

**ON KOLM'S USE OF EPISTEMIC COUNTERFACTUALS IN SOCIAL CHOICE THEORY**

by

John A. Weymark



**Working Paper No. 05-W18R**

June 2005

Revised May 2007

DEPARTMENT OF ECONOMICS  
VANDERBILT UNIVERSITY  
NASHVILLE, TN 37235

[www.vanderbilt.edu/econ](http://www.vanderbilt.edu/econ)

---

# On Kolm's Use of Epistemic Counterfactuals in Social Choice Theory

John A. Weymark

Department of Economics, Vanderbilt University, VU Station B #35189,  
2301 Vanderbilt Place, Nashville, TN 37235-1819, U.S.A.  
E-mail: john.veymark@vanderbilt.edu

## 1 Introduction

In a series of writings, Serge Kolm has examined the conceptual foundations of Arrovian social choice theory.<sup>1</sup> In its choice-theoretic formulation, an Arrovian social choice correspondence specifies the socially best alternatives from each admissible feasible set of alternatives as a function of the individual preferences over the universal set of alternatives. An issue that needs to be addressed when constructing a social choice correspondence is: For what feasible sets of alternatives and for what preference profiles are social decisions required? The rationale for the choice of this domain is one of the issues that Kolm discusses at some length. When considering resource allocation problems, social alternatives are meant to be complete descriptions of all the features of a social state relevant to choice, including future allocations of resources. It would then seem to follow that there is only one choice situation for which a social choice is required, the situation characterized by the actual feasible set and the actual preference profile. However, Arrow's famous impossibility theorem (see Arrow, 1951, 1963) requires the domain of a social choice correspondence to be reasonably rich. Indeed, many of Arrow's axioms are vacuous if there is only one feasible set and one preference profile. Arrow (1951, p. 24) appeals to uncertainty about what the actual choice environment is at the time social decisions are being made to justify having a non-singleton domain. Kolm (1996), however, believes that it is difficult to reconcile this rationale for the choice of domain with the adoption of the axioms Arrow proposes for relating choices in different choice environments.

Kolm (1993, 1995, 1996, 1997) has proposed a different rationale for considering more than one choice situation in the domain of an Arrovian social choice correspondence—his “epistemic counterfactual” principle. In this rationale, a social choice only needs to be made from the actual feasible set given the actual profile of preferences, but in order for this choice to be justified,

---

<sup>1</sup> See Kolm (1993, 1994, 1995, 1996, 1997).

good reasons must be provided for this choice. These reasons involve comparing the choice that is actually made with what would have been justifiably chosen in appropriate counterfactual feasible sets with appropriate counterfactual preference profiles. In other words, we need to consider a social choice rule that specifies what is chosen in a range of test situations in order to determine if the choice in the actual choice situation is justified. Thus, the process of justification involves a thought experiment in which hypothetical choice environments are envisaged, and the choices that are recommended in these environments provide grounds for regarding the choice in the actual situation as being justifiable. Together with the actual feasible set and the actual preference profile, these hypothetical choice situations constitute the domain for the social choice correspondence.

Kolm's epistemic counterfactual principle provides a rationale for including preference profiles and feasible sets in the domain of a social choice correspondence in addition to the actual profile and the actual feasible set. However, this principle does not provide any guidance as to what these alternative choice situations should be. The purpose of this article is to consider this issue. In my investigation, I shall simply take the premises of Kolm's argument (i.e., that there is only one actual social decision to be made, but this decision must be justified by the choices that would have been made in appropriate counterfactual choice situations) as given and focus on what his principle implies for the domain of a social choice correspondence.

The choice of an appropriate set of counterfactual choice situations, and hence the choice of an appropriate domain for the social choice correspondence, depends on the purpose for which the choice is being made. Kolm is interested in problems of justice. Impartiality is an essential feature of all theories of justice and, more generally, of theories of morality.<sup>2</sup> Given the absence of any interpersonal comparisons of well-being in the Arrovian set-up, it is perhaps expecting too much to ask an Arrovian social choice correspondence to embody principles of justice, but it is not asking too much for Arrovian social choices to be impartial. Accordingly, I shall suppose that the objective is to make an impartial social choice. Different concepts of impartiality impose different constraints on choice, and so will result in the consideration of different sets of counterfactual choice situations. For example, impartiality might be with respect to personal identity, generation, or conception of the good. The theory of justice proposed by Rawls (1971, 1993) is impartial in all three of these senses, while the theories of Harsanyi (1953, 1955, 1977), Vickrey (1945, 1960, 1961), and Hare (1952, 1963, 1981) are only impartial with respect to personal identity.

Harsanyi, Rawls, and Vickrey all make use of the device of a veil of ignorance to help determine substantive principles of justice. A veil characterizes

---

<sup>2</sup> Kolm argues that impartiality is implied by the requirement that a social choice should be rational (i.e., justifiable). See, for example, Kolm (1996, p. 185). My arguments do not depend on the validity of this claim.

the information that is morally relevant for making social decisions, with different conceptions of impartiality resulting in different specifications of the veil. Harsanyi and Vickrey have a “thin” veil of ignorance because personal identity is the only morally irrelevant information in their theories. Rawls has a thicker veil because his principles are not only to be impartial with respect to personal identity, they are also to be impartial with respect to generation and with respect to conception of the good. Here, I argue that a veil of ignorance is a natural way to generate the set of counterfactual choice situations required by Kolm's epistemic counterfactual principle when the objective is to make an impartial social choice. In my argument, the relevant Arrovian domain is the set of all choice situations that a moral agent behind a veil of ignorance thinks might be the actual choice situation outside the veil. The size of this domain is positively related to the thickness of the veil. Thus, my argument supplements Kolm's epistemic counterfactual principle by providing an account of how the appropriate domain for a social choice correspondence is to be determined, at least when the objective is to justify the actual social choice impartially.

Once this domain has been identified, one can then ask if it is rich enough for all of Arrow's axioms to be nonvacuous. I shall argue that this is the case when the domain is generated by a Rawlsian veil, but not when it is generated by a Harsanyi–Vickrey veil.

In the next section, I describe Arrow's formulation of the social choice problem and the choice-theoretic version of his impossibility theorem. In Section 3, I consider a number of rationales that have been proposed for including more than one choice situation in the domain of a social choice correspondence. In Section 4, I discuss Kolm's epistemic counterfactual principle. Section 5 sets out in a preliminary way my procedure for using a veil of ignorance to generate the set of counterfactual choice situations required by Kolm's principle. The following two sections examine the Harsanyi–Vickrey and Rawlsian veils, and how they might be used in my argument, in more detail. In Section 8, I consider the possibility of using an ideal observer theory, Hare's universal prescriptivism, or a particular concept of impartiality directly to generate the required domain for the social choice correspondence. Section 9 offers some concluding remarks.

## 2 Arrow's Theorem

Arrow's Theorem demonstrates that it is impossible for any collective decision-making procedure to satisfy a number of properties that *a priori* one might think any reasonable collective decision-making procedure should satisfy. Arrow's Theorem can be formulated in terms of either a social welfare function or a social choice correspondence. A *social welfare function* assigns a social ordering (i.e., a reflexive, complete, and transitive binary relation) of the alternatives to each admissible profile of individual preference orderings of the

alternatives. A *social choice correspondence* specifies, for each admissible feasible set of alternatives and each admissible profile of individual preferences on the universal set of alternatives, a subset of the feasible alternatives—the *social choice set*. For brevity, I often refer to a feasible set of alternatives as an *agenda* and refer to a combination of an agenda and a preference profile as a *choice situation*. The original versions of Arrow’s Theorem, as found in Arrow (1951, 1963), were expressed in terms of social welfare functions, but it is a straightforward exercise to restate Arrow’s Theorem in terms of social choice correspondences. Formal statements of both the welfare-theoretic and choice-theoretic versions of Arrow’s Theorem may be found in Donaldson and Weymark (1988) and Le Breton and Weymark (2005).

