Economics & Philosophy (2025), 1–18
doi:10.1017/S0266267125100473

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Future generations and the evidential veil of ignorance

Johan E. Gustafsson[1] and Andreas L. Mogensen[2]

[1]Department of Philosophy, The University of Texas at Austin, Austin, TX, USA and [2]Faculty of Philosophy, University of Oxford, Oxford, UK
Corresponding author: Andreas L. Mogensen; Email: andreas.mogensen@philosophy.ox.ac.uk

(Received 18 July 2023; revised 24 February 2025; accepted 24 March 2025)

Abstract
Ideal Contractualism views principles of justice as corresponding to what rational, mutually disinterested persons would collectively choose behind a veil of ignorance. It is well-known that Ideal Contractualism faces profound challenges in accounting for justice between generations. We present a unified solution to these problems that involves rejecting the assumption that the parties conceive of their choices as causally efficacious and assumes instead that the parties choose in light of the news value of their decision. And we explore what concrete principles would be chosen by the parties as governing intergenerational justice against the backdrop of this assumption.

Keywords: Social contract theory; evidential decision theory; population ethics; intergenerational justice; the veil of ignorance; dynamic choice

According to *Ideal Contractualism*, the requirements of justice are those principles for governing political institutions that rational, mutually disinterested persons would choose when they are deprived of knowledge about themselves that would allow them to favour their own interests over others (that is, when they are placed behind the *veil of ignorance*).[1] This theory faces some significant challenges when accounting for justice between generations. In this paper, we will explore how these challenges can be met.

Our first task will be to make clear just how significant these challenges are (section 1). In particular, we emphasize the following worry: because principles governing intergenerational justice are identity-affecting, the choices available to the parties seem to provide them with information about where they are in the order of generations. We go on to present a unified solution to this and the other major challenges facing the application of Ideal Contractualism to the intergenerational setting (section 2). We propose to reject the assumption that the parties conceive of

[1]Vickrey (1945, 329) and Rawls (1971, 136–42; 1999, 118–23; 2001, 85–9).

their choices as causally efficacious and to assume instead that they choose in light of the news value of their decision. With this general conception of the choice-situation in hand, we explore what concrete principle would be chosen by the parties as governing intergenerational saving (section 3). We argue that if the parties can be assumed to obey *ex ante* Pareto sequentially, they will choose in accordance with *ex post* Average Utilitarianism.

1. Problems for Ideal Contractualism

In this section, we outline the significant challenges faced by Ideal Contractualism in specifying a theory of intergenerational justice.

As is well-known, there are deep difficulties associated with attempting to imagine the contract situation as representing a compact between generations, with the contracting parties corresponding to all those persons who actually exist, be they past, present, or future.[2] Because of the far-going and profound influence of political institutions, the composition and size of the total population of everyone who will ever live varies depending on what principles govern those institutions. As a result, it seems we cannot coherently conceive of the choice among principles as to be made by an assembly of all those persons who will ever exist. We cannot coherently conceive of an assembly of individuals choosing among options if the choice of some of those options would make it the case that some of them do not exist, never have existed, and never will exist.

A potential solution is to imagine that some individuals behind the veil of ignorance may represent merely possible people who will never exist.[3] Among other problems discussed in the literature, the proposal arguably goes against the core Ideal Contractualist belief that principles of justice correspond to what those who are governed by the principles would together choose for themselves, by giving a say in the design of the principles to people who are not to be governed by them.[4] Moreover, insofar as merely possible people who may never exist are to choose among options that determine whether or not they will ever exist, it seems they must do so in part by comparing whether they are better off existing as opposed to never existing at all. It is far from clear, however, that such comparisons even make sense.[5]

We could instead imagine that the contracting parties know that they all belong to the same generation, although the veil of ignorance deprives them of knowledge about the identity of the particular generation to which they all belong.[6] Yet this makes it hard to see why the parties should agree to principles that require them to save anything at all on behalf of later generations, assuming that each aims to maximize their own standard of living. Given this aim, it would be best for the members of each generation to save nothing and consume everything.[7]

[2]See Rawls (1971, 139; 1999: 120), Barry (1977) and Heyd (1992: 44–5).

[3]Kavka (1975), Parfit (1984: 392), Barry (1989: 194–6), Mulgan (2006: 43), Gardiner (2009: 115) and Finneron-Burns (2017: 819).

[4]Compare Finneron-Burns (2017: 819).

[5]See Williams (1973: 87), Parfit (1984: 487), Broome (1993: 77) and Bykvist (2007).

[6]Following Rawls (1971: 138–40; 1999: 119–21).

[7]See Rawls (1971: 291–2; 1999: 254–5).

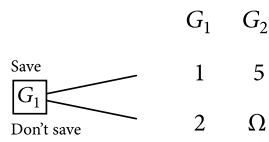
Jane English and John Rawls attempted to solve this problem via the introduction of an additional constraint on the choice of savings principle, proposing that "the correct principle is that which the members of any generation ... would adopt as the one their generation is to follow and as the principle that they would want preceding generations to have followed ... no matter how far back ... in time".[8] The idea is that the parties will agree to principles that require them to save on behalf of later generations, since the parties want any preceding generations to have done the same on their behalf.

This constraint has an ad hoc quality.[9] Nor is it altogether clear what it means for a principle of saving to be one that you would want preceding generations to have followed. If you are glad to be alive, does it follow that you would want preceding generations to have followed whatever principle of saving had to have been followed in order for you to come into existence? If the contracting parties expect to be happy to be alive, does it follow that they should choose such principles? Would it follow that whatever idiosyncratic saving schedule is a necessary condition of their existence is, by that very fact, just? This result seems absurd, but it is not clear how to avoid it.

Last but not least, this proposal fails to resolve what we perceive as an especially serious problem faced by Ideal Contractualism in the intergenerational setting. If the choice of principles at any given time determines the size of the subsequent population and no one can coherently be conceived as making choices that would bring about their own non-existence, then it seems impossible to keep the members of each generation ignorant of the generation to which they belong. Unless arbitrarily restricted, the choices available to the parties will inevitably provide them with information about who they are. Unsurprisingly, this leads to some unwelcome results. The problem is especially acute when we consider the choices available to members of the first generation.

