

Uncertainty behind the Veil of Ignorance

A. FAIK KURTULMUS

Sabanci University

This article argues that the decision problem in the original position should be characterized as a decision problem under uncertainty even when it is assumed that the denizens of the original position know that they have an equal chance of ending up in any given individual's place. It supports this claim by arguing that (a) the continuity axiom of decision theory does not hold between all of the outcomes the denizens of the original position face and that (b) neither us nor the denizens of the original position can know the exact point at which discontinuity sets in, because the language we employ in comparing different outcomes is ineradicably vague. It is also argued that the account underlying (b) can help proponents of superiority in value theory defend their view against arguments offered by Norcross and Griffin.

I

Harsanyi used a device similar to Rawls's veil of ignorance to generate utilitarian conclusions.¹ Both Rawls and Harsanyi's veils of ignorance are intended to model the idea of impartiality. Behind both veils of ignorance, individuals deliberate on the assumption that they could end up in anyone's position. However, Rawls's veil of ignorance is in one respect thicker. Behind Rawls's 'thick' veil of ignorance, individuals do not know the likelihood that they may end up in any given person's position. Behind Harsanyi's 'thin' veil of ignorance, they know that they have an equal chance of being in any person's position. This difference has substantial consequences. Under the thin veil of ignorance we have a decision problem under risk, in which case the maximization of expected utility is the preferred solution, and we end up with some form of utilitarianism. Under the thick veil of ignorance, we have a decision problem under uncertainty, in which case maximin, along with several other decision rules, is a plausible candidate as a solution.² Therefore, whether we use a thick or thin veil of ignorance is of great importance.

The goal behind introducing the veil of ignorance is to ensure that when considering and choosing different principles one does not favour oneself. Since the denizens of the original position lack knowledge of their particular features they cannot choose principles tailored to

¹ John C. Harsanyi, 'Cardinal Utility in Welfare Economics and in the Theory of Risk-taking', *The Journal of Political Economy* 61 (1953), pp. 434–5.

² When the probabilities of different outcomes are known to the decision-maker we have a *decision under risk*. When the decision-maker lacks such probability information we have a *decision under uncertainty*.

themselves. However, to achieve this goal we don't need a thick veil of ignorance. The thin veil of ignorance is enough to ensure this kind of impartiality. Accordingly, Rawls seems to lack a rationale for employing the thick veil of ignorance, other than the fact that it produces the results he prefers. This makes the whole procedure question begging. As Parfit puts it:

Rawls himself points out that, since there are different contractualist formulas, he must defend his particular formula. This formula, he writes, must be the one that is 'philosophically most favoured', because it 'best expresses the conditions that are widely thought reasonable to impose on the choice of principles'. Could Rawls claim that, compared with the Equal Chance Formula [Harsanyi's formula], his No Knowledge Formula better expresses these conditions? The answer, I believe, is No. Rawls's veil of ignorance is intended to ensure that, in choosing principles, we would be impartial. To achieve this aim, Rawls need not tell us to suppose that we have no knowledge of the probabilities. If we supposed that we had an equal chance of being in anyone's position, that would make us just as impartial. Since there is no other difference between the Equal Chance and No Knowledge Formulas, Rawls's No Knowledge Formula cannot be claimed to be in itself more plausible.³

The aim of this article is to defend the characterization of the decision problem in the original position (hereafter the OP) as one of uncertainty rather than of risk even when we have a thin veil of ignorance. Before I present my defence, I'd like to set out its limits. First, on some interpretations of probability, there is no difference between risk and uncertainty, because it is believed that all probability is subjective. Here, I'm assuming that this is not the case. Second, I do not seek to defend the claim that maximin is the correct decision rule in the OP, or that the difference principle would emerge out of the OP. All I seek to establish is that the decision problem the denizens of the OP face should be characterized as one under uncertainty rather than risk.

There will be a lot going on in this article. Let me therefore present an intuitive road map for the reader. My central argument will be based on three claims. The first claim is that the continuity axiom of decision theory does not hold for all outcomes. Very roughly, the continuity axiom says that for all outcomes o , there is a gamble, which has a possible outcome better than o and another possible outcome worse than o , that is preferable to o for certain. By denying the continuity axiom, we are saying that for some outcomes possible gains over them do not justify taking gambles where it is also possible that we will end up worse-off – no matter how small the probability of the worse outcome is. My second claim will be that the denizens of the OP cannot know precisely for

³ Derek Parfit, *On What Matters*, 2 vols. (Oxford, 2011), vol. 1, pp. 350–1. Nagel also makes the same objection. See Thomas Nagel, 'Rawls on Justice', *The Philosophical Review* 82 (1973), pp. 220–34, at 11–12.

which outcomes continuity does not hold. My third claim will be that, given these first two claims, the decision problem in the OP should be interpreted as a decision problem under uncertainty rather than risk even if the denizens of the OP know that they have an equal chance of ending up in any given individual's position.

The bulk of this article is dedicated to defending the second claim. I defend this claim in the context of discussing objections to superiority in value theory. In the next section, I introduce and motivate superiority and present a very forceful objection voiced by Griffin and Norcross. In section III, I offer my diagnosis of their objections. I argue that their objection does not show what they intend it to show, namely that superiority ought to be rejected. Instead, it shows that we are not able to determine precisely where the superiority claims apply and do not apply when we are looking at the kind of cases they use in their examples. I argue that our failure is due to the ineradicable vagueness of the terms we use in making sense of and comparing different outcomes. In section IV, I present my case for rejecting the continuity axiom of decision theory and argue that my diagnosis of what is going on in Griffin and Norcross's examples is applicable there too. Just as we cannot determine where superiority sets in between different objects of value because the language we employ is vague, we also cannot determine where discontinuity sets in. In section V, I tie up the different threads of my argument. I show that the decision problem in the OP should be characterized as a decision problem under uncertainty, because the denizens of the OP do not know where precisely discontinuity sets in.

