IMPARTIAL EVALUATION UNDER AMBIGUITY*

Abstract: How should an impartial social observer judge distributions of wellbeing across different individuals when there is uncertainty regarding the state of the world? I explore this question by imposing very weak conditions of rationality and benevolent sympathy on impartial betterness judgements under uncertainty. Although weak enough to be consistent with all the main theories of rationality, these conditions prove to be sufficient to rule out any heterogeneity in what is good for individuals, to require a neutral attitude to uncertainty on the part of the social observer and to require that both individual and social betterness be strongly separable.

1. Introduction: Harsanyi's Theorem

In his two famous papers of 1953 and 1955 defending Utilitarianism, Harsanyi draws on the same simple idea: that to determine what is morally best we should put ourselves into the shoes of an impartial, but benevolent, rational evaluator of states of affairs that differ in terms of the wellbeing of the various individuals within them.[i] Such an evaluator would, he argued, have preferences between states that reflected their impartial concern for the wellbeing of individuals and so offer guidance as to the preferences that we should have if we want to avoid giving special consideration to our own partial interests. Application of this thought experiment to situations of risk, in which the prospects to be evaluated are probability distributions over the different possible states of individual wellbeing, led Harsanyi to a startling and controversial conclusion:

rational moral evaluation in these circumstances must, if appropriately sensitive to the wellbeing of affected individuals, take an expectational Utilitarian form. That is, one such distribution should be judged better than another just in case the expected sum of individual wellbeing is greater under the former than the latter.

We need not agree with Harsanyi that adoption of a perspective of impartiality is constitutive of moral judgement in order to recognize the value of this type of thought experiment; in particular to the study of the evaluative judgements that guide policy interventions made under uncertainty about the implications for the wellbeing of individuals affected by them. But Harsanyi's results apply only to contexts in which the probabilities of wellbeing outcomes are known. So, in this paper, I will extend his thought experiment to contexts of uncertainty in which they are not known and, in particular, to those that are 'ambiguous' in the sense that the evaluator lacks the information and/or expertise to form precise subjective probabilities for all relevant contingencies. Many important public policy decisions must be made in contexts of this kind, but they have thus far received little attention in social ethics.[ii]

I will proceed as follows. In the section 2, I will set out the basic concepts and assumptions that will provide the framework for an investigation of impartial evaluation under uncertainty and then, in section 3, draw out some of if its core implications. In section 4, I will address the question of what form such evaluative judgements should take in conditions of ambiguity. Finally, in section 5, I will address the permissibility of attitudes to risk and uncertainty that are ruled out by expected utility theory and, in particular, those implying a violation of the Sure-thing Principle, the separability condition that is a central plank of the Bayesian theory of rational practical judgement. The remainder of the introduction will be devoted to putting this project into some context. Throughout the paper, the emphasis will be on explaining the significance of formal results and demonstrating why they follow from the adopted assumptions, often by using examples rather than mathematically more general proofs.[iii]

In the large literature inspired by Harsanyi's two papers, his thought experiment has been extended and modified in various ways. Versions of his 1955 axiomatic argument, the one that I will concentrate on in this paper, have been proven in a number of different frameworks for expected utility theory: in the von Neumann and Morgenstern one by, for instance, John Weymark, in the Savage one by Peter Hammond and Peter Fishburn, and in the Bolker-Jeffrey one by John Broome.[iv] Indeed, similar results exist for even more general frameworks, in which one or more of the rationality conditions on individuals and/or the social observer are weakened.[v] What is remarkable about this literature is the robustness of Harsanyi's conclusions: that the two requirements on the impartial evaluator of rationality and of benevolent responsiveness to what is good, in expectation, for individuals imply that her judgements take an expectational Utilitarian form or some generalization of it.

In these more general frameworks, the standard of rationality applied to moral evaluation is that of Bayesian decision theory (otherwise known as subjective expected utility theory). Benevolent responsive, on the other hand, is captured by what are known as the *ex ante* Pareto conditions on the relation between is best individuals and what is best overall (from the impartial perspective). These require in essence that no prospect can be better than another unless it is better for some individual. Together with the rationality condition, they imply that prospects must be ranked in accordance with the weighted sum of the expected wellbeing of individuals relative to a shared probability measure on states. (The assumption of impartiality, when it is meaningful in this framework, then forces the weights on individuals’ expected wellbeing to be equal.)

These results, like Harsanyi’s original ones, have been subject to various kinds of criticism. Two are particularly important here. The first is that the *ex ante* Pareto conditions impose insensitivity to inequality on the part of the social evaluator. (I will explain this criticism in more detail in the next section.) The second is that the rationality assumptions of expected utility theory are overly restrictive in at least two different ways. Firstly, and most importantly for my project, the requirement that decision makers assign precise probabilities to states of the world seems overly demanding, or even unreasonable, in situations in which they lack sufficient evidence to be able to do so in a non-arbitrary way. In such situations, known in decision theory as situations of *ambiguity,* individuals cannot determine unique expected utilities for prospects with any confidence and so need not follow the prescriptions of the Bayesian theory (I will discuss this in section 4). Secondly, expected utility theory disallows attitudes to risk and uncertainty that many find perfectly reasonable; for instance, those exhibited in a famous paradox due to Maurice Allais.[vi] In particular, it rules out certain forms of caution in the face of uncertainty that seem to some to be appropriate when the wellbeing of others is at stake (I will discuss this in section 5).

I will respond to the first type of objection by working with a much weaker condition of benevolent (Paretian) responsiveness on the part of the social evaluator to the goodness of individuals' prospects; one that is also sensitive to equality considerations. This will suffice to make room for non-Utilitarian forms of impartial judgement. In response to the second type of criticism, I will weaken the rationality assumptions of expected utility theory sufficiently to accommodate a wide range of rival theories of rational judgement under uncertainty. In principle, this response opens up a space for impartial social evaluation that is not expectational in form and which can exhibit forms of sensitivity to uncertainty disallowed by Harsanyi's theory.

The rationality constraints that I will adopt will be strictly weaker, not only than Harsanyi's, but than those imposed by any of the mainstream theories of rationality, whatever kind of uncertainty they are tailored to. These include not only the main theories of decision making under ambiguity, such as those of Gilboa and Schmeidler and of Klibanoff, Marinacci and Mukerji but also the main rival theories of rationality under risk and/or subjective uncertainty to expected utility theory, such as Quiggin's rank-dependent utility theory, Tversky and Kahnemann's cumulative prospect theory and Buchak's risk-weighted utility theory. [vii] This will give our conclusions very broad scope indeed.

This fact makes the conclusions themselves all the more surprising, even if they are prefigured to some extent in the more formal literature.[viii] It does not, on the face of it, seem unreasonable to allow that how an individual is affected by uncertainty may depend on characteristics peculiar to them. Nor that judgement as to what is best, either for a particular individual or overall, should be sensitive to the severity of any uncertainty about how well individuals will fare. But it turns out that the assumptions that I will adopt, weak though they may appear to be, are sufficient to rule out any difference in the way that individuals evaluate their prospects, any sensitivity on the part of individual and social evaluation to the presence of ambiguity, and any violation of a separability condition (called the Sure-thing principle) central to expected utility theory and which many have argued to be too restrictive. This means that acceptance of these weak constraints not only restricts the degree to which an impartial evaluator can exhibit sensitivity to the form or severity uncertainty that is faced, but also significantly restricts judgement as to what is best for any individual. This creates a dilemma for theories of social evaluation: either these restrictive implications must be accepted or we must abandon either the weak rationality requirements on impartial judgement or those of benevolent sensitivity to individual wellbeing.