The two approaches are closely related. A social welfare function can be used to construct a social choice correspondence as follows. For any choice situation, the social welfare function is first used to determine which social ordering of the universal set of alternatives is assigned to the profile of individual preferences. Then the social choice set is determined by maximizing this social ordering on the feasible set of alternatives. This is the procedure Arrow (1951, p. 26) uses to determine the social choice set in any admissible choice situation. While it is possible to generate a social choice correspondence in this fashion, it is not necessary to do so; one could instead determine the social choice set directly from knowledge of the choice situation without the intermediary of a social welfare function. This direct approach to making social choices is more general than the two-step procedure outlined above because there exist social choice correspondences for which the choices made from the various agendas for a given profile of preferences are not rationalizable by any ordering of the alternatives.

My concern is with social choice, not the social ranking of alternatives, and so I use the social choice formulation of Arrow’s problem. The *domain* of a social choice correspondence is the collection of admissible choice situations; i.e., the combinations of preference profiles and agendas for which a social choice is to be made. Arrow’s impossibility theorem assumes that the domain of the social choice correspondence is quite rich. In standard choice-theoretic versions of this theorem, the domain includes (a) all conceivable profiles of individual preference orderings as possible profiles and (b) all nonempty finite subsets (or at least all two- and three-alternative subsets) of the set of alternatives as possible agendas. Arrow (1963, p. 97), however, noted that his preference domain assumption is “unnecessarily strong.”

Kolm (1993, 1994, 1995, 1996, 1997) has been particularly critical of the assumption that there is an unrestricted domain of preference profiles, arguing that this results in “absurd” preferences being considered. For example, he notes that such a domain would require the consideration of preferences in which it is always preferable to have less of every good. However, it has been shown that the social welfare function version of Arrow’s Theorem holds on quite restricted domains, including domains that satisfy the kinds of assumptions normally made in economic models. For social choice correspondences,

the inconsistency of Arrow's axioms is less robust than in the welfare-theoretic case. On some natural domains, Arrow's axioms are consistent, whereas on some other natural domains, they are either inconsistent or only consistent with rather undesirable choice rules. See Le Breton and Weymark (2005). While these results demonstrate that one does not need to consider absurd choice situations to establish versions of Arrow's Theorem, they show that a reasonably rich domain is required if the properties Arrow proposed for a social choice correspondence are to have much force.

Given the domain assumptions described above, the choice-theoretic version of Arrow's Theorem shows that it is not possible to jointly satisfy Weak Pareto, Independence of Infeasible Alternatives, Arrow's Choice Axiom, and Nondictatorship.

*Weak Pareto* applies to each choice situation in the domain separately. This axiom says that a feasible alternative must not be chosen if there is another feasible alternative that everyone strictly prefers.

*Independence of Infeasible Alternatives* applies to pairs of choice situations that share a common agenda and in which both preference profiles agree on this agenda (but may differ on nonfeasible alternatives). This axiom requires the social choice set to be the same whenever the two choice situations are related in this fashion. Informally, the chosen alternatives only depend on the individual preferences for feasible alternatives.

*Arrow's Choice Axiom*, first introduced in Arrow (1959), requires that, holding the preference profile fixed, if the feasible set shrinks and there is some alternative that was originally chosen still feasible, then any of the originally chosen alternatives that remain feasible must continue to be chosen and no alternative that was originally rejected should now be chosen. Arrow's Choice Axiom is an example of what is known as a collective rationality assumption.

An individual is a *dictator* if for every choice situation, each alternative in the social choice set is one of this individual's most-preferred choices on the agenda. If the dictator has more than one best choice on an agenda, it is permissible (but not required) to select among them. *Nondictatorship* requires that no such individual exists.

Whether these axioms are reasonable or desirable constraints on social choice is not my concern here.<sup>3</sup> Kolm, in the works cited above, has expressed reservations about many features of Arrow's formulation of the social choice problem, including the axioms.

### 3 Rationales for Non-Singleton Domains

In both its choice-theoretic and welfare-theoretic formulations, Arrow's Theorem supposes that every conceivable profile of individual preference orderings

---

<sup>3</sup> Formal statements and a discussion of these axioms may be found in Le Breton and Weymark (2005).

is included in the domain. In the choice-theoretic version of the theorem, it is also supposed that choices must be made from all possible finite agendas (or all those agendas containing two or three alternatives). In one of his earliest contributions, Kolm (1969, p. 168) says that

... the object of a normative theory is to define and to determine the optimum (i.e. to point out the right public action) under the given, and unique, set of constraints (which may be uncertain, dated for the future, etc.); what should be done *if* there are other constraints is simply irrelevant to the problem.

Expressed in terms of a social choice correspondence, his point is that there is only one choice situation—the actual one—and this is the only relevant choice situation. Accepting this view would imply that the domain of the social choice correspondence should only contain the actual choice situation, in which case both Arrow's Choice Axiom and Independence of Infeasible Alternatives would be vacuous.

As we shall see in the next section, Kolm now believes that the domain should contain more than one choice situation. However, before presenting the details of Kolm's argument, it is useful to see how the domain issue is dealt with in Bergson (1938)–Samuelson (1947) welfare economics and in Arrovian social choice theory.

Little (1952), Samuelson (1967), and others have suggested that a fundamental difference between a Bergson–Samuelson social welfare function and an Arrovian social welfare function is that the former is defined for only a single preference profile, whereas the latter is defined for all conceivable profiles.<sup>4</sup> A Bergson–Samuelson social welfare function is a function that represents an ordering of the social alternatives.<sup>5</sup> Bergson and Samuelson use their kind of social welfare function to characterize the optimal choice from a given agenda when the alternatives are allocations of economic goods. Thus, they use the two-step procedure described above to determine the social choice correspondence. From their writings, it is not very clear whether a Bergson–Samuelson social welfare function is meant to be used to determine the social choice for different agendas or for just the current feasible set. In the subsequent literature, the ordering of alternatives given by a Bergson–Samuelson social welfare function is used to determine the social choice set from alternative agendas.<sup>6</sup> Thus, a Bergson–Samuelson social welfare function is the foundation for a

<sup>4</sup> See also Bergson (1954, p. 247). Little (1952, p. 423) acknowledges that he is indebted to Samuelson (in conversation) for this observation. Kemp and Ng (1976), Parks (1976), and Pollak (1979), among others, have established versions of Arrow's Theorem in its social welfare function formulation for a domain consisting of a single preference profile. These versions of the impossibility theorem utilize an intraprofile independence condition that Samuelson (1977) finds quite unreasonable.

<sup>5</sup> See Bergson (1938, p. 312) and Samuelson (1947, p. 221).

<sup>6</sup> Little (1952, p. 423) is explicit on this point.

social choice correspondence that has more than one choice situation in its domain, but each of these choice situations only differs in the agenda being considered, not in the preference profile.