Consider the following case, representing a choice to be faced by the members of the first generation:

Case One



Going down in this case represents the choice not to save and to leave no descendants. Going up represents the choice to bring into existence a second generation and to save on their behalf so that they can enjoy a much higher standard of living than is available to members of the first generation.

[8]Rawls (1993: 274), following English (1977: 98). See also Rawls (2001: 160).

[9]This concern may not have carried much weight with Rawls (1971: 141; 1999: 122), as he seems to admit to rigging the social-contract set-up to get what he wants out of it, writing: "We want to define the original position so that we get the desired solution."

Suppose that we try to put the members of G_1 behind the veil of ignorance and ask them to choose a principle governing this case.[10] They may reason as follows. We cannot coherently conceive of an assembly of individuals choosing among options if the choice of some of those options would make it the case that some of them do not exist, never have existed, and never will exist. But the members of G_2 (and any subsequent generations) depend for their existence on whether the parties now choose a principle that requires G_1 to go up (to save) as opposed to down (to not save). Therefore, the parties know that they must be in G_1. Insofar as each aims to maximize their own standard of living, each will prefer a principle whose implementation results in G_1 going down in Case One. This choice trivially satisfies the requirement that the parties must want preceding generations to have followed the same principle, no matter how far back in time. Thus the correct principle of intergenerational saving will apparently have the first generation save nothing for posterity and leave no descendants.[11]

It may be objected that the parties behind the veil of ignorance should not be imagined as able to make choices about what generations will exist. But note that choices about how to arrange society and about how much to save for future generations will predictably have an effect on the size and composition of the future population. And, when they do, cases structured like Case One are possible.

Some may think that this issue rests simply on a failure to keep in mind the basic elements of the Ideal Contractualism framework. It may seem natural to think that Ideal Contractualism straightforwardly justifies depriving the parties of the ability to choose such a principle, precisely because its availability allows the members of G_1 to work out who they are. Its availability is therefore incompatible with the assumption that principles of justice are to be chosen from behind a veil of ignorance. If we hold fast to that assumption, it may seem that any such principle must be off the table. We find this unpersuasive. The availability of this principle does not allow the members of G_1 to distinguish themselves from other existing people. After all, if it is chosen, then they are the only existing people, and we have just argued that this principle, if available, will be chosen. The veil of ignorance is intended to ensure that the parties do not have knowledge about themselves that would distinguish them from others. It is not intended to keep them from knowing something that is true of everyone who will ever live.[12]

[10]Note that we do not assume that the parties are told that they are either in G_1 or G_2, and thus explicitly informed that they are *not* in any of the subsequent generations that might succeed G_2. Rather, we imagine that they are told that the options depicted in Case 1 are or were open to G_1. They are then asked to choose a principle or set of principles governing this case.

[11]If we, like Rawls (1971: 131; 1999: 113), insist that principles of justice cannot be formulated using proper names or definite descriptions that pick out particular individuals, this might instead involve the parties choosing the principle according to which *every* generation is to save nothing and leave no descendants.

[12]Thus Rawls (1971: 136; 1999: 118) states that the aim of the veil of ignorance is to "nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage".

Others may object that this problem is illusory or irrelevant, since there has never existed a first generation.[13] The claim that a first generation never existed presumably appeals to the vagueness of the boundary between those of our remote biological ancestors to which requirements of justice do not apply and those more recent ancestors to which they do. Nevertheless, we cannot infer that no first generation existed from the fact that it is vague which generation was the first. Among theories of vagueness, both epistemicism and supervaluationism entail that we can truly say that there existed a first generation.[14] Moreover, even if there was no first generation, there could in principle have existed one, and we should want our theory of justice to be able to apply in possible worlds in which a first generation exists.[15]

Finally, some have objected that the problem posed by Case One can be avoided if we assume that the individuals who occupy the Original Position are to be imagined not as actual people in a state of ignorance, but as trustees who choose on their behalf.[16] Yet, interpreting the parties as people's trustees contradicts the contractualist tenet that "the fair terms of social cooperation are conceived as agreed to *by those engaged in it*".[17] Furthermore, it does not seem to help. The parties can infer by a similar argument that they are trustees for the first generation, provided that a trustee can only represent a person if she exists and that who she represents is independent of her choice.[18] Moreover, if a trustee could represent possible people who will never exist, then the trustee would need to be able to compare whether existence is better than non-existence for that possible person. As noted, it is far from clear whether such comparisons between existence and non-existence can be made.

2. Ideal Contractualism, Meet Evidential Decision Theory

The key insight that underwrites our proposed solution is the observation that the argument presented in the foregoing discussion of Case One implicitly assumes what we call

[13]Attas (2009: 205) writes: "The concept of a first generation, involving as it must a theological idea of creation, appears more suited to mythology than to philosophy."

[14]On epistemicism, see Williamson (1994: 185–215). On supervaluationism, see Keefe (2000: 152–201).

[15]It may be objected that this line of reasoning would be rejected by Rawlsians, since Rawls (1971: 159) dismisses the claim that moral conceptions should hold for all possible worlds, objecting that the space of possible worlds is so wide that the project of devising moral principles that cover it would "outrun human comprehension". Yet Rawls clearly did not think that reasoning about principles of intergenerational justice in a way that takes into account the existence of a first generation outruns human comprehension, since he does exactly that. For example, Rawls (1971: 291) appeals to the position of "the least fortunate first generation" in arguing that the Difference Principle should not be applied in the intergenerational setting.