2. Basic Framework

Let me start by setting up the problem more precisely. Our concern is impartial evaluation of what I will call *social outcomes* and *social prospects*: respectively distributions of wellbeing over a set of individuals and distributions of social outcomes over the set of possible states of the world. States should be understood to be combinations of features or factors that determine the wellbeing outcome of any action or policy, whatever these may be. Social prospects will be denoted by italicized Roman capital letters: *X*, *Y*, etc. For any individual *i* or state *s*, $X(i)$, $X(s)$, and $X(i,s)$ are, respectively, the distribution of wellbeing across states (the individual prospect) that *i* faces, the social outcome in *s* (i.e., the distribution of wellbeing across individuals in that state) and the outcome for *i* in *s* that *X* implies. For most purposes it will suffice to work with just two individuals and two to four states, allowing me to represent a social prospect by a table whose rows are individual prospects and columns are social outcomes and whose cells contain the wellbeing magnitude of the individual concerned in that state of the world. For instance, in Table 1, we see that in state 1, Jocelyn's wellbeing is of magnitude 1 and Kit's of magnitude 0, while in state 2, Jocelyn's wellbeing is of magnitude 5 and Kit's of magnitude 4.

[Insert Table 1 here]

Three remarks about the assumptions underlying the project. Firstly, I will make no assumption about the kind of uncertainty faced by evaluators. So, even if my main interest is in ambiguity, the results will apply also to situations in which probabilities are available either objectively or subjectively. Secondly, I will say very little in this paper about what individual wellbeing is other than that I take it to be an interpersonally comparable numerical measure of how good outcomes are for individuals. So, any conclusions of the investigation will hold for any theory of wellbeing that allows that it be numerically measure and compared. How to measure quantities of individual wellbeing, and to compare measurements for different individuals, is a non-trivial problem of course, but I will simply assume here that methods for doing so are available.[ix]

Finally, there will be no presumption in this paper that the evaluator of social prospects is perfectly well-informed about all moral and/or empirical facts relevant to determining which *distribution* of wellbeing across states is best for each individual. (By working with wellbeing, rather than the properties of outcomes that produces it, we remove any uncertainty about how good the outcome of a prospect is for a given individual in a given state.) Nor will I say anything about what these facts might be, though intuitively they will include how likely it is that any state will arise. Instead, the task will be to examine the constraints on such (potentially imperfect) evaluation that arise out of the commitment to impartiality and to other normative principles: in particular, to judging rationally under uncertainty and to a certain kind of sensitivity to the wellbeing of the individuals making up society.

Much of our focus will be on how judgements about the goodness of prospects should reflect uncertainty about the state of the world and especially what attitudes an evaluator can permissibly take to the presence or absence of uncertainty about the wellbeing outcomes of prospects. An evaluator will be said to be *neutral* with respect to risk/uncertainty about wellbeing if and only if they are indifferent between any two distributions of wellbeing that have the same expected wellbeings; for instance between the prospect of someone achieving a wellbeing of magnitude 5 for sure and that of them achieving a wellbeing of magnitude 10 with probability 0.5 and a wellbeing of zero with probability 0.5. On the other hand, they are said to be *averse* to risk/uncertainty about wellbeing iff they rank distributions with less spread over those with more; for instance, the former over the latter of the two distributions just mentioned.[x] Expected utility theory does not require any particular attitude to uncertainty about wellbeing but it does significantly restrict patterns of such attitudes, a feature that is frequently used to motivate rival, more permissive, theories.

The judgements of the impartial (but not omniscient) evaluator will be represented by a *betterness* ordering $\succsim$ of both social outcomes and social prospects, with the expressions $X \succsim Y$ and $X(s) \succsim Y(s)$ respectively saying that, in the judgement of the evaluator, social prospect *X* is at least as good as (i.e. weakly better than) social prospect *Y* and that the outcome of prospect *X* in state *s* is as at least as good as that of prospect *Y*.[xi] Correspondingly $X \succ Y$ means that social prospect *X* is strictly better than social prospect *Y* and $X \backsim Y$ that the two prospects are equally good. So too for expressions involving social outcomes.

Let’s take as given a set of possible well-being values, a set of individuals and a set of states of the world. For convenience, we assume that the first is simply an open interval of real numbers and that the corresponding sets of individual prospects, social outcomes and social prospects respectively contain every possible distribution of wellbeing to individuals, across states of the world, and across both. To capture the idea that it is only wellbeing, and how it is distributed, that matters in the evaluation of both individual and social prospects and of social outcomes, we define a betterness relation as follows.

> A **betterness** relation on a set of a wellbeing distributions is a complete, transitive and continuous relation on the set that is *monotonic* in wellbeing, i.e. a ranking of them such that no distribution is ranked higher than another unless it yields greater wellbeing in at least one state for at least one individual.

There are perfectly reasonable individual and social orderings that do not fit this definition of a betterness relation (the Leximin ordering of social outcomes, for instance, is not continuous[xii]). But, although building completeness, transitivity and continuity into the definition of betterness right from the outset does narrow the scope of my conclusions to some extent, it also serves to simplify subsequent discussion to a considerable degree. In particular, in combination with the characterisation of the domains on which these relations are defined, it ensures that for any distribution $X = (x^1, \ldots, x^m)$ in the domain of a betterness relation, there exists a real number e and corresponding distribution $E = (e, \ldots, e)$ in its domain, called the *equal-valued equivalent* of *X*, which is as good as X (on that betterness relation), a fact that will prove useful later on.

I will make three key assumptions about social betterness judgements as represented by betterness orderings of wellbeing distributions across both states and individuals. The first, unsurprisingly, is that they are impartial. Impartiality will be construed here as indifference on the part of the social evaluator between a wellbeing distribution across individuals and any permutation of it (a reshuffling of the wellbeing assignments to individuals).[xiii] For example, an impartial evaluator will regard an assignment of wellbeing 1 to Jocelyn and wellbeing 0 to Kit as equally good as an assignment of wellbeing 1 to Kit and wellbeing 0 to Jocelyn.

To make this more precise, let σ be a permutation on the set of individuals: a mapping from each individual to another that represents the reshuffling of them. Now, for any prospect *X* and state *s*, let $\sigma(X(s))$ be the social outcome defined by, for all individuals, i, $\sigma\big(X(s)\big)(i) = X(\sigma(i), s)$, so that $\sigma(X(s))$ is the social outcome obtained by a particular reshuffling of the wellbeing outcomes assigned by X to individuals. Then the social betterness relation is required to be impartial in the sense that it satisfies:

(Outcome Anonymity) $X(s) \backsim \sigma(X(s))$

Note that this characterization of impartiality implicitly presupposes that the value of wellbeing doesn't depend on the individual enjoying it, something that is not implied by the monotonicity of individual betterness but which is consonant with the comparability of individual wellbeing.

The second assumption is that the betterness ranking of social prospects is minimally consistent with the ranking of social outcomes. More precisely, I will assume that social betterness satisfies *State Dominance*: that, if in every state the outcome of one prospect is at least as good as the outcome of another, then the first is at least as good overall as the second. Formally:

(State Dominance) If for all states s, $X(s) \gtrsim Y(s), \text{then } X \gtrsim Y$

State Dominance is a property of all mainstream normative theories of rational judgement and, in that sense, is relatively uncontroversial. Most satisfy a somewhat stronger condition that requires in addition that if in every state the outcome of one prospect is at least as good as the outcome of another and in at least one state is strictly better, then the first prospect is strictly better overall than the second. But there are notable theories of rationality under ambiguity (such as the aforementioned theory of Gilboa and Schmeidler) that don't satisfy it and, in any case, the results in the paper do not require it. Even on our weaker formulation, State Dominance has an important implication, namely that the betterness relation is *weakly* separable across states. That is, if two prospects differ only in their outcomes in state *s*, then one is better than the other just in case its outcome is better, given that *s*. As is frequently observed, the plausibility of this assumption depends on our ability to individuate states sufficiently finely as to settle everything that matters, including any relevant characteristics (such as fairness) of the procedure used to achieve the outcome. More on this later.