This approach does not preclude considering different preference profiles. As Little (1952, pp. 423–424), defending the Bergson–Samuelson approach, puts it:

If tastes change, we may expect a new ordering of all the conceivable states; but we do not require that the difference between the new and the old ordering should bear any particular relation to the changes of taste which have occurred. We have, so to speak, a new world and a new order; and we do not demand correspondence between the change in the world and the change in the order.<sup>7</sup>

Furthermore, Little (1952, p. 423) says that:

If, for a given set of tastes, the environment varies we expect that the choices will be consistent in the sense that the choice function is derivable from a weak ordering of all social states.

Thus, each profile defines a new Bergson–Samuelson social welfare function that is used to determine the choices from the admissible agendas. Piecing together these choices for the alternative profiles and the alternative agendas, we obtain an Arrovian social choice correspondence (on a possibly restricted domain).<sup>8</sup>

Kolm (1996, p. 180) has suggested that one possible reason for considering more than one choice situation is that there is a single set of individuals, each with preferences that are temporally independent, facing a sequence of undated agendas (so that the universal set of alternatives is the same in each of the choice situations). With the further proviso that individual preferences do not change from period to period, we have the kind of situation Bergson–Samuelson social welfare functions were designed to handle. Different agendas arise through time as technologies and, hence, feasible sets change.<sup>9</sup>

It is also possible to regard the various possible agendas as arising not because there is a sequence of agendas, but because there is uncertainty about

<sup>7</sup> For further discussion of this point, see Fleurbaey and Mongin (2005, Section 2).

<sup>8</sup> The passages quoted above appear as part of Little's critique of Arrow's independence axiom. As is clear from these quotations, Little does not believe that this axiom is compelling, whereas he does believe that a collective rationality axiom that places cross-agenda restrictions on choice for a given preference profile, such as Arrow's Choice Axiom, is justifiable. His rationale for this position may be found in Little (1952, p. 424). For critical discussions of Little's arguments, see Kolm (1993, Section 13), Mongin and d'Aspremont (1998, p. 423), and Pattanaik (2005, Section 3.3).

<sup>9</sup> It is possible to interpret the rationale proposed by Little (1952, p. 423) for considering different agendas in this manner. Kolm (1993, Section 6) notes that this kind of rationale is inappropriate with dated alternatives.



what the actual agenda will be at the time social decisions are made. Restated in terms of preference profiles, this is the rationale Arrow used to justify the presence of alternative profiles in the domain of his social welfare function.

If we do not wish to require any prior knowledge of the tastes of individuals before specifying our social welfare function, that function will have to be defined for every logically possible set of individual orderings. . . . However, we may feel on some sort of a priori grounds that certain types of individual orderings need not be admissible. (Arrow, 1951, p. 24)

Implicit in Arrow's argument (as applied to social choice correspondences) is the assumption that while the social choice correspondence must be specified before the uncertainty is resolved, the actual implementation of the choice does not occur until the actual choice situation is known.<sup>10</sup>

#### 4 Kolm's Epistemic Counterfactual Principle

The epistemic counterfactual principle proposed by Kolm (1993, 1995, 1996, 1997) provides a further rationale for considering a non-singleton domain. In contrast to Arrow, Kolm assumes that the actual choice situation is known when any decision must be made. Nevertheless, while it is only necessary to make a choice in the actual choice situation, this choice needs to be justified by the choices that would have been made in certain counterfactual choice situations. There must be a good reason that justifies the actual choice and it is the choices that would have been made in these hypothetical choice situations that provide this reason.<sup>11</sup>

A natural way of interpreting Kolm's principle is as follows. To see whether the choice made in the actual choice situation is justified, we need to consider a social choice correspondence defined on an appropriate domain of choice situations. If, for example, our objective is to choose among the alternatives impartially, then we need to determine if the choice in the actual choice situation is consistent with making impartial decisions in a range of choice situations in which the roles of individuals are interchanged in some way. More generally, it is by testing our intuitions about what choices are reasonable in a range of

<sup>10</sup> Kolm believes that using Arrow's argument for having more than one choice situation in the domain makes it difficult to justify the adoption of the Arrowian axioms that place restrictions on how choices from different choice situations are related to each other. See, for example, Kolm (1996, Section 4.2) or Kolm (1997, Section 15.2).

<sup>11</sup> As Kolm notes, counterfactual choice situations can also provide, in principle, a role for the axioms that only apply when there is a non-singleton domain. Arrow (1997, pp. 4–5) has made a similar point, arguing that when there is only a single election, the application of his independence and collective rationality axioms require the use of counterfactuals.

choice situations that we can determine whether the actual choice is justified or not. In effect, the social choice correspondence provides the reasons for making the choices, both actual and hypothetical. Thus, we need to determine if the social choice correspondence conforms with the principles that we want to govern the collective decision-making process and, if so, whether these choices are consistent with our considered judgments. If they are, then this social choice correspondence provides a good reason for making the prescribed choice in the actual choice situation.

According to Kolm's epistemic counterfactual principle, the need to justify the actual social decision provides a reason for not restricting the domain of a social choice correspondence to the actual choice situation, but it does not, by itself, provide guidance as to which counterfactual choice situations should be considered. In other words, there remains the question of determining the appropriate domain for the social choice correspondence. Kolm (1993, Section 12) raises this issue, but does not provide a specific proposal for its solution. Without specifying the criteria by which the actual choice should be judged, it is not possible to determine the set of counterfactual choice situations that are needed to apply Kolm's principle. In order to show how Kolm's principle can be used to determine an appropriate domain for a social choice correspondence when the actual choice situation is known, in subsequent sections, I shall suppose that the objective is to choose alternatives impartially. Different concepts of impartiality will be seen to lead to different specifications of this domain.

It should be emphasized that the application of Kolm's epistemic counterfactual principle involves three distinct steps. First, the domain of the social choice correspondence must be specified. In other words, we first need to determine the set of "test" choice situations that should be considered in order to justify the choice in the actual choice situation. Second, a social choice correspondence with this domain is specified. The social choice set recommended in the actual choice situation by this correspondence is required to coincide with the actual choice set. Third, we need to determine if the social choice correspondence embodies the principle or principles that the decision-making procedure is meant to conform to. If it does and the choices that it recommends are consistent with our considered judgments, then the choice made in the actual choice situation has been justified. My concern here is with identifying the domain needed to apply Kolm's epistemic counterfactual principle when choices are to be made impartially, and not with whether a social choice correspondence on this domain does in fact justify the actual social choice set. According, I focus on the first step in this procedure.

Kolm has argued that his epistemic counterfactual principle should take a more specific form than I have described above. See, for example, Kolm (1996, p. 180). To justify the choice of  $x$  when the agenda is  $X$ , he argues that there must exist another agenda  $Y$  containing  $x$  in which some other alternative  $y$  is chosen. Presumably, it is permissible to vary the profile as well as the agenda in this thought experiment, although Kolm is not very clear on this point. One

can adopt the general version of the epistemic counterfactual principle without committing oneself to this specific version, and that is what I shall do here. The reason why I only adopt the general, and not the particular, features of Kolm's counterfactual argument can be illustrated with the following example.