[16]In his later work, Rawls (1993: 24–5) appears to adopt this kind of interpretation at points, writing that "the parties are ... to be seen as representatives of free and equal citizens" and that we are to think of each as "responsible for the essential interests of a free and equal citizen". It may be argued that this interpretation of the choice situation is present from the start, since Rawls (1971: 64) emphasizes that when his principles of justice refer to persons, "the reference is to representative persons holding the various social positions". But stipulating that the principles make reference to the positions of representative persons in society is not the same as conceiving of the parties in the Original Position as representatives of the people to be governed by such principles, rather than the people themselves.

[17]Rawls (1993: 23). The emphasis is ours. Compare Rawls (2001: 15): "the fair terms of social cooperation are to be given by an agreement entered into by those engaged in it".

[18]Compare Finneron-Burns (2017: 818).

> *The Causal View* By virtue of their agreement to a given body of principles behind the veil of ignorance, the parties consider themselves as causally determining what principles govern their political institutions.[19]

This assumption is crucial to the reasoning by which the parties were imagined as inferring that they belong to G_1. Given this assumption, the parties can infer that the existence of people who are not in G_1 depends causally on their choice of principles. Since their own existence does not depend causally on their choice of principles, it follows that the individuals who are to choose must belong to G_1.

While there may be many potential ways to block this argument, we propose to replace the Causal View with

> *The Evidential View* By virtue of their agreement to a given body of principles behind the veil of ignorance, the parties consider themselves as gaining evidence about what principles govern their political institutions and *not* as causally determining what principles govern their political institutions.

If we assume, in addition, that the parties obey *Evidential Decision Theory*, then the evidential connection between their agreement and what principles actually inhere in society will suffice to make the individuals invested in their agreement. Roughly speaking, Evidential Decision Theory instructs you to choose the option that you would most like to learn that you will choose.[20] The main rival theory is known as *Causal Decision Theory*; it instructs you to choose the act that can be expected to cause the best outcome.[21] It can be the case that learning that you will make some choice is good news even if your choice is merely correlated with, but does not cause, some desirable outcome. In such cases, Evidential and Causal Decision Theory are liable to disagree.[22]

Note that, behind the veil of ignorance, given the Evidential View, the choice of principles is not conceived by the parties as having any causal effect on society. The choice of principle only gives the parties evidence about what principle of justice is implemented in society. So, if the individuals follow Causal Decision Theory, they would have no preference for any principle given the Evidential View. If they follow Causal Decision Theory, the Causal View is needed for them to have any interest in the choice of principles.

Given the Evidential View, the parties cannot reason in the way described earlier. They cannot infer that they belong to G_1 from the premise that the existence of G_2 is causally dependent on their choice, since they treat that premise as false.

[19]Some people have objected that the parties in the Original Position cannot be conceived as causing anything, since the Original Position is merely a device of representation. Clearly, this is true in that the Original Position is merely a thought experiment and what happens in thought experiments does not causally affect the real world. Nonetheless, in describing a thought experiment, we must specify what causal powers, if any, the agents are imagined as having or knowing that they have. It is in this sense that the Causal View is to be understood.

[20]Jeffrey (1965: 1–6) and Ahmed (2014: 43–6).

[21]Gibbard and Harper (1978: 128), Lewis (1981: 11–12) and Joyce (1999: 4).

[22]Interested readers may wish to consult the more formal characterization of Evidential and Causal Decision Theory provided in the Appendix.

Furthermore, they cannot reason in an analogous fashion based on the corresponding evidential dependencies. They cannot infer that they must belong to G_1 from the premise that their choices provide them with evidence about whether or not G_2 exists but do not provide them with evidence about whether or not they themselves exist.[23] Since the 'evidence about' operator creates an opaque context, Leibniz's Law cannot be applied in the standard fashion in this setting. The parties cannot reasonably infer with high confidence that they are in G_1 unless they choose a principle whose acceptance is strong evidence that G_1 will go down in Case One. If they instead choose a principle whose acceptance is strong evidence that G_1 goes up in Case One, they should consider it possible that they are in G_2. And they may well prefer to learn that they might be members of G_2 rather than members of G_1, since the members of G_2 have a standard of living much higher than that achievable by the members of G_1.

We now explain how this same approach allows us to solve the other hurdles to the intergenerational application of Ideal Contractualism that were noted in the previous section.

Consider the following problem. Suppose that the contracting parties know that they all belong to the same generation, although the veil of ignorance deprives them of knowledge about the generation to which they all belong. As noted, this can seem to make it difficult to see why the parties should agree to save anything, assuming that each aims to maximize their own standard of living.[24]

This argument, however, does not go through if we assume that the parties obey Evidential Decision Theory. Even if there is nothing that you can do now to affect whether or not some desirable outcome occurs, your choice may be evidence that that desirable outcome occurs or has occurred. That choice may therefore be recommended by Evidential Decision Theory. In particular, within the framework of Ideal Contractualism, the choice of a principle requiring some positive rate of saving serves as evidence that previous generations have saved on our behalf, as we now explain.

Here is why. Since it is intended to define the contours of ideal justice, the choice situation imagined by Ideal Contractualism involves an assumption of full compliance.[25] This assumption extends across time, as well as holding within a generation. In choosing among principles, the parties assume that all persons have complied, currently comply, or will comply with whatever principles they choose. The parties do not believe that their choices causally determine the actions of past people. Rather, they treat their choice of principles as evidence that those principles define a just institutional order and assume that the principles governing their political institutions are, have been, and will be just. Therefore, even if they are purely self-interested, they have reason to choose principles that require some positive rate of saving.[26]

[23]To do so would be to commit a version of the *Masked Man Fallacy*. See Macintosh (1995: 529).

[24]As Rawls (1971: 292) observes (compare Rawls 1999: 254–5): "unless they care at least for their immediate successors, there is no reason for them to agree to undertake any saving whatever … Either earlier generations have saved or they have not; there is nothing the parties can do to affect it."

[25]Rawls (1971: 132; 1999: 114).