The final assumption concerns the relationship between the (overall) goodness of a prospect and how good it is for the individuals affected by it. Informally, the assumption requires that social betterness be positively sensitive to what is good for individuals; in this sense it may be regarded as a requirement of sympathetic benevolence on the part of the social evaluator. There are however a variety of ways of cashing this out formally. Let me first state the condition as it is usually formulated before discussing both its interpretation and how I propose to weaken it. For every individual *i*, let $\succsim_i$ be a weak betterness ordering of the set of individual prospects, with $X(i) \succsim_i Y(i)$ meaning that *i*'s prospect in *X* is at least as good as her prospect in *Y*. Then for all social prospects *X* and *Y*:

(Strong Pareto) If, for all individuals *i*, $X(i) \succsim_i Y(i)$, then $X \succsim Y$. Furthermore, if there exists some individual i^* such that $X(i^*) \succ_{i^*} Y(i^*)$, then $X \succ Y$.

There are three different interpretations that might be given of the notion of individual betterness at work here; that might be termed the subjective, objective and judgemental conceptions of individual betterness. On the *subjective* conception an individual's betterness ranking of prospects represents her subjective preferences between them, whether construed as the (informed, reflective) choices she would make between them if offered both, or as her subjective evaluation of them in terms of their desirability. On the *objective* conception, it represents the ordering of prospects in terms how good in fact they are for her. Finally, on the *judgemental* conception, it represents the impartial evaluator's judgements as to how good the prospects are for the individual.

The constraints on betterness yielded by these three interpretations have rather different rationales and implications. The subjective interpretation is the most common in the literature, perhaps because of the popularity of the view in economics and political science that one prospect is better for an individual than another just in case the individual prefers it, at least when meeting certain epistemic conditions such as being well-informed and of clear mind. The Strong Pareto condition it implies has its natural home in the theory of social aggregation where it functions as the 'democratic' principle of sensitivity to unanimity in the opinions of individuals (a weak version of consumer and/or voter sovereignty). But in the context of the betterness judgements of an impartial evaluator it is quite implausible without some heavy-duty epistemic conditions on individual preference; minimally including the requirement that they be as well-informed as the evaluator. In any case, as we shall see, this interpretation will turn out to be incompatible with the other assumptions we have made.

On the objective interpretation, Strong Pareto says that overall betterness supervenes positively on individual betterness, a condition that John Broome dubs the Principle of Personal Good in order to distinguish it from the preference-based interpretation of it.[xiv] Although the objective interpretation provides a compelling rationale for respecting unanimity in individual betterness judgements, it is not ideal for the purposes of this paper. For the overall betterness relation is intended to represent the judgements of a social evaluator, someone I assumed to be impartial but not necessarily perfectly informed about whatever facts determine how good a distribution of wellbeing is for an individual. The social evaluator so construed cannot ensure that she respects what is *in fact* best for individuals; what she must do is respect what *in her judgement* is best for them. That is why I adopt the third interpretation of individual betterness, whereby Strong Pareto says that the impartial evaluator's judgement as to the goodness of a distribution of prospects to individuals must supervene positively on her judgement as to how good each individual's prospect is for them.

Irrespective of the adopted interpretation, the principle is considered by many to be too strong. This is because Strong Pareto implies that no prospect can be better than another unless it is ranked higher by at least one individual. But as a consequence impartial social evaluation must be insensitive to inequalities in the final wellbeing of individuals. To see this, consider the two prospects displayed in Table 2.

[Insert Table 2 here]

Suppose that Jocelyn and Kit are indifferent between the prospects they face under I and II, perhaps because they both regard the two states as equally probable. Then by Strong Pareto, I and II are equally good overall. But in one respect II is better (say some): under II, whatever happens, Jocelyn and Kit will have equal wellbeing, whereas under I, whatever happens, one of them will be better off than the other. Defendants of the principle reply that if there is something better about equality it must be because it is better for the individuals concerned – in which case this should be taken care of by the wellbeing magnitudes. Critics retort that it should be possible to determine someone’s wellbeing without reference to how it compares to someone else’s, on pain of a circularity in the concept of wellbeing.

I will not attempt to resolve this dispute but instead, in acknowledgement of the case against Strong Pareto, work with a more restricted principle of sensitivity to individual betterness, one that is not subject to objections rooted in a concern for equality. The general idea is that the requirement that unanimity in individual betterness judgements be respected, should be restricted to cases concerning comparisons between prospects which do not differ in their equality characteristics. It is up for debate what the relevant equality characteristics are exactly, but here we assume only that it suffices that two prospects are such that, in *every* state of the world, either the social outcome of the first prospect amounts to a reshuffling of the individual wellbeing outcomes in the second, or that the social outcomes of both are perfectly equal (all individuals get the same wellbeing). This gives us the principle:

(Pareto for Equivalent Outcomes) Suppose that *either*:

(i) *Permutation*: For every state, s, there exists a permutation σ_s, such that $X(s) = \sigma_s\big(Y(s)\big)$, *or,*

(ii) *Equality*: For all individuals, i and j, and for every state, s, $X(i,s) = X(j,s)$ and $Y(i,s) = Y(j,s)$.

If, for all individuals *i*, $X(i) \succsim Y(i)$, then $X \succsim Y$. Furthermore, if there exists some individual i^* such that $X(i^*) \succ_{i^*} Y(i^*)$, then $X \succ Y$.

Pareto for Equivalent Outcomes would simply not apply in the previous example, for instance, because the distribution of wellbeing to Jocelyn and Kit in state 1 of prospect II cannot be obtained by reshuffling the distribution of wellbeing to them in state 1 of prospect I. Nor is the wellbeing of Jocelyn equal to that of Kit in state 1 of prospect I. So there is no entailment by this condition that prospects I and II are equally good overall.

On the face of it, Pareto for Equivalent Outcomes is a rather weak condition and, indeed, it is consistent with many theories of social welfare; including a number of *ex post* versions of Egalitarianism that violate Strong Pareto.[xv] It is notable however that although *ex post* Prioritarianism satisfies the first (Permutation) clause of the condition, standard versions violate the second (Equality) clause.[xvi] To see why this must be so, note that the Equality clause trivially applies when there is just one individual and so in this framework the social betterness relation will be the same as that of the (single) individual. But *ex post* Prioritarianism does not respect unanimity in individual betterness even in such trivial cases. For example, prospect (0,3) may be better for Jocelyn than prospect (1,1) when the two wellbeing outcomes are equally likely, but regarded as worse overall by an *ex post* Prioritarian theory that applies a transformation on wellbeing that is sufficiently concave (such as the square root function).[xvii] One may retort that such cases fall outside the scope of social ethics, but the point applies in the more clearly 'social' case in which both Jocelyn and Kit face exactly the same two prospects and evaluate them in the same way. I will return to this issue in section 6, but for now will take this is sufficient grounds for setting aside the *ex post* Prioritarian objection to the Equality clause.

This proposed weakening of Strong Pareto does not directly respond to a second prominent objection to it: that unanimity in individual goodness may be spurious in virtue of being an 'accidental' outcome of substantial differences that cancel one another out.[xviii] Suppose Jocelyn and Kit are evaluating two policies. Policy A is that Jocelyn does the washing up if it rains and Kit if it does not, while policy B is the opposite: that Kit does the washing up if it rains and Jocelyn if it does not. Now if Jocelyn thinks it more likely than not to rain and Kit just the opposite then both might evaluate B as better than A. But such agreement in their evaluations is morally irrelevant since based on conflicting beliefs and preferences. Without any doubt, this objection is telling against Pareto conditions when individual betterness is interpreted subjectively. But it is not so under either the objective or the judgemental interpretation for in these cases there should be no differences in the probabilities underpinning judgements of individual betterness. (This point will be developed in more detail in the next section.) So I take the objection to be grounds for rejecting the subjective interpretation in this context, but not for rejecting Pareto for Equivalent Outcomes.

