Impartiality and priority.
Part 1: the veil of ignorance*

Juan D. Moreno-Ternero
Yale University

John E. Roemer†
Yale University

February 9, 2005

Abstract

The veil of ignorance has been used often as a tool for recommending what justice requires with respect to the distribution of wealth. We complete Harsanyi's model of the veil of ignorance by appending information permitting interpersonal comparability of welfare. We show that the veil-of-ignorance conception of John Harsanyi, so completed, and Ronald Dworkin's, when modeled formally, recommend wealth allocations in conflict with the prominently espoused view that priority should be given to the worse off with respect to wealth allocation. (*JEL numbers: D63, D71*).

Keywords: Impartiality, Priority, Veil of ignorance.

*We thank G.A. Cohen, Marc Fleurbaey, Shelly Kagan, Edi Karni, Marco Mariotti, Klaus Nehring, Ariel Rubinstein, and Larry Temkin for helpful contributions to this paper. We also acknowledge the participants' contributions at seminars and conferences where the paper has been presented, at Rutgers University, the University of Ghent, the Ecole Polytechnique, Hitotsubashi University, Osaka University and Yale University. Moreno-Ternero's research was made possible thanks to a post-doctoral grant for advancement of research staff from the Generalitat Valenciana.

†**Corresponding Author**: John E. Roemer. Department of Political Science. Yale University. PO Box 208301. New Haven CT 06520-8301. e-mail: john.roemer@yale.edu

1 Introduction

The construct of the veil of ignorance has been of significant import in political philosophy during the last half century: three prominent writers–John Harsanyi, John Rawls, and Ronald Dworkin–have employed it in different forms. Although these three disagree on exactly how thick the veil should be, each uses it as a tool to enforce impartiality in the procedure that deduces what the worldly distribution of resources or wealth should be. The veil-of-ignorance model is putatively impartial because the 'soul' or 'souls' or 'parties' or 'observer' who contemplate(s) behind the veil are (is) deprived of precisely that information that the author deems to be morally arbitrary, to use the phrase of Rawlsian parlance.

From quite a different vantage point, another group of political philosophers (which has a non-empty intersection with the first group) has been concerned to argue that justice requires that priority be given to the worse off in the allocation of wealth. The most extreme form of priority is advocated by Rawls, for whom differences in amounts of primary goods (wealth among them) accruing to people are only morally permissible if they maximize the level of (or index of) primary goods accruing to the worst off (that is, she who is least endowed with primary goods). Rawls (1971) attempts, unsuccessfully in our view, to argue for this principle using a veil-of-ignorance (original position) construction.[1]

The difference principle has often been criticized as being too extreme, and Derek Parfit (1997) has coined the term *prioritarianism* for the view that the 'worse off' should be given priority over the 'better off' with respect to resource allocation, but that the former need not necessarily receive the extreme priority that characterizes maximin (the difference principle). In a welfarist setting, prioritarianism is usually characterized as a social welfare function with strictly convex upper contour sets. The boundaries of prioritarianism are maximin on one side, and utilitarianism on the other. (See, for

[1] See Roemer (1996) for one discussion of the inadequacy of Rawls's argument from the original position.

example, Roemer, 2004).

Other philosophers who would identify themselves with either a *prioritarian* or *egalitarian* or *maximin* view include Brian Barry (1989, 1995), G.A. Cohen (1992), Larry Temkin (1993), and Thomas Scanlon (1998). There are surely many more. We include together the three views here italicized because prioritarianism is a weakening of egalitarianism and the difference principle: if a rule is egalitarian or maximin it is surely prioritarian.[2] Those who advocate priority but not maximin do so usually because they consider the sacrifices of implementing the difference principle too great–sacrifices borne by the better off.

In this paper, we show that the veil of ignorance, when formulated in a rigorous way, is inconsistent with prioritarianism: to be precise, it will often recommend distributions of wealth that give priority to the better off.[3] We focus on the conceptions of the veil of ignorance outlined by Harsanyi (1953, 1977) and Dworkin (1981b). In the former case, we complete the theory that Harsanyi began and then show its anti-prioritarian consequences. In the latter case, we show the anti-prioritarian nature of a simple model reflecting Dworkin's insurance mechanism.

The concept of one person's being worse off than another is one, in the philosophical literature, which presupposes the possibility of making interpersonal welfare comparisons. Therefore, such comparisons will be assumed to be possible in our formulation of the problem. To anticipate what we define in what follows, we will say that one person is worse off, or less able, than another if the first transforms resources (for us, 'wealth') into welfare less efficiently than the other.