Suppose that there are two individuals who differ in their productive abilities and that there are two goods, jam and leisure. For concreteness, let person one be the low productivity individual and person two be the high productivity individual. A social alternative specifies how much of each of the two goods each person consumes. Both individuals have the same preferences for jam and leisure (which is a normal good), but person two earns more than person one if they both work the same number of hours. While the government knows that one person has a high productivity, it does not know which individual is the more productive. As a consequence, to pursue its redistributive goals, the government is not able use personalized lump-sum taxes and transfers. Instead, the government chooses a single tax schedule that determines the tax paid (or the transfer received) by any individual as a function of the income earned.

The model I have just described is the two-person version of the Mirrlees (1971) optimal income taxation problem, as presented in Stiglitz (1982). In this model, any feasible alternative has the property that the jam consumption of person one does not exceed the jam consumption of person two.<sup>12</sup> Let me further suppose that the unique social choice, call it  $x$ , involves giving person two strictly more jam consumption than person one (which turns out to also require that person two earns more than person one), perhaps because doing so results in both individuals consuming more jam than would be possible if jam consumption were equalized. Stiglitz (1982, p. 220) refers to this outcome as the "normal" case.

Assuming that one of our objectives is to treat individuals impartially, a natural counterfactual to this choice situation is obtained by permuting the individuals in this example. This involves not only permuting who has which preference and productivity, but also permuting who gets what consumption bundle in each feasible allocation. Symmetry considerations suggest that the social choice in the new choice situation, call it  $y$ , should be the allocation that is obtained by permuting the consumption bundles in  $x$ . This symmetry argument provides a counterfactual justification for choosing  $x$  in the actual choice situation—the choice of  $x$  is justified because if the roles of the individuals had been reversed, the symmetric allocation  $y$  would have been chosen. While this justification for choosing  $x$  conforms to the general form of the Kolm's epistemic counterfactual principle, it does not satisfy Kolm's more specific *desiderata* because  $x$  is not in the counterfactual agenda. Person two has more jam consumption in  $x$  than person one, but in the counterfactual situation person two has the lower productivity and so, as a consequence of

<sup>12</sup> This is only a necessary, and not a sufficient, condition for an alternative to be feasible.

the incentive constraints in this problem, cannot receive a higher jam consumption.<sup>13</sup>

## 5 A Veil of Ignorance as a Source of Counterfactuals for Impartial Social Choice

Kolm (1993, 1995, 1996, 1997) is interested in determining the extent to which the Arrovian paradigm, viewed as a way of formalizing normative decision-making, can contribute to social ethics. Kolm is particularly concerned with problems of distributive justice—how to reconcile the conflicting claims of the individuals in society. While it is commonly held that impartiality is an essential feature of all moral theories, including theories of justice, there is considerable disagreement as to what exactly is meant by impartiality and as to how impartiality should be incorporated into a moral theory. In this section, I propose a procedure for identifying an appropriate domain for an Arrovian social choice correspondence when the objective is to determine an impartial social choice.<sup>14</sup>

My procedure is based on the ways in which impartiality is incorporated into the original position theories of Harsanyi (1953, 1955, 1977), Rawls (1971, 1993), and Vickrey (1945, 1960, 1961). In each of these theories (as well as in many other ethical theories), principles of morality or justice are the principles that would be proposed by an individual who adopts an impartial perspective when making his or her recommendations. In the subsequent discussion, I shall refer to such an impartial individual as a *moral agent*. Following Kolm, I suppose that there is only one actual choice situation, but the choice made in this situation must be justified by the choices that would have been made in certain counterfactual choice situations. In my argument, the devices that have been used by Harsanyi, Rawls, and Vickrey to derive substantive moral principles are instead used to generate the domain for an Arrovian social choice problem.

The original position theories of Harsanyi, Rawls, and Vickrey suppose that principles of morality or justice are embodied in the decision rules that would be adopted in certain idealized *individual* choice problems. In these theories, social decisions are made by a rational moral agent. An *original position theory* supposes that a social decision (or the principle that is to be appealed to when making social decisions) is justified if it would be recommended by a rational, self-interested individual who is deprived of certain

<sup>13</sup> This example demonstrates that counterfactual reasoning need not take the specific form suggested by Kolm. Nevertheless, in other examples, choices may well be justified using counterfactual arguments that do satisfy Kolm's structural conditions.

<sup>14</sup> It is widely believed that ordinal preference information is an inadequate foundation for a theory of justice (or any other moral theory), so I merely require a social choice to be impartial, rather than to be fully just.

information, including the decision-maker's identity. An *original position* is simply any hypothetical decision problem with these features. The informational constraints facing a moral agent in an original position describe what Rawls calls a "veil of ignorance." The idea of an original position (although not the terminology) has its origin in the work of Harsanyi (1953, 1955) and Vickrey (1945), who independently proposed what is essentially the same description of an original position. Interest in original position theories became much more widespread with the appearance of Rawls' *A Theory of Justice* (Rawls, 1971). While sharing with Harsanyi and Vickrey the device of an original position, Rawls' description of an original position differs markedly from that of Harsanyi and Vickrey.<sup>15</sup>

Barry (1973, p. 10) has observed that the features characterizing an original position are of two kinds: "... those which concern knowledge and those which concern motivation." In Barry (1989, Chapter 9), theories of justice in which the substantive principles of justice are determined by the choices made by moral agents (whether or not the moral agents are self-interested or operate behind a veil of ignorance) are classified along these two dimensions.<sup>16</sup> In an original position theory, the moral agent pursues his or her own self-interest to the extent that this is possible given the information available and does so rationally in the sense of choosing appropriate means to achieve his or her objectives. Further, the moral agent is stripped of all morally irrelevant information, including his or her own identity, but is provided with all morally relevant information. Alternative original position theories differ in what they regard as morally relevant information.

Original position theories have been used to justify the choice of principles of justice, moral rules, social institutions, and particular social alternatives. Here, I suggest that the device of a veil of ignorance can be used to help generate the set of counterfactual choice situations required to justify the actual social decision made in an Arrovian social choice problem when this decision must be made impartially. Because I am only concerned with identifying the domain needed to apply Kolm's epistemic counterfactual principle, and not with specifying the social choice correspondence on this domain that is used in the justificatory process, the motivations ascribed to a moral agent behind the veil of ignorance are irrelevant, although, of course, these motivations play a fundamental justificatory role in original positions theories. The relevant domain for the problem I am considering depends only on the informational constraints that characterize the veil of ignorance. The veil of ignorance specifies the morally relevant information and it is the description of the veil, not the motivational assumption, that distinguishes the original

<sup>15</sup> In describing Rawls' original position in terms of a single moral agent choosing principles of justice from behind a veil of ignorance, I am following the description found, for example, in Rawls (1971, p. 139).

<sup>16</sup> Barry (1989) refers to all such theories as "original position theories," but it is more usual to use this term in the more restricted way that I do here.

position of Harsanyi and Vickrey from that of Rawls. Harsanyi and Vickrey disagree with Rawls as to which information is morally relevant for choice in the original position, with Harsanyi and Vickrey endowing moral agents with much better information about the particulars of the world outside the veil than does Rawls. These differences in what is viewed as morally relevant information largely account for the substantially different principles advocated by Harsanyi and Vickrey on the one hand and Rawls on the other.