[26]Although she does not invoke Evidential Decision Theory, this may well be what English (1977: 98) has in mind when she argues that since the choice of savings principle is made under the assumption of full

The Evidential View, accordingly, explains why the parties should choose principles that require them to save on behalf of future generations – even given that they know they all belong to the same generation. There is an obvious affinity between the Evidential View, so applied, and Rawls's suggestion that the parties are to choose a principle of saving that they would want previous generations to have followed. At the same time, the Evidential View serves to clarify and refine that idea so as to avoid problems with which it is otherwise beset. For example, earlier we noted that the contracting parties could well be said to want previous generations to have adhered to whatever idiosyncratic savings schedule represents a necessary condition of their own existence, whereas it is absurd to suppose that adherence to that savings schedule is, by that very fact, just. The Evidential View avoids this problem. The parties are assumed to obey Evidential Decision Theory and to choose a savings principle according to the news value of their decision. While their existence may depend causally on past people behaving in some particular way, nothing they could learn about the behaviour of past people through their choice of savings principle can raise or lower their confidence that they are currently alive, since their existence is already certain. The causal dependence of their existence on some particular schedule of past savings behaviour is irrelevant to their decision and drops out of consideration, as desired.

The Evidential View also undermines the motivation for rejecting the idea of a contract between generations in the first place. As set out in the previous section, that rejection was motivated by the thought that the composition and size of the total population of everyone who will ever live varies depending on what principles govern a society's political institutions, whereas we cannot coherently conceive of an assembly of persons choosing among a set of principles if adoption of some of those principles would result in some of those people having never existed in the first place. As should be clear at this point, the Evidential View does away with that problem and allows us to make sense of the the contract situation as representing a contract between generations, with the contracting parties corresponding to all those persons who actually exist, be they past, present, or future. Their choices do not causally affect what principles govern their shared institutions, and therefore cannot change the size and composition of the population of everyone who will ever live. Their choices can provide them with evidence about the size and composition of the total population of everyone who will ever live, but that is no obstacle to the assumption that the choosing parties together represent that very population.

3. What Principle Would the Parties Choose?

Given that we adopt the Evidential View, what principles of justice can the parties be expected to choose? In order to answer this question, it is necessary to make more detailed assumptions about the decision theory that describes the contracting parties.[27]

The Ideal Contractualism tradition contains varying assumptions on this point. In the intragenerational setting, Rawls argues that the parties decide in accordance with

compliance, "the choosers in the original position should assume that other generations save according to just principles, too. Then selecting a saving principle would not be contrary to their self-interest."

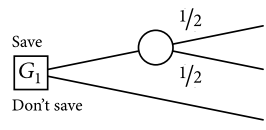
[27]A commitment to Evidential Decision Theory is, as explained in the Appendix, in principle compatible with a range of decision principles, and need not in itself commit us to expected utility theory.

the non-probabilistic *Maximin* rule. Harsanyi argues that the parties will obey *the Laplace Criterion*, which uses a uniform probability distribution. The depth of their disagreement is unclear, however. Harsanyi seems concerned with fundamental questions about rational decision making under uncertainty, whereas for Rawls' Maximin is merely a plausible "rule of thumb" appropriate under specific conditions.[28] These include the assumption that each decision maker "cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule".[29] Under the specific assumptions Rawls adopts, it is very plausible that the recommendations of the Laplace Criterion coincide with those of Maximin, as Rawls acknowledges.[30] Rawls rejects utilitarianism considered as the public charter on the basis of which the basic structure is to be organized.[31] Harsanyi's concerns are more fundamental, even meta-ethical, encompassing what it means to make a moral judgement or express a moral preference.

Notably, Rawls does not recommend Maximin as a decision rule for choosing the intergenerational rate of saving.[32] He seems thereby to concede that the particular conditions that justify Maximin as a rule of thumb fail to hold in this context. But he suggests nothing to replace it. We shall assume that the parties' decisions coincide with those that would be recommended in light of subjective expected utility theory – in its evidential guise, of course.[33]

Given this approach, the principle to be chosen by the parties will depend on how they asses the probability of belonging to generations that might possibly exist but whose existence cannot be guaranteed by the predictor's implementation of the chosen principles. The existence of these generations may depend, in addition, on one or more chance events. For instance, consider

Case Two



	G_1	G_2
	1	13
	1	Ω
	5	Ω

For simplicity, we assume that the generations in our examples are all the same size. In this case, how should the individuals assess the option of saving?

[28]Harsanyi (1975: 595–8, 605) and Rawls (1971: 155). Compare Rawls (1999: 134–5).

[29]Rawls (1971: 154; 1999: 134).

[30]Rawls (1971: 182; 1999: 158–9). Compare Harsanyi (1975: 606) and Bognar (2011: 335–9).

[31]See Rawls (1971: 182; 1999: 158).

[32]Specifically, Rawls (1971: 291) rejects the use of Maximin as a decision rule for choosing the intergenerational rate of saving because he believes it would require no saving at all, given that the members of the first generation will be worst off and cannot be made better off as a result of a schedule of savings.

[33]Consistent with the assumption that the parties decide on the basis of precise subjective probabilities, we could instead assume that the parties maximize *risk-weighted expected utility* with the most risk-avoidant attitude within reason, as recommended by Buchak (2017: 630–3). However, risk-weighted expected utility encounters well-known difficulties in dynamic contexts, which will specifically interest us here. For a discussion, see Buchak (2013: 170–200). For an argument against maximizing risk-weighted expected utility behind the veil of ignorance, see Nebel (2020).

A first idea is to retain probabilistic impartiality between the generations and to regard it equally likely that one is in G_1 as that one is in G_2. Since the individuals know that they exist, we shouldn't focus on expected well-being in variable-population cases behind the veil of ignorance but instead on conditional expected well-being. Let the *conditional expected well-being* of a person S from a prospect X be equal to S's expectation from X given that S exists.[34] Then the average conditional expected well-being from saving is $\big((1 + 1)/2 + 13\big)/2 = 7$, which is better than the average conditional expected well-being from not saving – that is, 5.