II

Standard accounts of value assume that for any objects of intrinsic value, a and b , which are positively valuable and where a is more valuable than b , there is some amount k such that $k b$ is more valuable than a . And in the case of objects of intrinsic disvalue, for any objects of intrinsic disvalue a and b where a is worse than b , there is an amount k such that $k b$ is worse than a . Some value theorists, however, have challenged this. They make the following claims.

Strong superiority (roughly): For some intrinsic goods a and b , any amount of a is better than any amount of b .

Weak superiority (roughly): For some intrinsic goods a and b , some amount of a is better than any amount of b .⁴

⁴ For a discussion of this distinction see Gustaf Arrhenius, 'Superiority in Value', *Philosophical Studies* 123 (2005), pp. 97–114.

There are many strong intuitions supporting superiority. For instance, it is very plausible to claim that ‘The badness of the premature death of one innocent person cannot be outweighed by the badness of any number of mild headaches’. Value theories that incorporate superiority would be able to accommodate such intuitively plausible value judgements. In addition to this, the incorporation of superiority would enable consequentialist theories to account better for our judgements about what we ought to do. We not only think that the badness of one death cannot be outweighed by some number of mild headaches, but also think that if we can save either one person from death or alleviate minor headaches we should pick the first option no matter how many headaches we can alleviate.

Some philosophers are sceptical of judgements of the goodness or badness of states of affairs. They deny that there is such a thing as a good outcome *simpliciter*. They, nevertheless, concede that the notion of good for someone is coherent. Those philosophers also have reason to consider superiority and challenges to it, because there are cases concerning a person’s good in which superiority is intuitively plausible. Consider the following example from Parfit:

Suppose that I can choose between two futures. I could live for another 100 years, all of an extremely high quality. Call this the Century of Ecstasy. I could instead live forever, with a life that would always be barely worth living. Though there would be nothing bad in this life, the only good things would be muzak and potatoes. Call this the Drab Eternity.

I believe that, of these two, the Century of Ecstasy would give me a better future.⁵

In sharing Parfit’s judgement that the Century of Ecstasy would give one a better future we are judging that some amount of life at a very high quality is better than any amount of life at a very low quality, and committing ourselves to superiority.

There is, then, a *prima facie* case in favour of superiority in both interpersonal and intrapersonal cases. However, there are also certain cases in which the commitment to superiority and other highly plausible claims results in a contradiction. One such case is due to Griffin. Griffin first makes a superiority claim:

Fifty years of life at a very high level of well-being – say, the level which makes possible satisfying personal relations, some understanding of what makes life worth while, appreciation of great beauty, the chance to accomplish something with one’s life – outranks any number of years at the level just barely worth

⁵ Derek Parfit, ‘Overpopulation and the Quality of Life’, *The Repugnant Conclusion: Essays on Population Ethics*, ed. J. Ryberg and T. Tännsjö (Dordrecht, 2004), pp. 7–22, at 17–18.

living – say, the level at which none of the former values are possible and one is left with just enough surplus of simple pleasure over pain to go on with it.⁶

He then starts a sequence in one dimension of value:

1. 'Fifty years at a very high level – say, with enjoyment of a few of the very best Rembrandts, Vermeers, and de Hoochs – might be outranked by fifty-five years at a slightly lower level – no Rembrandts, Vermeers, or de Hoochs, but the rest of the Dutch School.'

2. '[F]ifty-five years at that level might be outranked by sixty years at a slightly lower level – no Dutch School but a lot more of the nineteenth-century revival of the Dutch School.'⁷

We repeat these steps until all appreciation of beauty is lost, and we are left with kitsch objects. We then repeat similar sequences for other dimensions of value, which Griffin mentions in his superiority claim, at which point we end up contradicting the initial superiority claim.

Another similar sequence is offered by Norcross, who also observes that incorporating superiority would bring aggregationist principles more in line with our intuitions about what we ought to do. Norcross begins with the intuitively plausible superiority claim that was already mentioned:

Less: The state of affairs constituted by any number of fairly minor headaches is less bad than even one premature death of an innocent person.⁸

Norcross next introduces the following comparison, which we also think is true:

One death is less bad than n , $n > 1$, mutilations.

He now takes the second term of the previous comparison and compares it with some other misfortune that is less painful but suffered by more people:⁹

n mutilations are less bad than m , $m > n$, painful experiences that are less painful than mutilations.

We repeat this procedure where we compare the second element of each previous comparison with another less painful experience suffered by a greater number of people until we arrive at the following comparison:

y mild ankle sprains are less bad than z headaches.

⁶ James Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance*, (Oxford, 1986), p. 86.

⁷ Griffin, *Well-Being*, p. 86.

⁸ Alastair Norcross, 'Comparing Harms: Headaches and Human Lives', *Philosophy and Public Affairs* 26 (1997), pp. 135–67, at 137.

⁹ My formulation of Norcross's sequence is slightly different from the original.

Given the transitivity of better than, we are now forced to conclude that ‘one death is less bad than z headaches’ and contradict Less.

III

In both Norcross’s and Griffin’s examples, we begin with an intuitively plausible superiority claim of the form: ‘No amount of Z is better than any, or some, amount of A ’. Nevertheless, we find ourselves willing to grant a claim of the form ‘Some amount of B , which is only in slightly inferior to A in the aspect that’s relevant to our valuation, is better than A ’. Through similar steps involving gradual diminishments we arrive at the claim: ‘Some amount of Z , which is slightly inferior to Y in the aspect that is relevant to our valuation, is better than Y ’. Then, by transitivity we are required to concede that there is an amount of Z better than A , thereby contradicting our initial superiority claim. We have three ways to resolve the inconsistency: (a) we can reject the superiority claim; (b) we can reject transitivity; (c) we can reject one of the comparisons we make at one of the steps of the sequence. Both superiority and transitivity are intuitively very plausible. So, the best candidate for rejection is one of the comparisons we make.¹⁰

The sequences presented by Norcross and Griffin work similarly to the sorites paradox. A sorites paradox takes us from a premise, which we take to be true, through an inductive premise, or repeated applications of a conditional, which we again hold to be true, to a conclusion, which we take to be false. The paradox of the heap, which gives the sorites paradox its name, has the premise ‘1 grain of sand does not make a heap’, the inductive premise ‘For any n , if n grains of sand do not make a heap then $n + 1$ grains of sand do not make a heap’, and the conclusion, depending on where we stop, ‘100,000 grains of sand do not make a heap’. The argument is valid, the premises seem true, but the conclusion seems false. The paradox arises owing to the vagueness of the predicate ‘heap’.

