demography* 5, 104-121.

Herzer, Dierk, Holger Strulik, and Sebastian Vollmer (2012) The long-run determinants of fertility: one century of demographic change 1900–1999. *Journal of Economic Growth* 17, 357-385.

Horton, Joe (2017) The all or nothing problem. *Journal of Philosophy* 114, 94-104.

Huemer, Michael (2008) In defence of repugnance. *Mind* 117, 899-933.

Hurford, Peter (2013) Why I'm sceptical about unproven causes (and you should be too). <http://everydayutilitarian.com/essays/why-im-skeptical-about-unproven-causes-and-you-should-be-too/> Last accessed 13/03/2019.

Joyce, James (2005) How probabilities reflct evidence. *Philosophical Perspectives* 19, 153-178.

(2011) A defence of imprecise credences in inference and decision making. *Philosophical Perspectives* 24, 281-323.

Keefe, Rosanna (2000) *Theories of vagueness*. Cambridge: Cambridge University Press.

Knight, Frank (1921) *Risk, uncertainty, and profit*. Boston, MA: Hart, Schaffner, and Marx.

Kremer, Michael (1993) Population growth and technological change: one million BC to 1990. *The Quarterly Journal of Economics* 108, 681-716.

Lenman, James (2000) Consequentialism and cluelessness. *Philosophy and Public Affairs* 29, 342-370.

Levi, Isaac (1974) On indeterminate probabilities. *Journal of Philosophy* 71, 391-418.

Lorentzen, Peter, John McMillan, and Romain Wacziarg (2008) Death and development. *Journal of Economic Growth* 13, 81-124.

Lyon, Aidan (2017) Vague credences. *Synthese* 194, 3931-3954.

MacAskill, William (2015) *Doing good better*. London: the Guardian and Faber & Faber.

Mason, Elinor (2004) Consequentialism and the principle of indifference. *Utilitas* 16, 316-321.

McKeown, Thomas (1976) *The modern rise of population*. London: Edward Arnold.

Meadows, Dennis, Donella Meadows, Jørgen Randers, and William Behrens (1972) *Limits to growth*. New York, NY: Universe Books.

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

Murtin, Fabrice (2013) Long-term determinants of the demographic transition, 1870-2000. *Review of Economics and Statistics* 95, 617-631.

Nouri, Ali and Chyba, Christopher (2008) Biotechnology and biosecurity. In Bostrom and Cirkovic, eds. *Global catastrophic risks*, 450-480. Oxford: Oxford University Press.

Ord, Toby (2014) Overpopulation or underpopulation? In Goldin, ed. *Is the planet full*, 46-60. Oxford: Oxford University Press.

Pummer, Theron (2016) Whether and where to give. *Philosophy and Public Affairs* 44, 77-95.

Ramsey, Frank (1926) Truth and probability. In his *The foundations of mathematics and other logical essays*, 156-198. Abingdon: Routledge.

Rinard, Susanna (2015) A decision theory for imprecise credences. *Philosophers' Imprint* 15, 1-16.

Roodman, David (2014) The impact of life-saving interventions on fertility. <https://davidroodman.com/blog/2014/04/16/the-mortality-fertility-link/> Last accessed 14/03/2019.

Sahlin, Nils-Eric and Weirich, Paul (2014) Unsharp sharpness. *Theoria* 80, 100-103.

Sachs, Jeffrey (2003) Institutions don't rule: direct effects of geography on per capita income. NBER Working Paper No. 9490.

Savage, Leonard (1972) *The foundations of statistics, 2nd rev. ed*. New York, NY: Dover.

Schoenfield, Miriam (2012) Chilling out on epistemic rationality. *Philosophical Studies* 158, 197-219.

Schultz, Paul (1997) Demand for children in low income countries. In Rosenzweig and Stark, eds. *Handbook of population and family economics*, *volume 1, part A,* 349-430. Amsterdam: Elsevier.

Singer, Peter (2015) *The most good you can do: how effective altruism is changing ideas about living ethically.* New Haven, CT: Yale University Press.

Siniscalchi, Marciano (2009) Two out of three ain't bad: a comment on 'The ambiguity aversion literature: a critical assessment.' *Economics and Philosophy* 25, 335-356.

Shelton, James (2014) Taking exception: reduced mortality leads to population growth: an inconvenient truth. *Global Health: Science and Practice* 2, 135-138.

Smart, J. J. C. (1973) An outline of a system of utilitarian ethics. In Smart and Williams *Utilitarianism: for and against*, 3-75. Cambridge: Cambridge University Press.

Srinivasan, Amia (2015) Are we luminous? *Philosophy and Phenomenological Research* 90, 294-319.

Trautmann, Stefan, and van de Kuilen, Gijs (2015) Ambiguity attitudes. In Keren and Wu, eds. *The Wiley-Blackwell Handbook of Judgment and Decision Making*, 89-116. Oxford: John Wiley and Sons.

van Fraassen, Bas (1966) Singular terms, truth-value gaps, and free logic. *Journal of Philosophy* 63, 481-495.

(1990) Figures in a probability landscape. In Dunn and Gupta, eds. *Truth or consequences*, 345-356. Dordrecht: Springer.

von Neumann, John and Morgenstern, Oskar (1947) *Theory of games and economic behavior, 2nd ed.* Princeton, NJ: Princeton University Press.

Weirich, Paul (2004) *Realistic decision theory: rules for nonideal agents in nonideal circumstances*. Cambridge: Cambridge University Press.

Weatherson, Brian (1998) Decision making with imprecise probabilities. <http://brian.weatherson.org/vdt.pdf> Last accessed 14/03/2019.

White, Roger (2010) Evidential symmetry and mushy credence. *Oxford Studies in Epistemology* 3, 161-186.

Williamson, Timothy (2000) *Knowledge and its limits*. Oxford: Oxford University Press.

Willis, Amy (2012) Adolf Hitler 'nearly drowned as a child.' *The Telegraph* 06/01/2012. <https://www.telegraph.co.uk/history/world-war-two/8996576/Adolf-Hitler-nearly-drowned-as-a-child.html> Last accessed 14/03/2019.