A second idea is to regard the potential positions one might occupy in society as equally likely: G_1 given that chance goes up, G_2 given that chance goes up, and G_1 given that chance goes down. Taking these mutually exclusive and jointly exhaustive positions as equally likely, we get that the expected well-being of saving for the individuals is $(1 + 13 + 1)/3 = 5$, which is the same as the expected well-being of not saving.[35]

A third idea is that it's equally likely that chance goes up or down and, if chance goes up, it's equally likely that one is in G_1 or in G_2. According to this line of thinking, the individuals' expectations from saving is $\big(\big((1 + 13)/2\big) + 1\big)/2 = 4$, which is worse than their expectation from not saving, which is 5.[36]

Hence we get that saving in Case Two might be better than, worse than, or equally good as not saving depending on how we should assess uncertain prospects where some generations only have some chance of existing. Let us now assess the plausibility of these three alternative ideas.

Conditional Ex Ante Average Utilitarianism

The first idea, of retaining probabilistic impartiality between the generations by relying on conditional expected well-being, leads to

Conditional Ex Ante Average Utilitarianism A prospect x is choice-worthy if and only if there is no feasible prospect y such that $V_{CEAAU}(y) > V_{CEAAU}(x)$,

[34]Harsanyi quoted in Ng (1983: 168).

[35]The prospect of saving in Case Two is structurally similar to the Sleeping Beauty Problem; see Piccione and Rubinstein (1997: 12–13) and Elga (2000: 144). In the Sleeping Beauty Problem, there will be one waking day if the coin lands heads and two waking days if the coin lands tails but the agent cannot, on any waking day, tell which day it is. In the prospect of saving, there will be one generation if a coin (or some equivalent chance event) lands heads and two generations in the coin lands tails – and the parties cannot tell what generation they belong to. The main difference between the prospect of saving and the Sleeping Beauty Problem is that, in the latter, it is the same person that would enjoy the two observer days in case the coin lands tails, whereas, in the former, there is no overlap between the generations. Taking all potential positions as equally likely in the prospect of saving corresponds to taking the chance of tails to be 1/3 in the Sleeping Beauty Problem. This is worrying, because Evidential Decision Theorists are vulnerable to Dutch-book arguments if they do not accept the 1/2 answer; see Briggs (2010: 17–18) and Pettigrew (2020: 69–70). Since the two generations are composed of different agents, Case Two is structured more like Bostrom's (2002: 64) Incubator case, rather than the Sleeping Beauty Problem. Hence the particular Dutch book suggested by Briggs does not apply directly in Case Two, because it relies on the same agent being offered bets on each waking day. It may be objected, perhaps, that, if the 1/3 answer were correct in the standard Sleeping Beauty Problem, it would also be so if the agents waking up on Monday were not the same as the agents waking up on Tuesday. Note, moreover, that the 1/3 answer in Case Two is open to Bostrom's (2002: 124) Presumptuous Philosopher objection.

[36]This answer corresponds (see note 35) to taking the chance of tails to be 1/2 in the Sleeping Beauty Problem; see Lewis (2001: 171).

where

$V_{CEAAU}(x) =_{\text{df}}$ the average conditional expected well-being for possible people in x.

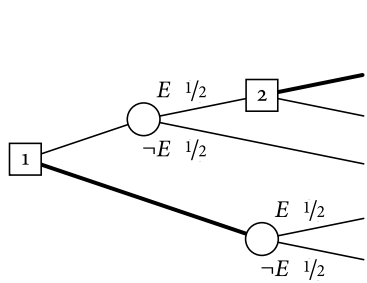
Notably, this principle leads to an extreme form of longtermism:[37] any future lives with positive probability (however small) count the same as any current lives and any certain life. There are many more possible people that we may bring about with some probability in the far future than in the near future. Given the enormous amount of possible people that have some chance of existing in the far future, this principle will give an enormous proportional weight to the long-term. Not only will the future matter more than the near future, it will matter in a strange way. Since the probability of a life is not taken into account, it will not be important to merely reduce the probability of bad lives or increase the probability of good lives. What matters is whether you can affect whether these lives would still be possible.

Even though Conditional *Ex Ante* Average Utilitarianism has some counter-intuitive implications, it might still makes sense from the perspective of the parties behind the veil of ignorance (given the Evidential View). Would they opt for this principle? It seems that they would not. Conditional *Ex Ante* Average Utilitarianism gives rise to significant problems in sequential cases. It violates the following principle:[38]

The Weak Sequential Statewise Pareto Principle If S and S' are two plans such that the same people exist in the outcomes of these plans and, in each state of nature, everyone has a higher well-being in the outcome of S than in the outcome of S', then S' is not followed if S available.

Consider

Case Three



G_1	G_2
7	1
3	Ω
11	Ω
2	Ω
10	Ω

At node 2, the conditional *ex ante* average is $(7 + 1)/2 = 4$ if we go up and 3 if we go down. Accordingly, Conditional *Ex Ante* Average Utilitarianism would go up at node 2. Taking that into account at node 1, the conditional *ex ante* average is $\big((7 + 11)/2 + 1\big)/2 = 5$ if we go up and $(2 + 10)/2 = 6$ if we go down. So Conditional *Ex Ante* Average Utilitarianism (using backward induction) would go

[37]On longtermism, see MacAskill (2022: 4–5) and Greaves and MacAskill (Forthcoming).

[38]See Gustafsson (2018: 599) and Kowalczyk (2023: 9) for similar principles.

down at node 1. This violates the Weak Sequential Statewise Pareto Principle. To see this, consider the following table of what happens if we follow the plan recommended by Conditional *Ex Ante* Average Utilitarianism and the plan consisting in doing the opposite of its recommendations:

	E happens		$\neg E$ happens	
	G_1	G_2	G_1	G_2
Up at node 1 and down at node 2	3	Ω	11	Ω
Down at node 1	2	Ω	10	Ω

Regardless of whether E happens, it holds that the plan consisting in going *up at node 1 and down at node 2* gives G_1 a higher well-being than the plan *down at node 1* and that G_2 will not exist.