Vague predicates, like ‘tall’, ‘red’, ‘heap’ and ‘child’, have borderline cases. Borderline cases are cases in which we cannot decide whether or not a predicate applies. For instance, we may not be able to decide whether someone is a child or not even though we know every observable fact about her. Another feature of vague predicates is their

¹⁰ My argument in this section owes a great debt to Gustaf Arrhenius and Wlodek Rabinowicz, ‘Value and Unacceptable Risk’, *Economics and Philosophy* 21 (2005), pp. 177–97; Mozaffar Qizilbash, ‘Transitivity and Vagueness’, *Economics and Philosophy* 21 (2005), pp. 109–31; Richard Tuck, ‘Is There a Free Rider Problem?’, *Rational Action: Studies in Philosophy and Social Science*, ed. R. Harrison (New York, 1979), pp. 147–56.

lack of sharp boundaries.¹¹ There isn't a sharp boundary which divides objects falling under a vague predicate and objects which do not fall under that predicate. For instance, there isn't a determinate age beyond which someone is no longer a child. A final feature of vague predicates is their susceptibility to Sorites paradoxes. Small differences in the age of a person do not make a difference with regard to the appropriateness of calling them a child. Accordingly we assent to the conditional premise ('If someone n years old is a child so is someone n years and a day old') that generates the Sorites paradox.¹²

Our assent to each comparison in Norcross–Griffin sequences is due to the vagueness of the terms we rely on in thinking about the different comparisons. This is, roughly, how vagueness affects our judgements in Norcross–Griffin sequences. The goodness or badness of some outcomes cannot be directly perceived, or judged, by us. That is, the badness or goodness of all outcomes, and how they compare to other outcomes, cannot always be read off just by being presented with them. We need to rely on redescrptions of outcomes which reduce them to more basic judgements. This process of redescription can be carried out consciously or unconsciously. Either way, these redescrptions contain vague predicates. Accordingly, we are unable to identify the specific points at which the predicates in our redescrptions no longer apply. As a result, we feel compelled to assent to the trade-offs in the consecutive steps of Norcross–Griffin sequences.

The redescription process can take place in two different ways. I think both processes actually take place, but either of them taking place is enough for my argument. On the first view, we think about the options we are comparing by consciously redescrbing them in natural language. We do this in order to make the options vivid and establish connections between the outcomes under consideration and our more basic judgements. This is supported by attending to the way we make actual comparisons. When I think about the comparisons in Norcross's sequence, I find myself formulating alternative descriptions of the options like the following: 'Very painful', 'Unbearably painful', 'Disabling', 'Impossible to ignore', 'Easy to ignore', 'One could carry on as usual despite it', etc.¹³ On the second view, the redescrptions do not take place in natural language, but in an innate language of thought,

¹¹ Epistemicists deny this. They argue that vague predicates have well-defined extensions and sharp boundaries, but their extensions are unknowable to us. For a defence of epistemicism see Timothy Williamson, *Vagueness* (London, 1994).

¹² In this paragraph, I'm following Rosanna Keefe and Peter Smith, 'Introduction', *Vagueness: A Reader*, ed. R. Keefe and P. Smith (Cambridge, MA, 1997), pp. 1–57, at 2–3.

¹³ Qizilbash makes a similar observation. See Qizilbash, 'Transitivity and Vagueness', pp. 119–20.

sometimes called *mentalese*, that functions as the medium of thought.¹⁴ Mentalese is like natural languages in that it employs words and these words combine to form mental sentences. Since the language of thought has to be translatable into natural languages – otherwise we wouldn't be able to express our thoughts or understand the thoughts of others – it has to contain vague predicates.¹⁵

On both views, we do not need to identify a single term that should bear the blame for the introduction of vagueness into our re-descriptions. It is unlikely that we rely on a single term when trying to make the alternatives we are comparing vivid and reason about them. What is more likely is that there are many terms we use in describing the alternatives which are susceptible to borderline cases.

It is useful to state explicitly the main claims of this account. First, when judging whether an option is better than, or superior to, another, we need to represent it mentally. Second, this medium of representation is a language. This language may be a natural language or may be mentalese. Third, this language, which acts as the medium of thought, contains vague predicates. Fourth, some of our value judgements are derived from more basic value judgements. Sometimes when deciding whether *A* is better than, or superior to, *B* our judgement will be inferential. It will have, roughly, the form *A* possesses properties *a*, *b*, *c* . . . ; *B* lacks these properties or possesses fewer of them; hence *A* is better than or superior to *B*. I do not claim that this process of reasoning needs to be carried out consciously.

Let us now illustrate this account using Griffin's example as our basis. For ease of exposition, let us assume that the only relevant value is the enjoyment of the arts and beauty. Then, the basic superiority claim is that: 'Fifty years of life spent in enjoyment of the arts and beauty is better than any duration of life spent in enjoyment of kitsch objects'. On my account, to evaluate the superiority claim and to make the comparisons in Griffin's sequence, we need to make it vivid and spell out 'enjoyment of art', 'experience of beauty' and 'enjoyment of kitsch'. Let's say the predicates n_1, n_2, \dots, n_m are the ones we use in our account of 'enjoyment of art' and 'experience of beauty'. Some of these predicates will be vague and lack sharp boundaries. As a result, we won't be able to determine the precise cut-off point when an object is no longer a work of art, or when our experience cannot be characterized as the experience of beauty.

¹⁴ The *locus classicus* is Jerry A. Fodor, *The Language of Thought* (Cambridge, MA, 1975).

¹⁵ For the argument that the language of thought has to be vague see Roy A. Sorensen, 'Vagueness within the Language of Thought', *The Philosophical Quarterly* 41 (1991), pp. 389–413.