We now have in place all the pieces of the framework that hereafter I will term "rational Paretian social evaluation" of prospects. In summary, it expresses the idea that an impartial wellbeing ranking of social prospects in terms of their comparative goodness should satisfy Outcome Anonymity, State Dominance and Pareto for Equivalent Outcomes. Since these conditions are relatively weak, it is a framework for social evaluation that many should find congenial. Indeed, I know of no mainstream view in social ethics other than *ex post* Prioritarianism that explicitly rejects any of them: the disputes between them concern primarily what other conditions should be satisfied by social betterness. Nonetheless, as we shall see, the framework very strongly constrains both individual and social evaluation of prospects: it disallows heterogeneity in individual good (section 3), it disallows sensitivity to ambiguity in either the individual and social betterness (section 4) and it implies separability of both individual and social betterness across events (section 5).

3. Homogeneity of Betterness

The assumption of an interpersonally comparable measure of wellbeing entails that a unit of wellbeing is as good for one individual as another. It does not follow however that any distribution of wellbeing across states is equally good for all individuals. Individuals themselves are likely to value prospects rather differently because of having different beliefs and different levels of aversion to the risk or uncertainty contained in them. Even if we set aside subjective differences, it doesn't seem unreasonable for the goodness of a prospect for an individual to depends on characteristics peculiar to them: in particular, their sensitivity to uncertainty. If financial security matters more to you than to me, then it could be better for me, but worse for you, to have a lottery ticket paying a $10,000 with probability 0.1 and nothing otherwise, than to have a guaranteed $800. Nonetheless, such heterogeneity in the sensitivity of individual betterness to how wellbeing is distributed across states is inconsistent with impartial rational Paretian evaluation, which entails that there can be no differences between individuals in how good a prospect is for them.

I will establish this claim in two steps. The first is to show that Outcome Anonymity, our formal expression of the requirement of impartiality with respect to wellbeing, is equivalent, in the presence of State Dominance, to a different impartiality condition, which I will dub Prospect Anonymity. The second will be to show that Prospect Anonymity and Pareto for Equivalent Outcomes jointly imply that all individual betterness relations are the same. Informally, Prospect Anonymity says that the social evaluator must be indifferent between any social prospect and one that is obtained by reshuffling the individual prospects occurring in it. More formally, let $\sigma(X)$ be the social prospect defined by $\sigma(X)(i,s) = \sigma\big(X(s)\big)(i)$. Then:

(Prospect Anonymity) $X \backsim \sigma(X)$

Now consider the distributions over three states and two people displayed in Table 3 below. One might suppose that prospect I could be better for both Jocelyn and Kit, not because they regard the final wellbeing outcomes denoted by a, b, etc., differently, but because they have different attitudes or sensitivities to the distribution of wellbeing across the states. The first step to establishing that this is not possible (in this framework) is to see that Outcome Anonymity and State Dominance imply Prospect Anonymity. For suppose that betterness is outcome anonymous. Then in every state the social outcomes in distributions I and II are equally good. So by State Dominance, social prospects I and II are equally good overall. Hence social betterness is prospect anonymous.

[Insert Table 3 here]

For the second step, let the certainty equivalent $C_i(X) = (c_i, \dots, c_i)$ of prospect X, for individual *i*, be the equal-valued equivalent of X under i's betterness relation. The magnitude c_i should be understood in this context as the amount of wellbeing that individual *i* must receive for sure in order to be as well-off as she is when facing uncertain prospect *X*. Now let (x, y, z) be any distribution over three states and let (j, j, j) be Jocelyn's certainty equivalent for (x, y, z). Then consider the social prospects I and II displayed in Table 4.

[Insert Table 4 here]

By Prospect Anonymity, prospects I and II are equally good overall and by construction they are equally good for Jocelyn. Then by Pareto for Equivalent Outcomes, it must be the case that I and II are equally good for Kit. For if they were not, this condition would imply that I was either strictly better or strictly worse than II, depending on whether it was strictly better or strictly worse for

Kit. It follows that (j, j, j) is Kit's certainty equivalent for (x, y, z) as well. But since this is true for any choice of distribution, this shows that Jocelyn and Kit have the same certainty equivalents for every prospect and, hence, have the same betterness ranking over prospects.

It is well-known that full Bayesian rationality together with Strong Pareto not only suffices for an expectational Utilitarian representation of social betterness, but also implies that individuals have the same subjective probabilities.[xix] Since Bayesianism also requires risk neutrality in utility and since, in these frameworks, utility is a cardinal measure of wellbeing, individuals assigning the same probabilities to states must rank prospects in the same way. What the result proven above shows is that such homogeneity of individual betterness holds even when we assume rationality conditions so weak that they do not imply that individuals have precise probabilities, let alone that individual goodness goes by expected wellbeing or that overall betterness is Utilitarian in form.

These results impose strong constraints on the interpretation of individual betterness rankings. In the first place they rule out any kind of subjective interpretation of them, including construing them as individual preferences. In conditions of uncertainty it is to be expected that, and reasonable for, individuals to assign different probabilities to the states of the world because of the different information that they hold. As a result, they would exhibit different preferences between actions whose wellbeing consequences are dependent on the state of the world. But since such differences are inconsistent with rational Paretian social evaluation, this interpretation must be rejected. This conclusion does not change if under conditions of ambiguity individuals are unable to assign precise probabilities to states, for it remains reasonable that they assign different imprecise probabilities (for instance) and, as a result, exhibit different preferences.

If betterness is construed objectively, on the other hand, it is to be expected that all individual’s betterness relations should depend on the same probability assignment to states. The same holds true for the interpretation of individual betterness as the judgements of the impartial evaluator regarding what is best for individuals, since the evaluator would apply the same probabilities (or other measure of their uncertainty) in all their judgements. What doesn’t follow directly from either of these interpretations however is that there can be no differences in individual betterness with regard to the (dis)value of uncertainty. That this follows from the rather weak assumptions of our framework is therefore quite surprising.

4. Social Evaluation under Ambiguity

We often face *ambiguity*: circumstances in which it is difficult to assign precise probabilities to all the events that interest us, because we lack sufficient evidence to determine them in a non-arbitrary way, for instance, or because experts and/or scientific theories disagree about the implications of the evidence we do have. The question that I want address in this section is: how should social prospects to be evaluated in circumstances of this kind?

Following Daniel Ellsberg’s seminal work, some argue that not only do individuals in fact evaluate prospects under ambiguity differently from under risk, but that such differences are prudentially rational (in the sense of being rationally permitted, though not required).[xx] In particular, the aversion to ambiguity that individuals display in their choices expresses a reasonable attitude of caution in the face of informational scarcity and/or scientific disagreement.[xxi] Recently Rowe and Voorhoeve have argued that social evaluation too can permissibly be sensitive both to the ambiguity contained in the distribution of wellbeing to individuals and to the ambiguity that individuals face in their individual prospects.[xxii]

This viewpoint is far from uncontroversial however. Bayesians point out that such aversion to ambiguity leads to several kinds of apparently irrational behaviour, including refusal of free information and forms of dynamic inconsistency.[xxiii] Others argue that, difficult as it may be, evaluators of social prospects must simply do their best to assign probabilities to relevant states so that prospects can be evaluated under ambiguity in the same way as under risk (when the probabilities are known).[xxiv] The issues this dispute raises are complex and I will not try to adjudicate it. Instead, I simply take its existence as providing grounds for interest in the question as to the implications of allowing for ambiguity aversion in social evaluation.

In the first of two 'paradoxes' that Ellsberg presents, he invites us to suppose that a ball will be drawn randomly from each of two different urns, both containing 10 balls that are either white or black in colour. The 'risky' urn contains exactly five black balls; the 'ambiguous' urn an unknown number of them. Let Bb be the state in which a black ball is drawn from both, Bw the state in which a black ball is drawn from the risky urn and a white ball from the ambiguous one, and so on. You are offered the option of betting on either black or white from either the risky urn or the ambiguous one with the payoffs in the dollar amounts shown in Table 5.