Veiled Average Utilitarianism

The second idea, taking the potential positions one might be in society as equally likely, leads to the following principle:[39]

Veiled Average Utilitarianism A prospect x is choice-worthy if and only if there is no feasible prospect y such that $V_{VAU}(y) > V_{VAU}(x)$,

where

$V_{VAU}(x) =_{\text{df}}$ the expected total well-being in x divided by the expected population size in x.

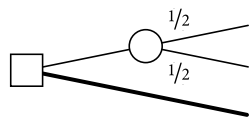
This principle avoids the extreme weight that Conditional *Ex Ante* Average Utilitarianism put on very unlikely future people. It will still support the conclusion that the far future matters much more than the near-term, given at least a moderate probability that there a lot more people in the future than there will be in the short term.

Would the parties behind the Evidential Veil opt for Veiled Average Utilitarianism? It seems that they would not. Veiled Average Utilitarianism violates, even in non-sequential cases,

The Conditional Ex Ante Pareto Principle If the same people possibly exists in prospects x and y and each one of them has a higher conditional expected well-being in x than in y, then y is not chosen if x could be chosen instead.

Consider, for instance,

Case Four



G_1	G_2
2	14
2	Ω
1	13

[39]Thomas (2016: 150).

If we go up, the expected total divided by expected size is $\big((2+14+2)/2\big)/\big((2+1)/2\big)=6$, and, if we go down, it is $(1+13)/2=7$. Hence Veiled Average Utilitarianism would go down. But this violates the Conditional *Ex Ante* Pareto Principle, since going up gives everyone a greater conditional expectation than going down.

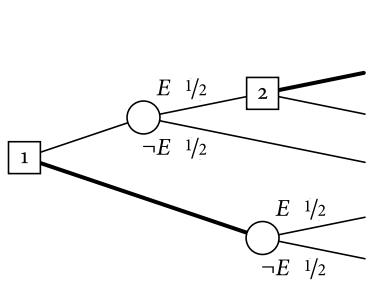
Conditional expectations in Case Four

	G_1	G_2
Up	$(2+2)/2=2$	14
Down	1	13

So each generation is doing worse, conditional on their existence, if Veiled Average Utilitarianism is followed.[40]

Veiled Average Utilitarianism also violates the Weak Sequential Statewise Pareto Principle. To see this, consider

Case Five



G_1	G_2
2	4
2	Ω
18	Ω
1	Ω
17	Ω

At node 2, the expected total divided by expected size is $(2+4)/2=3$ if we go up and 2 if we go down. So Veiled Average Utilitarianism would go up at node 2. Taking this into account at node 1, the expected total divided by expected size is $\big((2+4+18)/2\big)/\big((2+1)/2\big)=8$ if we go up and $(1+17)/2=9$ if we go down. Therefore, Veiled Average Utilitarianism (using backward induction) goes down at node 1. This violates the Weak Sequential Statewise Pareto Principle. To see this, consider the following table of what happens if we follow the plan recommended by Veiled Average Utilitarianism and the plan consisting in doing the opposite of its recommendations:

	E happens		¬*E* happens	
	G_1	G_2	G_1	G_2
Up at node 1 and down at node 2	2	Ω	18	Ω
Down at node 1	1	Ω	17	Ω

[40]This kind of result is hinted at in Nebel (2019: 339 n41).

Regardless of whether E happens, it holds that the plan consisting in going *up at node 1 and down at node 2* gives G_1 a higher well-being than the plan *down at node 1* and that G_2 will not exist.

Ex Post Average Utilitarianism

Finally, the third idea – to accept the probabilities of the chance events and, for each final outcome, to regard it as equally likely that one belongs to each of the generations in that outcome – leads to the principle

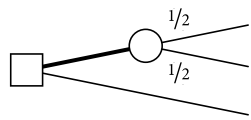
Ex Post Average Utilitarianism A prospect x is choice-worthy if and only if there is no feasible prospect y such that $V_{EPAU}(y) > V_{EPAU}(x)$,

where

$V_{EPAU}(x) =_{\text{df}}$ the expected average well-being in x.

Is this the principle we are looking for? Like Veiled Average Utilitarianism, it violates The Conditional *Ex Ante* Pareto Principle. To see this, consider

Case Six



G_1	G_2
1	1
9	Ω
6	2

If we go up in Case Six, the expected average well-being is $(1 + 9)/2 = 5$ and, if we go down it is $(6 + 2)/2 = 4$. Accordingly, *Ex Post* Average Utilitarianism would go up.

Conditional expectations in Case Six

	G_1	G_2
Up	$(1 + 9)/2 = 5$	1
Down	6	2

We have that *Ex Post* Average Utilitarianism violates Conditional *Ex Ante* Pareto, since going down gives everyone a higher conditional expectation than going up. But, unlike the reasoning that led to Conditional *Ex Ante* Average Utilitarianism, the reasoning that led to the *Ex Post* Average Utilitarianism does not attach any importance to conditional expectations. There seems to be little reason to care about this kind of dominance in conditional expected well-being if you do not think that you are equally likely to be anyone in a prospect.[41] In Case Six, you should think that it is more likely that you belong to G_1 if we go up than if we go down at the choice node. So the parties need not care about violations of the Conditional *Ex Ante* Pareto Principle.

[41]This point is analogous to Nozick's (1963: 230–1; 1969: 118–19) objection to statewise dominance in cases where there is probabilistic dependence between acts and states.

And, unlike the earlier principles, *Ex Post* Average Utilitarianism does not violate Sequential Statewise Pareto. Hence, of the three approaches, *Ex Post* Average Utilitarianism emerges as the only plausible choice.

4. Conclusion

Ideal Contractualism faces a number of significant challenges when applied to questions of intergenerational justice. These problems, however, can be solved by rejecting the Causal View and endorsing the Evidential View in its stead.

In addition, we have shown that it is possible to make significant progress in determining what particular principles the parties will endorse by reasoning about dynamic decision problems and assuming that the parties obey *Ex Ante* Pareto sequentially. This leads us to expect that the parties will choose in accordance with *Ex Post* Average Utilitarianism.