It is now possible to describe more precisely my procedure for using original position theories to generate the Arrovian domain needed to justify an impartial social choice. In my procedure, this domain consists of all the choice situations one might think could be the actual choice situation given the information available behind the veil of ignorance. The size of the domain obtained in this fashion depends on what is considered to be morally relevant information; the thicker is the veil of ignorance, the larger is the set of possible counterfactual choice situations. If, as in Harsanyi's and Vickrey's theories, there is a "thin" veil of ignorance, then the domain is not very rich. On the other hand, if, as in Rawls' theory, there is a "thick" veil of ignorance, then a rather large set of counterfactuals is obtained. In the next two sections, this general procedure is used to construct the domains implied by the Harsanyi–Vickrey and the Rawlsian veils.

## 6 A Harsanyi–Vickrey Veil

The first original position arguments were developed independently by Vickrey (1945) and Harsanyi (1953, 1955). The basic features of their original positions are similar, although they do differ in some respects, including the kinds of social alternatives that they consider.<sup>17</sup>

A moral agent behind the Harsanyi–Vickrey veil of ignorance chooses without knowing his or her actual identity, but with full knowledge of all other features of the social alternatives. For Harsanyi and Vickrey, personal identities are morally irrelevant, but everything else about social alternatives is morally relevant information. More precisely, when considering a social alternative, a moral agent imagines that he or she has an equal chance of being any particular person in society when the veil is lifted.<sup>18</sup> Thus, the thought

<sup>17</sup> Because Vickrey believes that interpersonal comparisons of utility are problematic, for the most part he assumes that all individuals share the same utility function. Harsanyi has written extensively on the nature of interpersonal utility comparisons (see, for example, Harsanyi, 1955), and so heterogeneity of preference plays a prominent role in his writings. More complete statements of their theories may be found in Harsanyi (1977) and Vickrey (1960, 1961).

<sup>18</sup> In different alternatives, the same individuals (or even the same number of individuals) need not be alive. While it is not difficult to deal with this complexity, for simplicity, in the subsequent discussion I restrict attention to a fixed group of individuals.

experiment replaces an actual social alternative with an uncertain prospect in which the moral agent has an equal chance of occupying any position in society.

To evaluate one of these uncertain prospects, the moral agent must empathize with each individual in turn, considering the worth of the underlying social alternative from each person's point of view. When imagining being person  $i$  in social alternative  $x$ , the moral agent evaluates the social alternative from person  $i$ 's perspective, complete with  $i$ 's tastes and objective circumstances. Harsanyi and Vickrey assume that the moral agent is self-interested. Although self-interested, because the moral agent does not know his or her identity, he or she must promote the interests of all individuals, and must do so impartially because it is equally likely that the agent will be any particular person when the veil is lifted. Harsanyi and Vickrey also assume that the moral agent ranks uncertain prospects in accordance with the principles of expected utility theory. Because uncertain prospects assign equal probability to each of the outcomes (i.e., to which position the moral agent occupies when the veil is removed), Harsanyi and Vickrey argue that different prospects, and hence different social alternatives, will be ranked by the average of the utilities obtained by each of the individuals. In this way, or so they argue, the Harsanyi–Vickrey original position results in a form of average utilitarianism.<sup>19</sup>

Recall that I am interested in identifying a set of counterfactual choice situations that can be used to justify the social choice from the actual agenda given the actual preference profile. What I want to borrow from the Harsanyi–Vickrey construction is (a) the idea that personal identities are morally irrelevant and (b) the idea that this irrelevance can be operationalized by supposing that a moral agent is ignorant of his or her true identity. Adapting the first of these ideas to choice situations leads one to the view that who has which preference in the actual preference profile is morally irrelevant and that it is also morally irrelevant who occupies which position in the various alternatives that make up the actual agenda. Rather than introducing uncertain prospects to capture the ignorance imposed on a moral agent, I instead include in the domain of the social choice correspondence all of the choice situations one might think could be the actual choice situation if one were ignorant of one's true identity. This domain, excluding the actual choice situation, is the set of counterfactual choice situations that are used to justify the actual social choice when the relevant impartiality is with respect to personal identity.

In constructing one of these counterfactual choice situations, the identities of the individuals are permuted in a particular way. For example, the identities of individuals one and two could be permuted, with all other identities unchanged. In this example, the preferences of the first two individuals in the actual preference profile are interchanged. In each of the alternatives

<sup>19</sup> Whether Harsanyi's utilitarian conclusions follow from his assumptions is a matter of considerable controversy. See Weymark (1991, 2005).

that make up the actual agenda, the roles of these two individuals are also switched; person one takes on two's objective circumstances and two's tastes and values in each alternative and *vice versa* for person two. More generally, a counterfactual choice situation is obtained by applying the same permutation of individuals to the actual preference profile and to all alternatives in the actual agenda.

The domain constructed in this way is not very rich. In general, no two admissible choice situations share a common preference profile, nor do they share a common agenda. For example, if the circumstances of each individual are unique, as in my jam-leisure example, then any nontrivial permutation of an agenda differs from the original agenda. It would then not be possible to find two choice situations with a common agenda, and so Independence of Infeasible Alternatives would be vacuous.<sup>20</sup> Because all agendas have the same number of alternatives, it is not possible to shrink the set of alternatives holding the profile fixed. As a consequence, Arrow's Choice Axiom is necessarily vacuous on the domain generated using this thought experiment. So while the need to justify the actual social choice in an impartial way provides an argument for considering a domain that includes hypothetical choice situations, it does not justify the choice of a rich enough domain for Arrow's independence and collective rationality assumptions to have any force if the relevant impartiality only concerns personal identity.

## 7 A Rawlsian Veil

In Rawls (1971, 1993), a moral agent in his version of an original position is charged with designing principles of justice that are to apply to the basic institutions in society. Compared with a moral agent behind the Harsanyi-Vickrey veil of ignorance, a Rawlsian moral agent is endowed with much less information. As is the case with Harsanyi and Vickrey, Rawls believes that personal identities in the distribution of advantages are morally irrelevant. Further, each person's particular conception of the good (comprehensive doctrine for ordering one's life) is morally irrelevant in Rawls' theory—in part because conceptions of the good are not independent of the institutions that help shape people's lives and in part because in each society "... there is a diversity of comprehensive doctrines, all perfectly reasonable." (Rawls, 1993, p. 24) Instead, Rawls argues that what is morally relevant is that there are

<sup>20</sup> If there are two individuals who share the same preference over the feasible alternatives in the actual profile and if these same two individuals are in exactly the same circumstances in each of the alternatives that make up the actual agenda, then permuting these two individuals will result in a new choice situation that can only differ from the actual choice situation in the preferences over nonfeasible alternatives. It is only in such very special circumstances that the independence axiom has any role to play on this kind of domain.



certain basic goods—primary goods and liberties—that are known to facilitate the advancement of any conception of the good and that the more of these goods one has, the better one is able to fulfill one’s life plans. More particularly, a moral agent does not know the features of his or her psychology that determine the content of his or her interests or that determine his or her willingness to take risks to promote these interests. Rawls also argues that the actual point in history at which social decisions are being made is morally irrelevant, for otherwise particular generations could use this information to the detriment of subsequent generations. In addition, Rawls does not provide a moral agent with the information needed to determine the relative likelihood of the various social outcomes.<sup>21</sup>

Thus, in the Rawlsian theory, impartiality with respect to personal identity is extended to include impartiality with respect to particular conceptions of the good and impartiality with respect to generations. All individuals are to have equal status as moral beings, regardless of one’s natural advantages, one’s social circumstances, or one’s conception of the good.