Since $n_1 \dots n_m$ contain vague predicates, which lack sharp boundaries, we are unable to tell at which specific point $n_1 \dots n_m$ no longer apply to given objects. We are committed to the superiority of a life spent in enjoyment of the arts to a life spent in enjoyment of only kitsch objects. That is, we are committed to the claim that a life that $n_1 \dots n_m$ are applicable to is superior to one that they are not applicable to. We are also willing to make trade-offs between quality and quantity as long as we are in the domain of beauty. At each step of Norcross–Griffin sequences, we ask whether $n_1 \dots n_m$ do apply to the objects under consideration. If they do, then we are willing to make trade-offs. Since, $n_1 \dots n_m$ contain vague predicates, there isn't a sharp cut-off point where we can say $n_1 \dots n_m$ no longer apply, even though $n_1 \dots n_m$ clearly apply to some objects, and clearly do not apply to others. Since our fundamental superiority claim is that the objects $n_1 \dots n_m$ apply to are superior to those $n_1 \dots n_m$ do not apply to, we cannot determine the precise point at which we should stop making trade-offs. For this reason, we find ourselves accepting the trade-offs offered at each consecutive step of Griffin–Norcross sequences.

The role of redescrptions in making the comparisons in Norcross–Griffin sequences confirms a point made by Dorsey in his insightful discussion of Norcross's sequence. Dorsey claims that 'the fulfilment of deliberative projects is strongly superior to hedonic values, though both are important for a complete account of human welfare'.¹⁶ He then offers the following diagnosis of Norcross's sequence. Norcross's sequence goes wrong in focusing only on one dimension of value: the intrinsic badness of physical discomfort *qua* physical discomfort. This is a mistake, because physical discomforts at certain levels 'become instrumentally injurious to our deliberative projects'.¹⁷ So, it is at the point where deliberative projects, which are superior to hedonic values, become impossible to carry out that the superiority claim kicks in.

This account and the discontinuity claim Dorsey offers are I believe correct. (In fact, I shall rely on a similar claim in the next section.) And it is confirmed by attending to how we redescribe the options in Norcross's sequence. Recall some of the terms we used in our descriptions: 'Unbearably painful', 'Disabling', 'Impossible to ignore', 'Easy to ignore', 'One could carry on as usual despite it'. They point to how the physical pain relates to one's life projects. So, the redescription process not only helps us get a handle on the magnitude of pains, but also directs our attention to the place of pains within a wider network of values. On Dorsey's proposal, we have a sharp cut-off point – the point at which physical pain becomes injurious to our deliberative projects.

¹⁶ Dale Dorsey, 'Headaches, Lives and Value', *Utilitas* 21 (2009), pp. 36–58, at 46.

¹⁷ Dorsey, 'Headaches, Lives and Value', p. 49.

However, if our account of Norcross–Griffin sequences is correct, where that point is on Norcross’s sequence cannot be specified with precision owing to the vagueness of the terms we rely on to spell out life projects, physical discomforts and the relationship between them.

I should point out that Dorsey’s proposal addresses a different puzzle from the one I’m addressing. I’m addressing the puzzle about why we would *assent* to the claims in each step of Norcross–Griffin sequences. Dorsey is addressing a puzzle about how superiority can arise when the only aspects of the objects under consideration that seem to be relevant to their value can be placed on a continuum.¹⁸ More fully stated, the objection Dorsey is addressing is the following: ‘If value is a function of discomfort or the intensity of pain of different experiences and these can be placed on a continuum how can there be jumps in value that superiority entails? If there is such a continuum wouldn’t we be imposing an arbitrary cut-off point when we make superiority claims?’ Dorsey’s response to this objection is that the objection rests on a mistaken assumption. The pains or discomforts have not only intrinsic disvalue but also instrumental disvalue. After a certain point, the physical pains or discomforts on Norcross’s sequence would hurt our deliberative projects and it is the deliberative projects that are the superior goods. Dorsey is suggesting that Norcross’s challenge draws support from a mistaken theory of value. This is true, but doesn’t explain all that we need to explain. Even if we revise our theory of value along the lines suggested by Dorsey, we would not be able to identify the point in Norcross’s sequence at which superiority kicks in. It does not explain why each individual comparison would seem intuitively plausible. Suppose we say with Dorsey that the sequence goes wrong at the step when deliberative projects are traded for other goods. Can we identify the precise point in the sequence at which this trade off first occurs? My suggestion is that we cannot, because of the vagueness of the terms we employ. Even if we know all the relevant facts about how pain or physical discomfort impacts on a person’s ability to pursue deliberative projects, and we have a fully worked out philosophical account of what counts as a deliberative project, there will be vagueness in the terms our analysis employs. As a result, we will be unable to identify a certain comparison in Norcross’s sequence as mistaken.

IV

With our diagnosis of what is going wrong in Norcross–Griffin sequences to hand we can now turn back to the OP. The denizens

¹⁸ I departed from Norcross’s original formulation to emphasize this point.

of the OP are deliberating in a probabilistic framework. They are not ranking outcomes that are certain. Instead they are considering possible outcomes with differing likelihoods. Accordingly, superiority isn't directly relevant to the decision problem in the OP. Nevertheless, there's a related thesis that is relevant: the continuity axiom of decision theory. Its denial is also susceptible to the kind of problems owing to vagueness to which superiority was susceptible. According to the continuity axiom, for all outcomes a , b , and c , where $a > b > c$, there is a probability p , $0 < p < 1$, such that $pa + (1 - p)c > b$.¹⁹ Intuitively, continuity says that for all outcomes, o , there is a lottery containing an outcome better than o and another outcome worse than o that should be preferred to o for certain.

Continuity has often been defended as a technical requirement. Preferences that do not satisfy the continuity axiom cannot be represented by a real-valued function.²⁰ However, it has substantive consequences and has not gone unchallenged. There are various counterexamples to the continuity axiom offered in the decision theory literature.²¹ Here is one common example. Suppose there are three outcomes. Outcome a gives you \$1000; outcome b gives you \$10; and in outcome c you are killed. Continuity requires that there is a probability, p , $0 < p < 1$, such that $pa + (1 - p)c > b$. That is, one ought to be willing to run a small risk of getting killed to end up with \$1,000 instead of \$10. This is highly counter-intuitive.