Average Utilitarianism is often dismissed due to a number of supposedly fatal objections.[42] These objections are based on intuitive ethical judgements. To count against Average Utilitarianism given Ideal Contractualism, these judgements must move the parties behind the evidential veil who are only concerned with securing their individual ends. But, behind the evidential veil, the parties will have good reasons, as we have seen, to disregard any considerations that conflicts with *Ex Post* Average Utilitarianism – and, more generally, the parties are unmoved by any ethical considerations as they just aim to secure their own ends. Yet there remains the worry that, even if these objections do not work if they are considered from within the framework of Ideal Contractualism, they may still be compelling objections to the whole of Ideal Contractualism.

References

Ahmed A. 2014. *Evidence, Decision and Causality*. Cambridge: Cambridge University Press.

Arrhenius G. 2000. Future Generations: A Challenge for Moral Theory. PhD thesis, Uppsala University.

Attas D. 2009. A Transgenerational Difference Principle. In *Intergenerational Justice*, ed. A. Gosseries and L.H. Meyer, 189–218. Oxford: Oxford University Press.

Barry B. 1977. Justice between Generations. In *Law, Morality, and Society: Essays in Honour of H. L. A. Hart*, ed. P.M.S. Hacker and J. Raz, 268–284. Oxford: Clarendon Press.

Barry B. 1989. *A Treatise on Social Justice Volume 1: Theories of Justice*. Berkeley: University of California Press.

Bognar G. 2011. Can the Maximin principle serve as a basis for climate change policy? *The Monist* **94**, 329–348.

Bostrom N. 2002. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. London: Routledge.

Briggs R. 2010. Putting a value on beauty. In *Oxford Studies in Epistemology, Volume* **3**, ed. T.S. Gendler and John Hawthorne, 3–34. Oxford: Oxford University Press.

Broome J. 1993. Goodness is reducible to betterness: the evil of death is the value of life. In *The Good and the Economical: Ethical Choices in Economics and Management*, ed. P. Koslowski and Y. Shionoya, 70–84. Berlin: Springer.

Buchak L. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Buchak L. 2017. Taking risks behind the veil of ignorance. *Ethics* **127**, 610–644.

Bykvist K. 2007. The benefits of coming into existence. *Philosophical Studies* **135**, 335–362.

Elga A. 2000. Self-locating belief and the Sleeping Beauty problem. *Analysis* **60**, 143–147.

[42]For example, the Utility Monster (Nozick 1974: 41; Arrhenius 2000: 53–4), the Egyptology Objection (McMahan 1981: 115; Parfit 1984: 420) and the Sadistic Conclusion (Arrhenius 2000: 54).

English J. 1977. Justice between Generations. *Philosophical Studies* **31**, 91–104.
Finneron-Burns E. 2017. The intergenerational original position. *Social Theory and Practice* **43**, 805–823.
Gärdenfors P. 1979. Forecasts, decisions and uncertain probabilities. *Erkenntnis* **82**: 159–181.
Gardiner S.M. 2009. A contract on future generations? In *Intergenerational Justice*, ed. Axel Gosseries and L.H. Meyer, 77–118. New York: Oxford University Press.
Gibbard A. and W.L. Harper 1978. Counterfactuals and two kinds of expected utility. In *Foundations and Applications of Decision Theory*, vol. **I**, ed. C.A. Hooker, J.J. Leach and E.F. McClennen, 125–162. Dordrecht: Reidel.
Greaves H. and W. MacAskill Forthcoming. The case for strong longtermism. In *Essays on Longtermism: Present Action for the Distant Future*, ed. H. Greaves, J. Barrett and D. Thorstad. Oxford: Oxford University Press.
Gustafsson J.E. 2018. The difference principle would not be chosen behind the veil of ignorance. *Journal of Philosophy* **115**, 588–604.
Harsanyi J.C. 1975. Can the Maximin principle serve as a basis for morality? A critique of John Rawls's theory. *American Political Science Review* **69**, 594–606.
Heyd D. 1992. *Genethics: Moral Issues in the Creation of People*. Berkeley: University of California Press.
Huntley N., R. Hable and M.C.M. Troffaes 2014. Decision making. In *Introduction to Imprecise Probabilities*, ed. T. Augustin, F.P.A. Coolen, G. de Cooman and M.C.M. Troffaes, 190–206. New York: Wiley.
Jeffrey R.C. 1965. *The Logic of Decision*. New York: McGraw-Hill.
Joyce J.M. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Joyce J.M. 2005. How probabilities reflect evidence. *Philosophical Perspectives* **19**, 153–178.
Joyce J.M. 2010. A defense of imprecise credences in inference and decision making. *Philosophical Perspectives* **24**, 281–323.
Kavka G.S. 1975. Rawls on average and total utility. *Philosophical Studies* **27**, 237–253.
Keefe R. 2000. *Theories of Vagueness*. Cambridge: Cambridge University Press.
Kowalczyk K. 2023. People in suitcases. *Journal of Moral Philosophy* **20**, 3–30.
Levi I. 1974. On indeterminate probabilities. *The Journal of Philosophy* **71**, 391–418.
Lewis D. 1981. Causal decision theory. *Australasian Journal of Philosophy* **59**, 5–30.
Lewis D. 2001. Sleeping Beauty: reply to Elga. *Analysis* **61**, 171–176.
MacAskill W. 2022. *What We Owe the Future*. New York: Basic Books.
Macintosh J. 1995. Masked man fallacy. In *The Oxford Companion to Philosophy*, ed. T. Honderich, 529. New York: Oxford University Press.
McMahan J. 1981. Problems of population theory. *Ethics* **92**, 96–127.
Mulgan T. 2006. *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations*. Oxford: Clarendon Press.
Nebel J.M. 2019. An intrapersonal addition paradox. *Ethics* **129**, 309–343.
Nebel J.M. 2020. Rank-weighted utilitarianism and the veil of ignorance. *Ethics* **131**, 87–106.
Ng Y.-K. 1983. Some broader issues of social choice. In *Social Choice and Welfare*, ed. P.K. Pattanaik and M. Salles, 151–173. Amsterdam: North-Holland.
Nozick R. 1963. The Normative Theory of Individual Choice. PhD thesis, Princeton University.
Nozick R. 1969. Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of his Sixty-Fifth Birthday*, ed. N. Rescher, 114–146. Dordrecht: Reidel.
Nozick R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
Parfit D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
Pettigrew R. 2020. *Dutch Book Arguments*. Cambridge: Cambridge University Press.
Piccione M. and A. Rubinstein 1997. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* **20**, 3–24.
Rawls J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rawls J. 1993. *Political Liberalism*. New York: Columbia University Press.
Rawls J. 1999. *A Theory of Justice*, revised edn. Cambridge, MA: Harvard University Press.
Rawls J. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.
Schoenfield M. 2012. Chilling out on epistemic rationality: a defense of imprecise credences (and other imprecise doxastic attitudes). *Philosophical Studies* **158**, 197–219.
Thomas J.T. 2016. *Topics in Populations Ethics*. PhD thesis, University of Oxford.