As in Harsanyi’s and Vickrey’s description of an original position, Rawls supposes that a moral agent is self-interested (in the sense of wanting to advance his or her particular conception of the good whatever it should turn out to be when the veil is lifted). He argues that a moral agent with these interests situated behind his veil will choose principles of justice that (a) give priority to the equality of basic liberties, (b) arrange social and economic advantages so as to maximize the interests of the most disadvantaged group in society as measured by an index of primary goods (the *difference principle*), and (c) have positions and offices open to all individuals under conditions of fair equality of opportunity.<sup>22</sup> As is the case with the Harsanyi–Vickrey original position, the only relevant feature of Rawls’ theory for my argument is his description of the veil of ignorance (and not his principles of justice or the motivations he ascribes to a moral agent).

Because Rawls has a much thicker veil than do Harsanyi and Vickrey, there are many more possibilities for what the world looks like outside the veil. Hence, my procedure for determining the domain of a social choice correspondence generates a much richer set of counterfactual choice situations when the Rawlsian veil is used instead of the Harsanyi–Vickrey veil. To facilitate comparison with the domain generated by the Harsanyi–Vickrey veil, I

<sup>21</sup> This description of the information available behind the veil of ignorance is largely drawn from the account given in Rawls (1971, p. 137). While information relating to particular individuals outside a Rawlsian veil is restricted behind the veil as described above, Rawls provides a moral agent with general information about human societies, such as principles of individual behaviour and social interaction.

<sup>22</sup> The derivation of Rawls’ principles from his formulation of an original position is widely held to be problematic, as is the use of the difference principle to regulate social and economic affairs. See, for example, Barry (1973, 1989). The appropriateness of Rawls’ formulation of the veil is also the subject of some controversy. See, for example, Hare (1973) and Nagel (1973).

adapt the Rawlsian veil to deal with the problem of impartial resource allocation for a fixed population (over the history of the world), rather than the more general problem of determining principles of justice for regulating the structure of social institutions.

For concreteness, let me suppose that social alternatives are allocations of public and private goods. The actual choice situation is described by the preferences of individuals over allocations and by the set of allocations that are feasible given the endowment of resources and given the technologies available for producing goods and services. Information that is morally irrelevant for Harsanyi and Vickrey is also morally irrelevant for Rawls, so each of the choice situations obtained by permuting individuals in the actual choice situation is also a choice situation that must be included in the domain of the social choice correspondence. In other words, the domain generated by the Rawlsian veil must include the domain generated by the Harsanyi–Vickrey veil. Because the detailed conceptions of the good are morally irrelevant, but increases in primary goods are known to advance anyone's goals, it is reasonable to suppose that the domain of possible preference profiles for the counterfactual justification procedure includes (but may not be limited to) all profiles of preferences that are increasing in the personal consumption of all private and public goods. Because the state of development is morally irrelevant, any feasible set of allocations that can be obtained from some conceivable technologies using resources that do not exceed the earth's (or perhaps the universe's) initial endowment is an agenda that should be included in the domain.

Essentially the same domain was considered by Donaldson and Weymark (1988), but for different reasons. Donaldson and Weymark were interested in determining the consistency of the Arrovian axioms when natural economic restrictions are placed on the admissible preference profiles and on the admissible agendas. With minor and unimportant differences, the domain I have identified using a Rawlsian veil is the domain considered in their Theorem 1.<sup>23</sup> This domain is rich enough for both Arrow's Choice Axiom and Independence of Infeasible Alternatives to play substantive roles. By holding preferences fixed, but changing the technologies available to firms, it is possible to shrink the agenda in such a way that Arrow's Choice Axiom is nonvacuous. Because a counterfactual preference profile does not need to be a permutation of the actual profile, it is possible to change preferences (while holding the

<sup>23</sup> Formally, Donaldson and Weymark's universal set of alternatives is the nonnegative orthant in a Euclidean space. Each component of this space corresponds either to the consumption of a public good or to some individual's consumption of a private good. Any subset of the orthant that is compact, comprehensive (i.e., satisfies free disposal), and has a nonempty interior is a feasible agenda. An agenda can be interpreted as being the set of all feasible allocations for an economy, with different agendas resulting from different resource endowments and/or different production technologies. Various preference domains satisfy Donaldson and Weymark's assumptions, including domains in which standard economic restrictions on preferences are assumed.

agenda fixed) in such a way that Independence of Infeasible Alternatives is nonvacuous. Thus, the domain considered by Donaldson and Weymark can be interpreted as being the domain generated by a Rawlsian veil, and this domain is rich enough for the Arrovian independence and collective rationality conditions to restrict the kind of social choice correspondences that can be considered.

## 8 Ideal Observer Theories and Hare's Universal Prescriptivism

Ideal observer theories and Hare's universal prescriptivism share with the Harsanyi–Vickrey version of an original position theory the idea that the relevant impartiality is with respect to personal identity. However, these theories differ from the Harsanyi–Vickrey original position in the thought experiments used to achieve impartiality on the part of a moral agent. One could argue that it is the kind of impartiality that a moral theory exhibits, and not the particular way in which a theory embodies this impartiality, that determines the set of counterfactual choice situations that should be used to justify the choice in the actual choice situation. If this view is accepted, then ideal observer theories, universal prescriptivism, and the Harsanyi–Vickrey original position all identify the same set of counterfactuals. While essentially endorsing this view, in this section I argue that the device of a veil of ignorance is the most natural way of conceptualizing the choice of the appropriate counterfactuals.

Hare (1972, p. 168) characterizes an *ideal observer theory* as follows:

The ideal observer theory . . . holds that in considering what we ought to do, we have to conform our thought to what would be said by a person who had access to complete knowledge of all the facts, was absolutely clear in his thinking, was impartial between all the parties affected by the action, and yet equally benevolent to them all.

In an ideal observer theory, the moral agent is a perfect altruist, giving each person's interests equal consideration. There is no veil of ignorance in an ideal observer theory. On the contrary, an ideal observer knows the interests of all individuals and knows his or her own identity. However, the ideal observer's full knowledge of personal identities is not something that is taken advantage of for personal gain because of the observer's benevolence.

Ideal observer theories have been used to justify classical utilitarianism—the principle that holds that alternatives are to be ordered in terms of the sum of individuals' utilities. Classical utilitarianism follows from the requirement that the observer is equally benevolent to the interests of each individual. Because of this fact, Hare (1972, p. 169) has argued that original position theories of the sort advocated by Harsanyi and Vickrey are essentially equivalent to an ideal observer theory, at least for fixed population problems. So we see that utilitarian conclusions emerge from either the combination of self-interest with

ignorance about one's own identity or the combination of perfect benevolence with full knowledge of each person's identity. The Harsanyi–Vickrey original position and the ideal observer are simply two different devices for incorporating impartiality with respect to personal identity into a moral theory.

Hare's universal prescriptivism received its first statement in *The Language of Morals* (Hare, 1952). This theory has been further developed and refined in many of Hare's later writings, most notably in Hare (1963, 1981). Hare obtains substantive moral principles—in particular, a form of classical utilitarianism—by elucidating the meaning of moral words and the nature of moral reasoning.

Gorr (1985, p. 116) provides a nice summary of Hare's theory.