[Insert Table 5 here]

Ellsberg postulated that many people would be indifferent between Risky Black and Risky White and between Ambiguous Black and Ambiguous White but strictly prefer Risky Black to Ambiguous Black and Risky White to Ambiguous White. Countless experiments have confirmed his hypothesis.[xxv] This pattern of preference has come to be known in the literature as *ambiguity aversion*: a preference for actions which have known or empirically well-supported probabilities of success over alternatives which have unknown or poorly-supported ones. Such a pattern of preferences is ruled out by expected utility theory however. To see this, note that the postulated indifferences are rational according to expected utility theory only if the agent regards a draw of a black or white ball, from either urn, to be equally probable. But if they do, it follows that (they believe that) a black ball is as likely to be drawn from an ambiguous urn as a risky one, and so an expected utility maximiser should be indifferent as to which urn she bets on.[xxvi] That many people are not indifferent between the two is evidence, Ellsberg suggested, not of irrationality but of an unwillingness to assign any probability at all to a draw of a black/white ball from the ambiguous urn. Agents preferring to bet on the risky urn are displaying an aversion to acts whose expected utilities cannot be calculated because of ignorance of the relevant probabilities.

Outside of the laboratory we rarely confront cases as stark as the one Ellsberg imagined, but situations in which for a variety of reasons we lack precise probabilities for relevant states, or in which we lack confidence in our probabilistic estimates, are common. Consider a doctor who must decide which of two treatments to prescribe to their patient. One treatment has been used tested and applied extensively, on a variety of different classes of patient and under a variety of conditions. Based on the rich evidence available, the probability of treatment success is estimated by the doctor to be 85%. The other treatment is a novel one, based on recent theoretical discoveries, but although it was successful in trials in 88% of cases it has been applied far less extensively. Despite the higher observed success rates of the second treatment, it would not be unreasonable for the doctor to opt for the first on the grounds that trial evidence is insufficient to determine with confidence a precise probability of success for her patient. All sorts of features of the conditions on which the trials were conducted and characteristics of the trial population may make the reported success rates a poor predictor of the effect of the treatment on her own patient.

Examples like this one give some initial plausibility to the claim that ambiguity aversion is a permissible attitude to take when one's decisions have ambiguous consequences for others. Two thoughts give further support to this claim. Firstly, in circumstances in which the decision maker lacks sufficient evidence to determine a unique probability assignment in a non-arbitrary way, she cannot be required to do so (ought implies can). And secondly, whether or not she is able somehow to determine a unique set of probability values for relevant contingencies, it is not rationally required that she base her choice of action solely on these probabilities when her evidence does not rule out other ones.[xxvii]

Surprisingly, however, Paretian social evaluation turns out to preclude any sensitivity to ambiguity in either individual or social betterness. To see this, recall the betting scenario imagined by Ellsberg in his first paradox and suppose that our two individuals both prefer Risky Black and Risky White to Ambiguous Black and Ambiguous White in line with his hypothesis. Now consider the social prospects displayed in Table 6.

[Insert Table 6 here]

In prospect II, both Jocelyn and Kit face a bet on the ambiguous urn, in prospect I they both face a bet on the risky one. So they both prefer the prospect they face in I than in II. On the other hand, the social evaluator must be indifferent between the two, for the outcome of prospect II in each state of the world is a permutation of the outcome of prospect I. So Outcome Anonymity requires she should regard the outcomes of the prospects as equally good in every state of the world. Hence by State Dominance, she should be indifferent between the two prospects. But note that I and II do not differ in terms of the equality of the distributions of wellbeing to Jocelyn and Kit because, as per clause (i) of our condition, the social outcomes of prospect II are permutations of those in prospect I. So, by Pareto for Equivalent Outcomes, I is better than II. Contradiction.

Now this argument would not apply if Jocelyn and Kit had different betterness rankings of I and II; for example, if they had different sensibilities to ambiguity. But in section 3 we showed that this is not possible. It follows that our two individuals cannot have ambiguity averse betterness relations in Ellsberg’s set-up. And the argument generalizes without complication to cases involving more than two individuals, simply by considering social prospects that are just like I and II except for the fact that all individuals other than Jocelyn and Kit face the same individual prospects under the two alternatives.

What about the social evaluator? Suppose that she judges a bet on a risky urn to be better than a bet on an ambiguous urn, where the two urns have exactly the same outcomes. For instance, suppose she judges that Risky Black is better than Ambiguous Black, where these are given in Table 7 and with $X = (x_j, x_k)$ being any distribution of wellbeing to Jocelyn and Kit.

[Insert Table 7 here]

Let $E(X) = (e, e)$ be the equal-valued equivalent of social outcome X under the social betterness relation. Then it follows by State Dominance that Risky Black and Ambiguous Black are respectively equally good as the prospects Risky Black* and Ambiguous Black* displayed in Table 8.

[Insert Table 8 here]

Earlier we established that Risky Black* and Ambiguous Black* are equally good for both Jocelyn and Kit. But Risky Black* and Ambiguous Black* do not differ in their equality characteristics because, as per clause (ii) of Pareto for Equivalent Outcomes, the social outcomes of both prospects are perfectly equal in all states. So it follows from this condition that they must be equally good overall. Hence, by transitivity, Risky Black and Ambiguous Black are equally good overall. This shows that social betterness cannot be sensitive to ambiguity. More precisely, it would seem that the distinction between risk and ambiguity is of no significance to impartial social evaluation.

5. Separability

It was observed earlier on that State Dominance implies that the betterness relation is *weakly* separable, i.e., if two prospects differ only in their outcomes in state s, then one is better than the other just in case its outcome is better, given that s. Now Bayesian decision theory imposes a stronger condition than this, known as the Sure-thing Principle, which implies that betterness is *strongly* separable, i.e., that if two prospects have the same outcome in some event E (an event being just a set of states), then one prospect is better than the other just in case its outcome is better, given that E is not the case. To state the principle more formally, let $X_E Y$ be the prospect such that $X_E Y(s) = X(s)$ for all states $s \in E$ and $X_E Y(s) = Y(s)$ for all states $s \notin E$. Then for all prospects X, $\hat{X}$, Y and Z and any state E:

(Sure-thing Principle) $X_E Y \gtrsim \hat{X}_E Y \Leftrightarrow X_E Z \gtrsim \hat{X}_E Z$

The Sure-thing Principle is the subject of considerable normative controversy and some notable decision theories violate it. These include not just the decision theories for ambiguity mentioned before, but also rival models of individual decision making under risk to expected utility theory, such as cumulative prospect theory, rank dependent utility theory and risk-weighted expected utility theory.[xxviii] The debate has had only limited impact on social ethics thus far but recently Lara Buchak has argued that the forms of risk aversion that these models permit, and which are excluded by expected utility theory, serve to motivate, via a Harsanyi-style argument about choice behind the veil of ignorance, social betterness judgement that gives priority to the less well-off.[xxix] (Both Simon Blessenohl and Jake Nebel criticise her argument drawing on results that to some extent prefigure the ones in this section, but which assume Strong Pareto).[xxx]

The controversy around the status of the Sure-thing Principle centres on the rational permissibility of a pattern of preferences that is frequently exhibited in a decision problem known as the Allais Paradox.[xxxi] Consider a set of states {R,B,Y}, with probabilities 0.1, 0.01 and 0.89 respectively, and compare the four distributions displayed in Table 9.

[Insert Table 9 here]

Now the Sure-thing principle implies that prospect I is better than prospect II iff prospect III is better than prospect IV. In fact, however, in experiments where the numbers denote millions of dollars, many people prefer I to II and IV to III, giving reasons such as that choosing II over I means foregoing the certainty of a million dollars in order to gain a very slight chance of 4 million while choosing III over IV does not mean foregoing a large amount with certainty. The more general thought is that how good an outcome is in some event can depend on what the outcome would have been in other events. But this kind of counterfactual dependence of the goodness of the wellbeing outcomes of a prospect in some event on what the wellbeing outcomes of that prospect are in other events is ruled out by the Sure-thing Principle. And so, critics argue, the principle disallows patterns of betterness judgements that are in fact perfectly reasonable.