van Fraassen B.C. 1990. Figures in a probability landscape. In *Truth or Consequences: Essays in Honor of Nuel Belnap*, ed. J.M. Dunn and A. Gupta, 345–356. Dordrecht: Kluwer.
Vickrey W. 1945. Measuring marginal utility by reactions to risk. *Econometrica* **13**, 319–333.
Williams B. 1973. *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge: Cambridge University Press.
Williamson T. 1994. *Vagueness*. London: Routledge.

Appendix: Evidential and Causal Decision Theory

We can formalize Evidential and Causal Decision Theory in terms of expected utility theory in accordance with the framework proposed by David Lewis.[43] Let S be the set of possible worlds. Let a K-partition of S be a partition such that each $k \in K$ fully specifies the causal dependence of each possible outcome on each available act. Then, if $\{k_1, k_2, \ldots, k_m\}$ is a K-partition of S, $P(\cdot)$ is the agent's rational credence function, and $u(\cdot)$ is her utility function, then an available act a is rationally permissible according to Evidential Decision Theory if and only if there is no available act a' such that

$$\sum_{i=1}^{m} P(k_i|a)u(k_i \wedge a) < \sum_{i=1}^{m} P(k_i|a')u(k_i \wedge a')$$

By contrast, an available act a is rationally permissible according to Causal Decision Theory if and only if there is no available act a' such that

$$\sum_{i=1}^{m} P(k_i)u(k_i \wedge a) < \sum_{i=1}^{m} P(k_i)u(k_i \wedge a')$$

In other words, Causal Decision Theory asks us to compute expected utilities using the unconditional prior probabilities of each $k \in K$, whereas Evidential Decision Theory asks us to compute expected utilities using the corresponding conditional probabilities.

We emphasize that the distinction between Evidential and Causal Decision Theory has more general significance and need not assume that agents maximize expected utility relative to a unique subjective probability function, which may be unacceptable to Rawlsians. In particular, consider agents whose beliefs are modelled not by a single probability function, but by a set of probability functions (a so-called *representor*), corresponding to the different chance hypotheses left open by her evidence.[44] Many different decision principles for agents whose beliefs are so-represented have been proposed.[45] But virtually all of these principles entail that the agent's preferences over acts depend on the expected utility of the available acts with respect to the elements of the representor. In computing the expected utilities of the available acts relative to any element of the representor, a decision may then be faced as to whether to compute the evidentiary expected utility or the causal expected utility.

Consider, in particular, an agent who acts in accordance with the *Maximin-Expected-Utility Principle*, which instructs her to choose the option with the greatest minimum expected utility relative to the probability functions in her representor.[46] This principle should be of especial interest to Rawlsians, since it reduces to Maximin in the case where the agent's beliefs, as represented by sets of probabilities, are maximally spread out for each state. Note, then, that in order to compute the minimum expected utility of a given action, we can use either the evidential expected utility, such that the Maximin-Expected-Utility Principle is given by the rule that an act $a \in A$ is rationally permissible if and only if there is no a' such that

$$\min_{P(\cdot)\in R}\left(\sum_{i=1}^{m} P(k_i|a)u(k_i \wedge a)\right) < \min_{P(\cdot)\in R}\left(\sum_{i=1}^{m} P(k_i|a')u(k_i \wedge a')\right)$$

[43]Lewis (1981: 12).
[44]Levi (1974: 241–2), van Fraassen (1990: 349), Joyce (2005: 156; 2010) and Schoenfield (2012).
[45]Huntley *et al.* (2014).
[46]Gärdenfors (1979: 169).

or the causal expected utility, where it is given by the rule that act $a \in A$ is rationally permissible if and only if there is no a' such that

$$\min_{P(\cdot)\in R}\left(\sum_{i=1}^{m} P(k_i)u(k_i \wedge a)\right) < \min_{P(\cdot)\in R}\left(\sum_{i=1}^{m} P(k_i)u(k_i \wedge a')\right).$$

Johan E. Gustafsson is a research scientist at the University of Texas at Austin. He is the author of the book *Money-Pump Arguments* and works on decision theory, ethics, political philosophy and metaphysics. Email: johan.eric.gustafsson@gmail.com. URL: https://johanegustafsson.net.

Andreas L. Mogensen is a senior research fellow in the Faculty of Philosophy at the University of Oxford. He has published on issues such as the demands of beneficence, decision-making under deep uncertainty, population ethics, and discounting in the context of climate policy, with publications appearing in journals including *Noûs*, *Philosophy and Phenomenological Research*, *The Journal of Philosophy* and *Philosophical Studies*. URL: https://andreasmogensen.com/.

Cite this article: Gustafsson JE and Mogensen AL. Future generations and the evidential veil of ignorance. *Economics and Philosophy*. https://doi.org/10.1017/S0266267125100473