To sincerely affirm a moral judgment is, on Hare's view, to prescribe acting in accordance with a universal moral principle from which, in conjunction with statements specifying one's beliefs concerning the relevant facts, the judgment can be derived. To in turn determine whether one can prescribe acting in accordance with a universal principle is to determine whether one would actually choose to perform that action if one knew that one would have to play, in a series of possible worlds otherwise identical to the actual world, the role of *each* person (including oneself) who would be affected. Moreover, it is not enough that one simply imagines oneself, with one's *own* interests, in the place of those other persons—rather, one must imagine oneself as being in their place while having, in turn, *their* interests and desires.

Hare's moral agent is not ignorant of what role he or she occupies, as is the case behind a veil of ignorance; rather, this is information that is known, but it is morally irrelevant information, and so ignored. As a consequence, each person's interests are given equal weight. In order to promote each person's interests, it is necessary to know what these interests are. The moral agent obtains this knowledge by empathetic identification, imagining what it is like to be each person in turn.<sup>24</sup>

In Hare's theory, impartiality is achieved by the requirement that moral prescriptions be universalizable to all situations, both actual and hypothetical, that share with the situation under consideration the same universal features. The interests of all are advanced because in promoting one's own interests, one is constrained by the principle of universality to promote everyone else's interests as well. Because the thought experiment that underlies universal prescriptivism embodies the impartiality, equal benevolence, and knowledge of an ideal observer theory, Hare (1972, p. 171) argues that these two theories are for all practical purposes the same in their consequences. The idea of empathetically identifying with each individual and then giving each person's interests equal weight when aggregating is conceptually similar to a Harsanyi–Vickrey lottery, the main difference being that utilities are added in Hare's theory, but are averaged in the theories of Harsanyi and Vickrey. For a fixed

<sup>24</sup> See Hare (1981, Chapter 5).

population, classical and average utilitarianism order alternatives in the same way.

After commenting that the devices used in original position and impartial observer theories to ensure impartiality are less important than the purpose for which they are adopted, Kymlicka (1991, p. 194) goes on to observe that:

Impartial consideration can also be generated without any special devices at all, just by asking agents to give equal consideration to others notwithstanding their knowledge of, and ability to promote, their own good.<sup>25</sup>

The view expressed by Kymlicka can be adapted to deal with Arrovian social choice when the purpose is to make impartial decisions. That is to say, the kind of impartiality required by one's moral theory can be used directly (a) to identify the appropriate set of counterfactual choice situations and (b) to place restrictions on how the choices made in the admissible choice situations so identified are related to each other. This is done without the need for special devices, such as a veil of ignorance, which force a moral agent to adopt an impartial perspective. According to this view, because the Harsanyi–Vickrey original position, an ideal observer theory, and Hare's universal prescriptivism all employ the same notion of impartiality, they should all identify the same counterfactual choice situations. On the other hand, Rawls' original position should identify a different set of counterfactuals because the impartiality demanded by his theory is more all-encompassing.

In my procedure for generating the domain for an Arrovian social choice correspondence, the domain consists of all the choice situations that one might think is the actual choice situation given the information available behind a veil of ignorance. With the Harsanyi–Vickrey veil, this construction identifies the hypothetical choice situations that differ from the actual choice situation only in the names of individuals. This is the set of choice situations that needs to be considered to justify the actual social choice set when the basis for the justification is impartiality with respect to personal identity.

Alternatively, rather than using the device of a veil of ignorance to generate this set of counterfactual choice situations, one could instead simply say that this kind of impartiality requires that the choice(s) in any hypothetical choice situation that is obtained from the actual choice situation by permuting the identities of the individuals must be the permuted alternative(s) corresponding to the actual choice(s). In this argument, the very concept of "impartiality with respect to personal identity" identifies the relevant set of counterfactual choice situations, independent of how this concept is embodied in a moral theory. As in Hare's theory, we are guided by the meaning of moral words, in this case, the meaning of "impartiality with respect to personal identity." Similarly, other notions of impartiality, such as impartiality with respect to

<sup>25</sup> Ackerman (1994, p. 369) makes a similar point in arguing against the use of a Rawlsian veil.

generations or with respect to conceptions of the good, can be used to identify different sets of counterfactual choice situations without appealing to devices like a veil of ignorance.

While it is possible to use this device-independent approach in order to defend the choice of domain for the purposes of applying Kolm's epistemic counterfactual principle, because of the nature of the counterfactuals considered in the social choice problem, it is nevertheless the case that a veil of ignorance is the most natural way to conceptualize the choice of the relevant set of counterfactual choice situations. Behind a veil, it is natural to think in terms of alternative choice situations—all the choice situations compatible with the information allowed by the veil. In contrast, with an ideal observer theory or with Hare's universal prescriptivism, the moral agent *knows* what the actual choice situation is. As a consequence, neither an ideal observer theory nor the idea of a moral agent arriving at universal prescriptions by a process of empathetic identification lead naturally to the consideration of alternative hypothetical *choice situations*. While an impartial observer or one of Hare's moral agents imagines him- or herself in the hypothetical circumstance of being someone else, this is not a hypothetical choice situation; i.e., it is not a combination of a preference profile and an agenda.

Hence, while the Harsanyi–Vickrey original position, an ideal observer theory, and Hare's universal prescriptivism utilize the same notion of impartiality and could in principle be used to identify the same set of counterfactuals, it is the veil argument that most simply and naturally leads one to think in terms of counterfactual choice situations. Similarly, incorporating impartiality into a moral theory by directly placing constraints on the kinds of arguments that can be advanced does not naturally lead one to consider counterfactual choice situations, as is the case with the device of a veil of ignorance.

## 9 Concluding Remarks

According to Kolm's epistemic counterfactual principle, the actual social choice must be justified by the choices that would have been made in appropriate counterfactual choice situations. Together with the actual choice situation, these counterfactual choice situations constitute the domain for an Arrovian social choice correspondence. I have argued that Kolm's principle needs to be supplemented by an account of how the relevant set of counterfactual choice situations is to be determined. I have suggested that an investigation of the constraints imposed on choice by the need to be impartial provides such an account when the objective is to make collective decisions impartially. I have further suggested that the most natural way to capture the requirements of impartiality in this context is to use a veil of ignorance. A veil restricts information to that which is morally relevant, thereby forcing a moral agent to take an impartial perspective. In my argument, the relevant Arrovian domain consists of all the choice situations that a moral agent

behind a veil of ignorance thinks might be the actual choice situation. The use of a veil is compatible with quite different notions of impartiality, and so with quite different moral theories.<sup>26</sup> Here, I have used suitably reformulated versions of the veils proposed by Harsanyi, Rawls, and Vickrey to illustrate my procedure for generating counterfactual choice situations.

The use of original position arguments is not without controversy. Barry (1989, Section 41), Kolm (1996, Section 3), and Kolm (1997, Chapter 8) have argued that no original position theory—not just the original position theories of Harsanyi, Rawls, and Vickrey—can provide a satisfactory basis for determining principles of justice because the use of an original position necessarily results in the choice of principles that permit the sacrifice of some people’s interests for the interests of others without at the same time providing a compelling reason for anyone to accept such sacrifices as being morally justified. Even if one finds their arguments compelling, one can consistently use original position arguments as I have done here. My argument only makes use of one part of the description of an original position—the veil of ignorance. It is the veil, not the assumption that a moral agent is self-interested, that is needed to identify my set of counterfactual choice situations. Once these counterfactual choice situations have been identified, the veil can be lifted before making the choices. I escape the criticisms of original position arguments advanced by Barry and Kolm because I do not require that the choices proposed by a moral agent are the ones that would have been chosen by pursuing *self-interest* behind a veil of ignorance.