There's also a standard way of dealing with such objections. Kreps offers it right after presenting the previous putative counterexample to continuity.

¹⁹ Here's a brief illustration of why superiority does not entail the rejection of the continuity axiom. Suppose we have three intrinsically valuable goods a , b , and c with values 10, 3, and 1. Suppose that after the first unit of c , the value of each additional n^{th} unit of c is 2^{-n} . In this case, b is superior to c . No matter how many units of c we have their value will be less than 3. However, if there's a lottery giving us a with probability 0.5 and c with probability 0.5 we should pick this lottery over b for certain. We have superiority, but continuity isn't violated.

²⁰ Preferences that do not satisfy the continuity axiom but are complete and transitive, and satisfy the independence axiom, can be represented by assigning a vector of expected utilities to each gamble where these vectors are ordered lexicographically. It's worth noting that attributing a lexicographic value function to the denizens of the OP doesn't entail that the principle which emerges out of the OP will be lexicographic too. Or more precisely, the rejection of continuity gives us lexically ordered principles, but it doesn't ensure that the principles themselves are lexical like the difference principle. For an overview of lexicographic decision theory, where the continuity axiom is dropped, and some applications, see L. Blume, A. Brandenburger and E. Dekel, 'An Overview of Lexicographic Choice under Uncertainty', *Annals of Operations Research* 19 (1989), pp. 231–46.

²¹ A. A. Alchian, 'The Meaning of Utility Measurement', *The American Economic Review* 43 (1953), pp. 26–50, at 36–7; J. S. Chipman, 'The Foundations of Utility', *Econometrica* 28 (1960), pp. 193–224, at 221; R. D. Luce and H. Raiffa, *Games and Decisions* (New York, 1989), p. 27.

Suppose I told you that you could either have \$10 right now, or, if you were willing to drive five miles (pick some location five miles away from where you are), an envelope with \$1000 was waiting for you. Most people would get out their car keys at such a prospect, even though driving the five miles increases ever so slightly the chances of a fatal accident. So perhaps the axiom is not so bad normatively as may seem at first sight.²²

We can make two preliminary responses. First, we should observe that Kreps's second example shows only that it might not be irrational to take the gamble that was offered in his first example. It does not establish the stronger claim that continuity is required by rationality. Second, we ought to take it as a warning that we should not be too hasty in judging that continuity does not hold between certain outcomes. All that the example shows, if it shows anything, is that the rejection of continuity in Kreps's first example was over-hasty.

Nevertheless, there's something puzzling. What explains the divergence of our intuitions in these two cases? Kreps's second example transforms the first example into the kind of choice we face in everyday life. We, of course, need to do several things that might increase the chance of us dying in order to pursue our goals, make a living, etc. It is implausible to deny that we ought to take such risks. Clarifying the nature of the discontinuity claim behind our rejection of trade-offs between life and money can help us deal with Kreps's second example and address the kind of risks we take in everyday life. The key point is that continuity between life and money should be rejected conditionally. It is only if you are guaranteed a certain level of income and wealth that continuity breaks down. After all, we want to live longer for the values life embodies. For the enjoyment and attainment of these values – and more fundamentally, for sustenance – a certain level of wealth is a prerequisite. The rejection of continuity between life, extra income and death, which guides our response to Kreps's first example, should be interpreted as the rejection of continuity between a certain kind of life (one that embodies certain values which are only possible given some level of income and some amount of risk taking), extra income and death. For those who are committed to the conditional rejection of continuity between life and money, one *does* need to take risks that will increase the risk of death when those risks are necessary for securing the conditions that give life its value or come with activities that are a part of a good life. It isn't the duration of time one spends alive but the goods a life embodies that matter. Cases like Kreps's second example elicit intuitions about such choices where we need to take risks to secure the goods that are a precondition for a good life. We need to distinguish between the risks that are part and parcel of our efforts

²² D. M. Kreps, *Notes on the Theory of Choice* (Boulder, CO, 1988), pp. 45–6.

to lead a good life, and those that are not. Both of Kreps's examples lack the necessary details for us to determine whether the conditional rejection of continuity applies to them. In these examples, we don't know whether the risk is one we need to take as part of a plan of actions aimed at living a decent life or not. Since the act involved in Kreps's second example is similar to our everyday choices we fill out the details of the case differently from his first example. The conditional nature of the discontinuity claim and the different way we fill out the cases account for the divergence in our intuitions.

What is the discontinuity claim relevant to the OP? The denizens of the OP are examining principles that will determine their long-term expectations of primary goods. Each social position that results from the application of a principle they choose entails, in effect, a certain life prospect. Overall life prospects are cases in which the possibility of discontinuity is highly plausible. That is, it is more plausible to expect discontinuity between lives taken as a whole rather than between experiences of shorter duration. The basic, and admittedly very rough, idea is that continuity does not hold when one option is a life in which *S* is guaranteed the ability to do something with his life, and the other option offers either (a) such a life plus some further goods, or (b) a life in which *S* cannot do something with his life, even if this latter possibility is very unlikely. It is carrying out plans of life and projects we think worth pursuing and forming deep personal relations that give life meaning. A life containing these values is to be preferred to a life in which they are absent, even though a life in which they are largely absent may, nevertheless, be one worth living. Rawls's primary goods are instrumental to us carrying out our plans of life, and with more primary goods we can pursue our plans and projects with greater ease or pursue more ambitious plans and projects. However, the possibility of having more primary goods does not justify taking gambles where the other possibility is losing the capacity to pursue plans and projects, and to form deep personal relations.

The account of how vague predicates figure in our evaluation of different alternatives that we presented in section III is applicable to the decision problem in the OP, because the outcomes under consideration are very complex and we can get a handle on them only by re-describing them. There are many points at which vagueness can creep into these descriptions. Accordingly, if continuity does not hold for all of the outcomes in the OP we will not be able to tell the precise point at which discontinuity sets in.