In fact violations of strong separability, in either individual or social betterness judgements, are simply ruled out by our assumptions regarding impartial social evaluation. For if overall betterness is impartial and satisfies both State Dominance and Pareto for Equivalent Outcomes then both overall betterness judgement *and* individual betterness judgement must respect Sure-thing Principle. We can demonstrate why this is so by means of the example used to describe the Allais paradox. Suppose that Jocelyn and Kit both have the following separability-violating betterness rankings with respect to distributions of wellbeing:

1. (1,1,1) > (4,0,1)
2. (4,0,0) > (1,1,0)

Now consider the corresponding social prospects displayed in Table 10 below. Since the individual prospects of Jocelyn and Kit are both strictly better for them in social prospect I than in social prospect II, and since I and II do not differ in terms of equality (the outcomes of II are permutations of those in I), Pareto for Equivalent Outcomes implies that I is strictly better overall than II. But in all three states the social outcome of prospect II is a permutation of the social outcome in prospect I. So by Outcome Anonymity and State Dominance, I and II are equally good. Contradiction. So, contrary to hypothesis, Jocelyn and Kit cannot both have these Allais-style betterness rankings. But we know from before (section 3) that they must have the same betterness rankings, so there cannot be any case in which just one of them has betterness judgements violating the Sure-thing Principle.

[Insert Table 10 here]

The argument is easily generalized both to any violation of the Sure-thing Principle and to any number of individuals (greater than 1). Recall again that all individual betterness relations must be the same. Suppose that for all individuals *i*, $X_E Y \succ_i \hat{X}_E Y$ but that $\hat{X}_E Z \succsim_i X_E Z$, in violation of the Sure-thing Principle. Consider two social prospects I and II such that for two individuals, say individuals 1 and 2, $I(1) = X_E Y$, $I(2) = \hat{X}_E Z$, $II(1) = \hat{X}_E Y$, and $I(2) = X_E Z$ and for all other individuals *i*, $I(i) = II(i)$. By construction in every state of the world the social outcome of prospect II is reshuffling of the social outcome of prospect I. So clause (i) of Pareto for Equivalent Outcomes is satisfied. Hence it follows from this condition that I is strictly better than II. But since I can be obtained from II by permuting the outcomes in each state, Outcome Anonymity and State Dominance imply that they are equally good. Contradiction.

Now suppose that social betterness violates the Sure-thing Principle, i.e. for some event E and social prospects, X, $\hat{X}$, *Y* and *Z*, it is the case that $X_E Y \succ \hat{X}_E Y$ but that $\hat{X}_E Z \succsim X_E Z$. Let $I = X_E Y$,

$II = \hat{X}_E Y$, $III = X_E Z$, and $IV = \hat{X}_E Z$ and I*, II*, III* and IV* be the corresponding prospects obtained by replacing each social outcome in I, II, III and IV by their equal-valued equivalents. Then by definition of the equality equivalent, I* is strictly better than II* but III* is at least as good as IV*. Note that all individual prospects in I* are the same, all individual prospects in II* are the same, etc., and recall that the individual betterness relations are the same. Since the social outcomes in both I* and II* are perfectly equal in all states, clause (ii) of Pareto for Equivalent Outcomes is satisfied. Hence, since I* is strictly better overall than II*, it requires that for at least one individual I* is strictly better than II* and hence for all individuals it must be so. But we have seen that individual betterness satisfies the Sure-thing Principle and so it must be case that that for all individuals, IV* is strictly better than III*. So it follows by Pareto for Equivalent Outcomes that IV* is strictly better overall than III*. Contradiction. We can conclude that social, as well as individual, betterness must respect the Sure-thing Principle.

6. Concluding Discussion

We have seen that if social betterness is impartial, rational and minimally benevolent (in the sense captured by Pareto for Equivalent Outcomes) then it follows that there can be no heterogeneity in individual goodness, that neither individual nor social betterness is permitted to be sensitive to ambiguity and that both must satisfy the Sure-thing Principle. What should we make of these implications?

As a requirement on moral or social judgement, impartiality as characterized by Outcome Anonymity is no doubt contestable (one might for instance think that partiality towards those with which one has certain kinds of familial or social relationships is justified). But it is surely not contestable that it is *permissible* for social evaluators to seek to be impartial in this sense. But if they do, then they face a trilemma regarding which of the following three desiderata on overall betterness they should reject:

1. That it allow for forms of risk and/or ambiguity aversion that are ruled out by expected utility theory, but which are regarded as permissible in relevant circumstances by rival theories of rationality.
2. That it respect the State Dominance condition.
3. That it respect unanimous individual betterness judgements in cases without implications for equality.

It is clear enough what the Bayesian response would be to this trilemma. For the results of the paper can be read as a vindication of their position: that the principles of expected utility provide the standard of rational evaluation for both individual and social good under all conditions of uncertainty. So, on the Bayesian view, (1) should be rejected because rival theories of rationality are wrong about what forms of sensitivity to uncertainty are permissible. Bayesians who nonetheless recognize the intuitive force of the Ellsberg and Allais paradoxes can reconcile their theory to them by arguing that no violation of their theory is exhibited in the situations they describe provided that proper care is exercised in identifying all wellbeing-relevant features of outcomes.[xxxii] In particular, both the Allais and Ellsberg preferences can be accommodated within a somewhat broader Bayesian theory that treats objective chances as just such features.[xxxiii] The results of this paper may then be read as showing that the right way to allow moral sensitivity to the harm that uncertainty imposes on individuals (however severe it may be) is through an enrichment of the Bayesian theory rather than through its rejection.

Those who advocate rival theories of rationality to the Bayesian one are in the much more uncomfortable position of being able to defend their view only at the cost of having to deny either (2) or (3). Rejection of State Dominance might seem attractive since it faces a well-known objection (due originally to Peter Diamond): that, together with impartiality, it implies that the that the two prospects displayed in Table 11 are equally good (because in both states the social outcome of one prospect is a permutation of the other's), even though intuitively prospect II is fairer than prospect I.[xxxiv]

[Insert Table 11 here]

Likewise, they jointly imply Prospect Anonymity, the condition that says that overall betterness is insensitive to how prospects are distributed to individuals. Since intuitively such insensitivity is consistent with impartiality only if prospects are equally good for all individuals, rejection of State Dominance for social or overall betterness is the only plausible way of blocking the implication that individual betterness relations are the same. And doing so allows for forms of *ex ante* egalitarianism that are consistent with at least some forms of uncertainty aversion.[xxxv]

Rejection of State Dominance is not an option for the main rival theories of rationality to the Bayesian one, however, since they all depend on it. So these theories must accept that individual betterness relations are all the same and find a way of reconciling State Dominance with the intuition that fairness matters (as evidenced by prospect II seeming to be better than prospect I). The most promising way of doing this is to take a leaf out of the Bayesian playbook and argue that wellbeing values must capture *all* the benefits and harms to an individual in a state, including any that derive from modal facts about what would have happened had some other state been the true one and which support fairness claims. But doing so raises the question: if wellbeing values capture all the benefits and harms to individuals, must they not therefore incorporate all those deriving from the uncertainty they face? If the answer is 'yes', then the usual case for permitting individual betterness to display forms of uncertainty aversion ruled out by expected utility theory collapses. Advocates of non-Bayesian theories of rationality thus face a secondary dilemma. They can defend the relevance of their theory for social betterness but only at the cost of ceding that it is not fully adequate as a theory of *individual* betterness. Or they can insist that the harms deriving from uncertainty are not fully measured by wellbeing values, but at the cost of ceding that theirs is not adequate as a theory of *social* betterness.

Rejection of Pareto under Equivalent Outcomes looks more promising, for the axiom has the effect of forcing the social evaluator to ignore the manner in which individual prospects can combine to eliminate uncertainty about social outcomes.[xxxvi] Consider, for example the prospects displayed in Table 12.