The particular procedure that I have proposed here to construct a domain for an Arrovian social choice correspondence provides one possible way of operationalizing Kolm’s principle. The appropriateness of this procedure is, however, dependent on my assumption that the only social objective is to make choices impartially. With different objectives, different counterfactual choice situations are needed to apply Kolm’s epistemic counterfactual principle.

## Acknowledgements

I have benefitted from comments received from Ken Binmore, Andrew Irvine, Serge Kolm, Philippe Mongin, and the participants in the Conference on Social Ethics and Normative Economics held in honour of Serge Kolm at the Université de Caen, May 18–19, 2007. I am particularly grateful to Peter Valentyne for his extensive written comments. This article develops some ideas that were only briefly sketched in Weymark (1996). Research support from the Social Sciences and Humanities Research Council of Canada, the Hampton Fund of the University of British Columbia, and the College of Arts and Science at Vanderbilt University is gratefully acknowledged.

<sup>26</sup> Rawls (1971, p. 121) conjectures that for all of the major theories of justice, it is possible to specify an original position that would lead to the choice of its principles.

## References

- Ackerman, B. (1994). Political liberalisms. *Journal of Philosophy*, **91**, 364–386.
- Arrow, K. J. (1951). *Social Choice and Individual Values*. John Wiley & Sons, New York.
- Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, **26**, 121–127.
- Arrow, K. J. (1963). *Social Choice and Individual Values*. John Wiley & Sons, New York, second edition.
- Arrow, K. J. (1997). The functions of social choice theory. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Social Choice Re-Examined, Volume 1*, pages 3–9. Macmillan Press, London and Basingstoke.
- Barry, B. (1973). *The Liberal Theory of Justice*. Clarendon Press, Oxford.
- Barry, B. (1989). *A Treatise on Social Justice. Volume 1: Theories of Justice*. University of California Press, Berkeley and Los Angeles.
- Bergson, A. (1938). A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics*, **52**, 310–334.
- Bergson, A. (1954). On the concept of social welfare. *Quarterly Journal of Economics*, **63**, 233–252.
- Donaldson, D. and Weymark, J. A. (1988). Social choice in economic environments. *Journal of Economic Theory*, **46**, 291–308.
- Fleurbaey, M. and Mongin, P. (2005). The news of the death of welfare economics is greatly exaggerated. *Social Choice and Welfare*, **25**, 381–418.
- Gorr, M. (1985). Reason, impartiality and utilitarianism. In N. T. Potter and M. Timmons, editors, *Morality and Universality*, pages 115–138. D. Reidel, Dordrecht.
- Hare, R. M. (1952). *The Language of Morals*. Clarendon Press, Oxford.
- Hare, R. M. (1963). *Freedom and Reason*. Clarendon Press, Oxford.
- Hare, R. M. (1972). Rules of war and moral reasoning. *Philosophy and Public Affairs*, **1**, 166–181.
- Hare, R. M. (1973). Rawls' theory of justice—I and II. *Philosophical Quarterly*, **23**, 144–155 and 241–251.
- Hare, R. M. (1981). *Moral Thinking*. Clarendon Press, Oxford.
- Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, **61**, 434–435.
- Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, **63**, 309–321.
- Harsanyi, J. C. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press, Cambridge.
- Kemp, M. C. and Ng, Y.-K. (1976). On the existence of social welfare functions, social orderings and social decision functions. *Economica*, **43**, 59–66.
- Kolm, S.-C. (1969). The optimal production of social justice. In J. Margolis and H. Guitton, editors, *Public Economics: An Analysis of Public Production and Consumption and Their Relations to the Private Sector*, pages 145–200. Macmillan Press, London and Basingstoke.
- Kolm, S.-C. (1993). What sense social choice? unpublished manuscript, Conseil Général des Ponts et Chaussées, Paris.
- Kolm, S.-C. (1994). Rational normative economics vs 'social welfare' and 'social choice'. *European Economic Review*, **38**, 721–730.



- Kolm, S.-C. (1995). Meanings and rationality in social choice theory. In D. Andler, P. Banerjee, M. Chaudhury, and O. Guillaume, editors, *Facets of Rationality*, pages 79–103. Sage Publications, New Dehli.
- Kolm, S.-C. (1996). Rational just social choice. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Social Choice Re-Examined, Volume 2*, pages 167–196. Macmillan Press, London and Basingstoke.
- Kolm, S.-C. (1997). *Modern Theories of Justice*. MIT Press, Cambridge, MA.
- Kymlicka, W. (1991). The social contract tradition. In P. Singer, editor, *A Companion to Ethics*, pages 186–196. Basil Blackwell, Oxford.
- Le Breton, M. and Weymark, J. A. (2005). Arrovian social choice theory on economic domains. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare, Volume 2*. North-Holland, Amsterdam. forthcoming.
- Little, I. M. D. (1952). Social choice and individual values. *Journal of Political Economy*, **60**, 422–432.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, **38**, 175–208.
- Mongin, P. and d’Aspremont, C. (1998). Utility theory and ethics. In S. Barberà, P. J. Hammond, and C. Seidl, editors, *Handbook of Utility Theory. Volume 1: Principles*, pages 371–481. Kluwer Academic Publishers, Boston.
- Nagel, T. (1973). Rawls on justice. *Philosophical Review*, **82**, 220–234.
- Parks, R. P. (1976). An impossibility theorem for fixed preferences: A dictatorial Bergson-Samuelson welfare function. *Review of Economic Studies*, **43**, 447–450.
- Pattanaik, P. K. (2005). Little and Bergson on Arrow’s concept of social welfare. *Social Choice and Welfare*, **25**, 369–379.
- Pollak, R. A. (1979). Bergson–Samuelson social welfare functions and the theory of social choice. *Quarterly Journal of Economics*, **93**, 73–90.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge, MA.
- Rawls, J. (1993). *Political Liberalism*. Columbia University Press, New York.
- Samuelson, P. A. (1947). *Foundations of Economic Analysis*. Harvard University Press, Cambridge, MA.
- Samuelson, P. A. (1967). Arrow’s mathematical politics. In S. Hook, editor, *Human Values and Economic Policy*, pages 41–51. New York University Press, New York.
- Samuelson, P. A. (1977). Reaffirming the existence of “reasonable” Bergson-Samuelson social welfare functions. *Economica*, **44**, 81–88.
- Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, **17**, 213–240.
- Vickrey, W. (1945). Measuring marginal utility by reactions to risk. *Econometrica*, **13**, 319–333.
- Vickrey, W. (1960). Utility, strategy, and social decision rules. *Quarterly Journal of Economics*, **74**, 507–535.
- Vickrey, W. (1961). Risk, utility, and social policy. *Social Research*, **28**, 205–217.
- Weymark, J. A. (1991). A reconsideration of the Harsanyi–Sen debate on utilitarianism. In J. Elster and J. E. Roemer, editors, *Interpersonal Comparisons of Well-Being*, pages 255–320. Cambridge University Press, Cambridge.
- Weymark, J. A. (1996). Discussion of Kolm’s paper. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Social Choice Re-Examined, Volume 2*, pages 197–201. Macmillan Press, London and Basingstoke.

Weymark, J. A. (2005). Measurement theory and the foundations of utilitarianism. *Social Choice and Welfare*, **25**, 527–555.