I have relied on a very intuitive grasp of the idea of doing something with one's life and I have not offered a philosophical analysis of this idea. Given this, we would say that my claim about doing something with one's life is vague. However, this is not the sense of vagueness that

is pertinent. In the absence of a philosophical analysis of what doing something with one's life amounts to, we have vagueness in the sense of under-specificity. This sense of vagueness does not have anything to do with borderline cases. The sense of vagueness that is relevant to our argument is vagueness in the sense of having borderline cases in which it is unclear whether the predicate applies or not and the predicate is susceptible to a sorites sequence. With a good philosophical analysis, we can get rid of vagueness in the sense of under-specificity, but we cannot get rid of linguistic vagueness. Even our best analysis will involve vague predicates that have borderline cases.²³ If even our best analysis of the concepts that figure in our discontinuity claim will contain vague predicates, we won't be able to identify precisely the point after which a person can be said to be not pursuing plans and projects that give a life meaning.

We are justified in imputing the inability to know precisely where discontinuity sets in to the denizens of the OP, because our inability to know precisely where discontinuity sets in is due to the *ineradicable* vagueness of natural languages. This shortcoming is not something we can overcome with better reasoning and philosophical analysis. In this respect, it is crucially different from arguments like the following: 'In our reasoning, we are prone to make errors in probabilistic reasoning. We should, therefore, impute these shortcomings to the denizens of the OP'. These shortcomings are ones that we can overcome. Since the original position is supposed to do justificatory work, it cannot incorporate such shortcomings. The OP makes a connection between individual rationality and justice. When doing that it has to rely on the best account of rationality on offer and correct for the ways in which actual people fall short of that ideal. When these shortcomings are ones that cannot be corrected for, the next best thing is an account that works with an awareness of these shortcomings. This is why it is legitimate to impute our inability to know the precise point where discontinuity sets in to the denizens of the OP.

V

Let us, finally, apply the account of discontinuity we have developed to the decision problem in the OP. We begin by noting an ambiguity in discussions of the decision problem in the OP. The decision problem in the OP is often discussed as if primary goods were a theoretical construct analogous to utility. Maximinng one's expectation of

²³ We can get rid of vagueness in this sense by introducing a new vocabulary that does not contain vague predicates. However, that will mean eliminating our existing concepts and replacing them with other ones.

primary goods and maximin one's expectation of utility are treated as being equivalent. This way of looking at the decision problem in the OP brings into sharp relief the difference between Rawls's principles and utilitarianism. However, this assumed equivalence is misleading. Rawls, in fact, attributes a certain utility function to the denizens of the OP, which determines their valuation of primary goods, and he relies on this function in his argument for his principles of justice.²⁴ (I shall refer to this function as a value function to distinguish it from an individual's subjective utility function.)

These are some of the ways in which Rawls specifies the value function of the denizens of the OP. Its denizens know that 'they normally prefer more primary social goods rather than less'.²⁵ They value the interests protected by Rawls's first principle more than they value those promoted by his second principle.²⁶ A person in the OP has 'a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following maximin' and 'the rejected alternatives have outcomes that one can hardly accept'.²⁷

Once we acknowledge that the denizens of the OP have a value function, we can distinguish between two different characterizations of the decision problem they face. The decision problem that the denizens of the OP face can, initially, be characterized as it is in table 1. An outcome p , which is given in terms of primary goods, corresponds to each pair consisting of a principle of distribution and an individual position.

Since there is a value function which determines the value of each outcome for the denizens of the OP, the decision problem needs another step, which transforms outcomes in terms of primary goods to values. Once the values are known, the decision problem will be as presented in table 2.

Once this intermediate step becomes apparent, the question of the nature of the value function that takes us from primary goods to value arises. Given that we are deciding on the value function of the denizens of the OP, we can specify the function in such a way that maximin one's expectation of primary goods is the correct decision rule even under conditions of risk. That is, we can specify the value function in such a way that the outcome which maximizes expected value is the one which maximins primary goods. Rawls notes that this would be the

²⁴ Rawls concedes that this is the case. See John Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA, 2001), p. 107.

²⁵ John Rawls, *A Theory of Justice*, rev. edn. (Oxford, 1999), p. 123.

²⁶ Rawls, *Theory*, p. 131

²⁷ Rawls, *Theory*, p. 134.

Table 1. *The primary goods characterization of the decision problem*

| | | Individual positions | | | | |
|------------|-------|----------------------|----------|---|---|----------|
| | | i_1 | i_2 | . | . | i_n |
| Principles | A_1 | p_{11} | p_{12} | . | . | p_{1n} |
| | A_2 | p_{21} | p_{22} | . | . | p_{2n} |
| | . | . | . | . | . | . |
| | . | . | . | . | . | . |
| | A_n | p_{n1} | p_{n2} | . | . | p_{nn} |

Table 2. *The value characterization of the decision problem*

| | | Individual positions | | | | |
|------------|-------|----------------------|----------|---|---|----------|
| | | i_1 | i_2 | . | . | i_n |
| Principles | A_1 | u_{11} | u_{12} | . | . | u_{1n} |
| | A_2 | u_{21} | u_{22} | . | . | u_{2n} |
| | . | . | . | . | . | . |
| | . | . | . | . | . | . |
| | A_n | u_{n1} | u_{n2} | . | . | u_{nn} |

case if the value function has a sharp bend at the point guaranteed by maximin.²⁸ Of course, specifying the value function in this way would make the OP effectively redundant.

What we need is a way of specifying the value function of the denizens of the OP that is acceptable on its own, and does not make the argument from the OP trivial. I have so far argued that (i) continuity must be rejected, and (ii) we, and a fortiori the denizens of the OP, cannot know where discontinuity sets in. If these claims are correct, then there is a way of specifying the value function which is acceptable on its own, and justifies the imputation of uncertainty rather than risk to the OP. Here is why given these two claims, the decision problem in the OP should be characterized as one under uncertainty.