[Insert Table 12 here]

In prospect I both Jocelyn and Kit face uncertainty about their wellbeing, but the social evaluator does not face any uncertainty about the goodness of the distribution of wellbeing to them. This is because, in both states, one individual has wellbeing one and another wellbeing zero and an impartial evaluator does not care which individual it is that gets wellbeing one. In such circumstances, should the social evaluator take into account the attitudes to uncertainty exhibited by Jocelyn and Kit? Suppose for instance that in the prospects displayed below, State 1 and State 2 are equiprobable (or, if it’s a situation of ambiguity, suppose that they are indistinguishable) and that for both Jocelyn and Kit, prospect II is strictly better than prospect I, in virtue of the fact that both are averse to the uncertainty contained in the latter.

Now the social evaluator too might regard prospect II as strictly better than prospect I because the social outcomes in the latter are more equal. But this reason is quite independent of those motivating Jocelyn’s and Kit’s ranking of the prospects. Indeed, a social evaluator who did not care about equality should regard the two prospects as equally good. More precisely if they regard the outcomes (0,1) and (0.5, 0.5) as equally good then State Dominance implies that I and II are also equally good. But this conflicts with any requirement of sensitivity to the (unanimous) uncertainty aversion of the individuals affected. Of course, the social evaluator may (and perhaps must) care about equality. But in this case the underlying tension between impartial rationality and benevolence is defused only if the social evaluator is inequality averse to *precisely* the degree that Jocelyn and Kit are uncertainty averse. And it is far from apparent why this should be required.

In this simple example, Strong Pareto is required to turn the tension between the requirements of sympathetic benevolence and minimal rationality into full-blown contradiction. (Which, in a nutshell, is why any theory that, like standard versions of *ex post* Prioritarianism, imposes both uncertainty neutrality in individual betterness and inequality aversion in social betterness must reject Strong Pareto.) But in the demonstrations given in earlier sections that both individual and social betterness must be ambiguity neutral and satisfy strong separability, more complicated prospects were constructed for which application of Pareto under Equivalent Outcomes sufficed to derive a similar contradiction. And this fact might point to grounds for rejecting this axiom, or at least one of the two clauses of it.

What might such grounds be? In weakening the requirement for sympathetic benevolence, I granted that when two prospects had different equality characteristics then this provided the social evaluator with a reason to overrule unanimities in individual betterness, but that in the absence of such differences they should be respected. Now allowing that uncertainty can impose harms (or benefits) on individuals in a manner that is not adequately captured by the Bayesian theory does not affect this argument, unless the kind of harm involved is morally irrelevant. This latter thought is not likely to be palatable to the advocate of non-Bayesian theories however.

There is another possibility: that Pareto under Equivalent Outcomes does not correctly identify all the cases in which the prospects being compared differ in characteristics that are irrelevant at the individual level but not the social one. The case for the first (Permutation) clause seems pretty solid because from the perspective of an impartial evaluator who ignores the identities of the individuals who are affected, prospects whose social outcomes are permutations of each other’s are essentially the same. And the sufficiency of this clause implies that, absent an argument for the moral irrelevance of the harms to individuals of uncertainty, we must accept that individual betterness should not be sensitive to uncertainty in any of its forms (including both risk and ambiguity) in ways that are incompatible with expected utility theory. On the other hand, the fact that *ex post* Prioritarianism conflicts with the second (Equality) clause of the axiom offers grounds for thinking that even when the social outcomes of prospects are perfectly equal in all states of the world they may differ in morally relevant ways. And dropping the Equality clause would open up the possibility of theories that allow for non-standard attitudes to uncertainty in social betterness (including, of course, *ex post* Prioritarian ones). But to take such a path requires identifying the respects in which such prospects might differ in morally relevant ways and it is not clear what those might be.

A final possibility. This paper has implicitly followed the majority of those working on rational judgement under ambiguity in treating ambiguity aversion as a property of a *complete* betterness relation on prospects. But one might regard the assumption of completeness as implausible for situations of ambiguity since, in such situations, the social evaluator would not be in a position to judge which of two prospects is better overall or whether they are equally good. Indeed, a number of authors have argued that, to the contrary, the rational response to lack of evidence or to disagreement amongst experts is partial suspension of judgement.[xxxvii] If this is correct, then models of impartial social evaluation should work with incomplete betterness relations for individuals (or, equivalently, *sets* of betterness rankings) and make corresponding adjustments to the conditions expressing the requirements of rationality and benevolence. Some early work of this kind was done by Isaac Levi within Sen’s social welfare functional approach and there has been recent work on incomplete social betterness rankings within a framework close to that of this paper.[xxxviii] But exploring this issue further must be left for another day.

Let me finish with some brief remarks about what the implications would be of accepting the full framework of Paretian social evaluation postulated in this paper for the question of what form social betterness judgements should take. We have seen that this framework blocks a certain class of departure from Harsanyi’s expectational Utilitarianism because it doesn’t allow the social evaluator to make judgements that are sensitive to risk or ambiguity (in ways not allowed by his theory). Nonetheless, it should be noted that it does *not* follow from our results that social evaluation must be Utilitarian. In fact, we have not even shown that either individual or social evaluation *must* take an expectational form, i.e., that the goodness of a prospect is a probability weighted average of the goodness of its outcomes. On the other hand, since our results *do* rule out the sorts of attitudes to risk and uncertainty that typically serve to motivate non-expectational theories, there are no compelling grounds not to assume that betterness judgements conform with the principles of expected utility theory. Even if do so, however, Harsanyi’s Utilitarian conclusion does not follow: in essence, because the adopted condition of sympathetic benevolence is much weaker than that necessitated by Utilitarianism (which implies satisfaction of the Strong Pareto condition). This leaves room open for many alternative expectational theories, notably including all the impartial members of the wide class of *ex post* Egalitarian theories characterized by Marc Fleurbaey which value a prospect as the expected value of the equally distributed equivalents of its possible social outcomes.[xxxix] To decide between these theories, however, we would need to determine what sort of strengthening of our baseline Paretian unanimity condition was justified, another task that is beyond the scope of this paper.

[*] I am very grateful to the Princeton University Centre for Human Values (UCHV) for the support that it provided for the year in which the research for this paper was conducted. This paper has benefited enormously from discussions with Alex Voorhoeve and with Marc Fleurbaey who, in particular, steered me towards the weaker Pareto condition. I am also grateful for comments on earlier drafts from Matt Adler, Greg Bognar, Campbell Brown, Paul Égre, Brian Hill, Orri Stéfansson, Peter Vallentyne and two anonymous referees for *Ethics* and to the audiences of the UCHV Fellows’ seminar at Princeton University, the Duke Philosophy colloquium and the LSE Choice Group.

[i] John C. Harsanyi, *“*Cardinal utility in welfare economics and in the theory of risk taking,*” Journal of Political Economy* 61 (1953)*:* 434–35 and John C. Harsanyi, “Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility,” *Journal of Political Economy* 63 *(*1955): 309–321.

[ii] A notable exception being Tom Rowe and Alexander Voorhoeve, “Egalitarianism under severe uncertainty,” *Philosophy and Public Affairs* 46 (2018): 239-268

[iii] This will also serve, I hope, to explain the significance of other results found in the more formal literature on this topic.