²⁸ For Rawls's discussion of this point and a graph illustrating this possibility see Rawls, *Justice as Fairness*, pp. 107–9. There are a few other steps to his argument which I've ignored here.

Suppose the denizens of the OP know that they have an equal chance of ending up at any individual position. Also assume that they know their society's level of economic development.²⁹ This does not suffice for them to know which principles are preferable. To know this, they need to consult the value function. As I've argued, the value function we attribute to the denizens of the OP must violate the continuity axiom. This entails that the denizens of OP should reject certain lotteries. Furthermore, we do not know where precisely discontinuity sets in. There is, therefore, an uncertainty as to which lotteries the denizens of OP should reject. Unlike Rawls's OP, in this case, the uncertainty is due to the nature of the value function and our inability to determine where precisely discontinuity sets in. The denizens of the OP know that they have an equal chance of ending up as any individual, but they do not know exactly how each outcome maps on to each value (the mapping of each p_{ij} to each v_{ij}). More precisely, they know that there are certain outcomes which they should avoid, lotteries they should reject, but they do not know precisely which outcomes and lotteries these are.

The situation of the denizens of the OP can be likened to the following. Suppose you are ill and unless you receive treatment you will die. You have been told that a small amount of Drug A, which is a powdered substance, can cure you. If you take less than the required dose of Drug A you will die. However, there are no side-effects if you take too much of it. You know how much a gram of Drug A costs. There are three lotteries you can choose from. Each lottery has three possible outcomes all of which are equally likely. The money you may win from these lotteries is your only source of money obtaining drug A. The first lottery has as its prizes £60, £50 and £40. The second lottery has as its prizes £80, £80 and £10. The third lottery has as its prizes £110, £30, and £10. Which lottery should you choose?

Viewed in one way, in this example, there's no uncertainty about the outcomes. You are in a position to calculate the expected values, in money terms, of the lotteries. However, there's uncertainty as to which lottery gives you the greatest chance of receiving the treatment you need, which is your primary concern. If for instance, the required dose can be bought for an amount between £40 and £10, then the first lottery gives you the greatest chance. If the required dose costs £100, or more, the third lottery gives you the greatest chance. Given the vagueness of the instructions you have been provided you don't know which of the outcomes gives you the greatest chance of receiving treatment. (There is also an issue regarding the comparison class. To set this aside, we

²⁹ This assumption departs from Rawls's characterization of the veil of ignorance. I make this assumption to emphasize the robustness of the argument we're developing here.

can suppose that you have been in similar contexts before and you can figure out the comparison class.)

In this example, you have probability estimates for the likelihood of ending up with any given amount of money. So, if we focus on money, you face a decision problem under risk. Similarly, the denizens of the OP, who know they have an equal chance of being anyone, do not face uncertainty regarding the outcomes in terms of primary goods – they know the probability of ending up with any given amount of primary goods. Nonetheless, there's uncertainty about the value of the outcomes. In both cases, there's a threshold that the decision-maker would like to pass, but does not know where precisely that threshold is. Hence the uncertainty.

Having presented the intuitive argument for the characterization of the decision problem in the OP as one under uncertainty – despite the availability of probability estimates – we can present a more formal account. Let us suppose that the denizens of the OP are behind a thin veil of ignorance like Harsanyi's. They know the probability of ending in any single position. Suppose there are two principles of distribution, *A* and *B*, being considered by the denizens of the OP. Let $o_1 \dots o_n$ denote different amounts of primary goods individuals can end up with, where every element in the sequence contains progressively lower amounts of primary goods such that $o_1 > o_2 > \dots > o_n$.

Suppose we think that continuity does not hold between o_2 and o_n . There is no probability p , $0 < p < 1$, such that $po_1 + (1 - p)o_n > o_2$. That is, gains above o_2 do not justify taking gambles where one of the outcomes is o_n no matter how small the likelihood of ending up with o_n is. Let's say we are considering two principles of distribution. One of them guarantees o_2 . Under the other principle we may end up with either o_1 or o_n . Given the existence of discontinuity between o_2 and o_n , the denizens of the OP should pick the first principle no matter how small the probability of ending up with o_n under the second principle is.

Since we cannot determine precisely where discontinuity sets in, we can construct a Norcross–Griffin sequence that will lead us to contradict this discontinuity claim by making us agree to trade-offs on consecutive outcomes on the sequence. Suppose we are asked whether there's a probability p , $0 < p < 1$, such that $po_1 + (1 - p)o_3 > o_2$. Our response is that there is such a probability. We then repeat this procedure for each consecutive outcome, and end up contradicting our initial discontinuity claim. If our account of the real lesson of the Norcross–Griffin sequence is correct, there is a point where one of these comparisons is wrong, but we do not know which. Let's call the outcome where discontinuity actually sets in o_c . Let p_{ai} , $1 > p_{ai} > 0$, represent the probability of ending up with some outcome, o_i , under principle *A*. Let p_{bi} , $1 > p_{bi} > 0$,

represent the probability of ending up with some outcome, o_i , under principle B . Since we do not know where o_c is we do not know the values of $\sum_{i=1}^c p_{ai}$, $\sum_{i=c}^n p_{ai}$, $\sum_{i=1}^c p_{bi}$, and $\sum_{i=c}^n p_{bi}$, even though we know the values of each p_{ai} and p_{bi} . In other words, we do not know the probability of ending up above or below the point where discontinuity sets in under these two principles even though we know the probability of ending up with any given share of primary goods under both principles. Under the description of the decision problem that matters, we have a decision problem under uncertainty.

VI

In this section, I'd like to consider a possible solution to the problem I've been examining and show its weakness. The basic idea I've been labouring is this:

1. Continuity does not hold when one option is a life in which S is guaranteed the ability to do something with her life, and the other option offers either (a) such a life plus some further goods, or (b) a life in which S cannot do something with her life, even if this latter possibility is very unlikely.
2. This claim should be unpacked as: continuity does not hold when one option is a life that has the set of properties P , and the other option offers either (a) such a life plus some further goods, or (b) a life which lacks the set of properties P , even if this latter possibility is very unlikely.
3. Given the vagueness of our concepts we are unable to tell the exact cut-off point where a life no longer possesses the set of properties P .
4. As a result, we are unable to tell where discontinuity sets in.
5. The same point applies to the denizens of the OP. Denizens of the OP know that continuity does not hold between certain outcomes. However they do not know where the cut-off point is. For this reason their decision can be characterized as one under conditions of uncertainty even though they have all the relevant information that we can provide them.

The proposed solution I'd like to consider, which takes its cue from supervaluationist accounts of vagueness, is as follows.³⁰ Our concepts may be vague, but there are cases in which we are able to give a definite answer as to whether an object falls under a certain concept. For instance, we can classify some collections of sand grains as definitely

³⁰ For a helpful discussion of supervaluationism see Rosanna Keefe, *Theories of Vagueness* (Cambridge, 2000).

being heaps, some as definitely not being heaps and some as being neither definitely heaps nor definitely not heaps. In the latter cases, the claim that the vague predicate applies to an object lacks a truth value; it is neither true nor false. On these views, then, we have a threefold classification. The proposal I'm considering takes this threefold classification and applies it to the OP. Some outcomes are definitely above the threshold where discontinuity sets in, some are definitely below the threshold and some are neither definitely above the threshold nor definitely below the threshold. The denizens of the OP, behind the thin veil of ignorance, can make use of this threefold classification, and thereby remove the uncertainty in the decision problem they face. Since they know which outcomes are definitely above the threshold, they can determine which principles will enable them to end up definitely with an outcome over the threshold, and what the probabilities for these outcomes are.

As I indicated before, I wish to remain non-committal about how to solve the decision problem in the OP as I have characterized it. My aim is to defend a characterization of the decision problem as one under uncertainty. So, what I want to consider is whether this proposal can eliminate all uncertainty. The answer is that it cannot, because there is higher-order vagueness. We've seen that vague predicates lack sharp boundaries and admit of borderline cases. Similarly, 'cases that are borderline' lack sharp boundaries; there is no sharp boundary between definitely *F* and neither definitely *F* nor definitely not-*F*. Just as there isn't an *n*th second of a person's life after which we'll say 'He's no longer a child', there isn't an *m*th second in a person's life after which we'll say 'He is no longer *definitely* a child, from this moment on he's a borderline case'. The same point applies in the OP. Even though there will be outcomes which we think are definitely above the threshold, we will not be able to specify the exact point above which outcomes are definitely above the threshold.

Another objection that we need to consider is the following: 'Fine, there's vagueness, and higher-order vagueness, and because of these we don't precisely know which outcomes are (definitely) above the threshold or which ones are (definitely) below the threshold. Nevertheless, we do know that some outcomes are definitely above the threshold. So, why shouldn't the denizens of the OP set one such outcome as the guaranteed minimum and treat the decision problem for outcomes above it as one under risk?' As I have mentioned, I don't seek to defend a certain decision rule, but the whole article would be pointless if, despite all I have said, the correct decision rule was maximizing expected utility. Therefore, I need to show why adopting this proposal would take us away from maximizing expected utility. The proposal does not say that we should set the guaranteed

minimum at the point with the *least* amount of primary goods that guarantees being over the threshold. I have maintained that we can't know which outcome this is given the existence of first-order and higher-order vagueness. The proposal we're considering suggests setting the minimum at a point where we are certain that we are above the threshold. This is an important concession, because when the guaranteed minimum is not set at the least amount that is necessary for going above the threshold we are not maximizing expected utility. Suppose that the point at which discontinuity actually sets in is lower than where we have set our guaranteed minimum. Then, by lowering this guaranteed minimum and treating the decision problem above it as one of risk, we could increase the expected utility. Accordingly, if we adopt this proposal we would have departed from maximizing expected utility.

Even though this proposed solution is unable to remove the vagueness and uncertainty in the decision problem in the OP, it suggests one way in which we can enrich the information available to the denizens of the OP. Even though they do not know where discontinuity sets in, they may have a rough estimate. For instance, they may know that o_c , the point discontinuity sets in, is closer to o_n than it is to o_1 , or that it is closer to o_1 than it is to o_n . Information like this can offer considerations in favour of specific decision rules.

VII

Critics of the OP, such as Nagel and Parfit, have queried whether Rawls has non-question-begging grounds to opt for the thick veil of ignorance instead of the thin veil of ignorance and thereby transform the decision problem in the OP from one under risk to one under uncertainty. In this article, I have not offered reasons for adopting the thick veil of ignorance. Instead, I offered reasons for the characterization of the decision problem in the OP as one under uncertainty even if the thin veil of ignorance is employed. According to the account offered here the decision problem in the OP should be characterized as one under uncertainty, because discontinuity holds for the outcomes the denizens of the OP face and they cannot know precisely where discontinuity sets in.

I have not offered a defence of maximin as the correct decision rule given my characterization of the decision problem in the OP. That would require a separate argument. Even though I have not defended a specific decision rule, the characterization of the decision problem in the OP I have offered enables us to make the following observation. The decision problem in the OP and the principles of distribution chosen in the OP are sensitive to the outcomes under consideration. For instance,

if the denizens of the OP were facing outcomes for which continuity holds, the decision problem would be one under conditions of risk. In cases where discontinuity holds, the decision of the denizens of the OP will be sensitive to their rough estimate as to where discontinuity sets in. If they believe discontinuity sets in at a point closer to the worst outcomes, the principles they choose will reflect this. Because it is sensitive to the outcomes under consideration, the OP could be used to accommodate different principles of distribution.³¹ Given the variety in our intuitions about which distributions are acceptable, and the failure of a single distributive principle to capture our intuitions, this is a welcome result worth exploring.³²

afaikkurtulmus@sabanciuniv.edu

³¹ For a very helpful discussion of such hybrid views see Paula Casal, 'Why Sufficiency Is Not Enough', *Ethics* 117 (2007), pp. 296–326.

³² I am grateful to David Miller, Adam Swift, Nicholas Vrousalis, Andrew Williams, and two anonymous referees for this journal for their helpful comments on previous versions of this article.