[iv] John Weymark, “A Reconsideration of the Harsanyi–Sen debate on utilitarianism.” In Jon Elster and John Roemer, eds., *Interpersonal Comparisons of Well-Being*. (Cambridge, UK: Cambridge University Press, 1991); Peter Fishburn, "On Harsanyi's utilitarian cardinal welfare theorem," *Theory and Decision* 17 (1984): 21-28; Peter Hammond, “Ex-

Ante and Ex-Post Welfare Optimality under Uncertainty," *Economica* 48 (1981): 235-250; John Broome, "Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism," *The Review of Economic Studies* 57 (1990): 477–502

[v] Marc Fleurbaey, "Two variants of Harsanyi's aggregation theorem," *Economics Letters* 105(2009): 300–302; Marc Fleurbaey, "Assessing risky social situations," *Journal of Political Economy* 118 (2010):649–80; David McCarthy, Kalle Mikkola and Teruji Thomas, "Utilitarianism with and without expected utility," *Journal of Mathematical Economics* 87 (2020): 77-113 and Philippe Mongin and Marcus Pivato, *"*Ranking multidimensional alternatives and uncertain prospects,*" Journal of Economic Theory* 157 (2015): 146–71.

[vi] Maurice Allais, "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine," *Econometrica* 21 (1953): 503–546.

[vii] Itzhak Gilboa and David Schmeidler, "Maxmin expected utility with nonunique prior," *Journal of mathematical economics*, 18 (1989) : 141-153; Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji, "A smooth model of decision making under ambiguity," *Econometrica*, 73 (2005): 1849-1892; John Quiggin, *Generalized Expected Utility Theory: The Rank-Dependent Model* (Dordrecht: Kluwer, 1993); Amos Tversky and Peter Wakker, "Risk attitudes and decision weights," *Econometrica* 63 (1995): 1255–1280; Lara Buchak, *Risk and Rationality* (Oxford: Oxford University Press, 2013).

[viii] In particular, by Mongin and Pivato *"*Ranking multidimensional alternatives and uncertain prospects". The main difference between their results and those of this paper stem from the adoption here of a weaker Pareto condition.

[ix] See Matthew Adler, *Measuring Social Welfare: an introduction* (New York: Oxford University Press, 2019) for a recent survey and discussion.

[x] These definitions must be generalized for situations of ambiguity, but the details do not affect the arguments that I will be making.

[xi] I am implicitly assuming here that the social evaluator's judgement of the goodness of a social outcome is independent of the state in which it is realized. This is done for reasons of simplicity and none of the results that follow depend on it.

[xii] Leximin ranks one social outcome over another if the wellbeing of the least well-off individual is higher in the former than the latter. If they are equally well-off it ranks one over the other if the wellbeing of the second least well-off individual is higher in the former than the latter. And so on.

[xiii] Campbell Brown has pointed out to me that my results require only that social betterness satisfies a weaker notion of impartiality: indifference between a social prospect and any involutory permutation of it, i.e., a permutation of it that is its own inverse.

[xiv] John Broome*, Weighing Goods. Equality, Uncertainty and Time* (Oxford: Oxford University Press, 1991)

[xv] A prominent class of *ex post* Egalitarian theories – those that maximise the expectation of an equally distributed equivalent – are partially characterized by their satisfaction of Pareto for Equivalent Outcomes. Indeed, it is this fact that marks them out from *ex post* Prioritarian theories that maximise the sum of a concave function of individual wellbeing (see Fleurbaey "Two variants of Harsanyi's aggregation theorem").

[xvi] By 'standard' versions I mean those that accept the so-called Bernoulli condition: that the goodness of an individual prospect is measured by the expectation of wellbeing that it induces.

[xvii] See Fleurbaey, "Two variants of Harsanyi's aggregation theorem" and Michael Otsuka and Alexander Voorhoeve, "Why it Matters that Some are Worse Off than Others: An Argument Against the Priority View," *Philosophy and Public Affairs* 37 (2009): 171-99 for more detailed versions of this criticism and Matthew Adler and Nils Holtug, "Prioritarianism: A response to critics," *Politics, Philosophy and Economics* 18 (2019): 101-144 for a defence of *ex post* Prioritarianism against it.

[xviii] See Philippe Mongin, "Spurious unanimity and the Pareto Principle," *Economics and Philosophy* 32 (2016): 511-532.

[xix] See Philippe Mongin, "Consistent Bayesian aggregation," *Journal of Economic Theory* 66 (1995): 313–51, Richard Bradley, "Bayesian Utilitarianism and Probability Homogeneity," *Social Choice and Welfare* 24 (2005): 221-251 and Broome, *Weighing Goods*.

[xx] Daniel Ellsberg, "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics* 75 (1961): 643-669.

[xxi] See Itzhak Gilboa and Massimo Marinacci, "Ambiguity and the Bayesian Paradigm," in *Readings in Formal Epistemology*, eds. Horatio Arló-Costa, Vincent Hendricks and Johan van Benthem, (Dordrecht: Springer, 2016); Isaac Levi, *Hard Choices: Decision Making under Unresolved Conflict* (Cambridge: Cambridge University Press, 1986); James Joyce, "A defence of imprecise credences in inference and decision making," *Philosophical Perspectives* 24 (2010): 281-323 and Richard Bradley, *Decision Theory with a Human Face* (Cambridge: Cambridge University Press, 2017).

[xxii] Rowe and Voorhoeve, "Egalitarianism under severe uncertainty".

[xxiii] See Nabil Al-Najjar and Jonathan Weinstein, "The ambiguity aversion literature: A critical assessment," *Economics & Philosophy* 25 (2009): 249–84 and Marc Fleurbaey, "Welfare economics, risk and uncertainty," *Canadian Journal of Economics* 51 (2018): 5–40

[xxiv] John Broome, *Climate Matters: Ethics in a Warming World* (New York: WW Norton & Company, 2012).

[xxv] Stefan Trautmann and Gijs van de Kuilen, "Ambiguity attitudes," in *The Wiley-Blackwell Handbook of Judgment and Decision Making*, Volume 1, eds. Gideon Keren and George Wu, (Chichester, UK: Wiley-Blackwell, 2015).

[xxvi] More formally, the preferences postulated by Ellsberg violate the Sure-thing Principle, the main separability condition of standard subjective expected utility theory. Ambiguity aversion is not necessary for such a violation however and so I defer discussion of the Sure-thing Principle till the next section.

[xxvii] See Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler, "Is It Always Rational to Satisfy Savage's Axioms?," *Economics and Philosophy* 25 (2009): 285-296.

[xxviii] See Amos Tversky and Daniel Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty* 5 (1992): 297–323, Buchak, *Risk and Rationality* and Quiggin, *Generalized Expected Utility Theory*.

[xxix] Lara Buchak, "Taking Risks behind the Veil of Ignorance," *Ethics* 127 (2017): 610–44.

[xxx] Simon Blessenohl, "Risk Attitudes and Social Choice," *Ethics* 130 (2020): 485-513 and Jacob Nebel, "Rank-Weighted Utilitarianism and the Veil of Ignorance," *Ethics* 131 (2020): 87–106.

[xxxi] Because it was first presented in Allais, "Le comportement de l'homme rationnel devant le risque".

[xxxii] See Broome*, Weighing Goods* for an argument of just this kind.

[xxxiii] This is demonstrated in H. Orri Stefánsson and Richard Bradley, "What is Risk Aversion?," *British Journal for the Philosophy of Science* 70 (2019): 77-102

[xxxiv] He presents the paradox in Peter Diamond, "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: Comment," *Journal of Political Economy* 75 (1967): 765–66.

[xxxv] Such as that axiomatized in Larry Epstein and Uzi Segal, "Quadratic Social Welfare Functions," *Journal of Political Economy* 100 (1992): 691–712.

[xxxvi] Rowe and Voorhoeve, "Egalitarianism under severe uncertainty" make just this criticism of Strong Pareto.

[xxxvii] For instance, Levi, *Hard Choices*; Joyce, “A defence of imprecise credences in inference and decision making,” and Bradley, *Decision Theory with a Human Face*.

[xxxviii] See for instance Isaac Levi, "Pareto Unanimity and Consensus," *The Journal of Philosophy* 87 (1990): 481-92 and Eric Danan, Thibault Gajdos, Brian Hill, and Jean-Marc Tallon, "Robust Social Decisions," *American Economic Review*, 106 (2016): 2407-25

[xxxix] Fleurbaey, “Assessing risky social situations,”