[2] One could argue that egalitarianism does not imply priority, in the sense that $(2,2)$ is more egalitarian than $(3,4)$, but the worse off person is better off in the second allocation than in the first. Thus priority could recommend $(3,4)$ but equality $(2,2)$. One could, however, also argue that in $(2,2)$ the first person is given greater priority than in $(3,4)$. We pursue this no further.

[3] An early form of this work is available in Roemer (2002); that article has an error, which is corrected here.

There is a recent literature which also is concerned with completing Harsanyi's veil-of-ignorance approach (Karni, 1998, 2003; Dhillon and Mertens, 1999; and Segal, 2000).[4] In the conclusion, we briefly contrast our approach to this literature.

2 The Harsanyi veil of ignorance

2.1 Harsanyi's original first step

In 1953, John Harsanyi proposed the first precise model of the veil of ignorance.[5] Suppose there are n individuals, each of whom possesses von Neumann-Morgenstern (vNM) preferences over wealth lotteries. Denote the vNM utility functions on wealth for these people by $v^1, v^2, ..., v^n$. There is an amount of wealth $\overline{W}$ to be divided among them. What is the just division? Harsanyi proposes to conceptualize a single impartial observer (IO) who will become one of these people, with equal probability of becoming each one. How would such an observer allocate the wealth?

The IO's data, for Harsanyi, consist of the set $\{v^1, v^2, ..., v^n, \overline{W}\}$.

Denote by (i, W) the *extended prospect* that means 'becoming person i with wealth W.' Harsanyi proposes that the IO, to solve his problem, must itself possess a vNM utility function U defined on extended prospects. (That is, it must be able to evaluate lotteries on extended prospects.) We can then represent the 'birth lottery' through which the IO becomes a particular person, and in which the distribution of wealth among the individuals is $(W^1, W^2, ..., W^n)$, by

$$l = \left(\frac{1}{n} \circ (1, W^1), \frac{1}{n} \circ (2, W^2), ..., \frac{1}{n} \circ (n, W^n)\right).$$

This is to be read, "With probability $1/n$, the extended prospect $(1, W^1)$ is realized (and the IO becomes person 1 with wealth W^1), with probability

[4] We thank a referee for alerting us to the connections between our work and this literature.

[5] See Harsanyi (1977) or Weymark (1989) for a more detailed presentation of this model.

$1/n$ the extended prospect $(2, W^2)$ is realized, and so on". Now the utility the IO receives from this lottery is, by the expected utility property, equal to:

$$\sum_{i=1}^{n} \frac{1}{n} \cdot U(i, W^i) \tag{1}$$

and so the IO needs only find the distribution of wealth that maximizes expression (1) subject to the constraint that $\sum W^i = \overline{W}$. That distribution is the one it would choose, and therefore, that justice recommends.

The problem, then, is to deduce what the function U is. Harsanyi takes an axiomatic approach to this problem. He assumes what he calls:

The Principle of Acceptance*: For each fixed $i \in \{1, ..., n\}$, the function $U(i, \cdot)$ represents the same vNM preferences on* wealth *lotteries as $v^i(\cdot)$ represents.*

Now the vNM theorem tells us that any two vNM utility functions that represent the same preferences must be positive affine transformations of each other. Therefore:

For all W and i, there exist $a^i > 0$ and b^i such that

$$U(i, W) = a^i \cdot v^i(W) + b^i \tag{2}$$

Substituting formulae (2) into (1), we have that

$$\sum_{i=1}^{n} \frac{1}{n} \cdot U(i, W^i) = \sum_{i=1}^{n} \frac{1}{n} \left(a^i \cdot v^i(W^i) + b^i\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} a^i \cdot v^i(W^i) + \frac{1}{n} \cdot \sum_{i=1}^{n} b^i \tag{3}$$

Maximizing the right-hand side of (3) is equivalent to maximizing $\sum a^i \cdot v^i(W^i)$. That is the end of Harsanyi's argument: the IO must maximize *some* positive weighted sum of the vNM utilities of the individual persons.

2.2 Amending Harsanyi's first step to allow for interpersonal comparisons of welfare

Harsanyi's argument is in our view unfinished, for he has provided no well-argued way of determining the values of the positive numbers $\{a^i : i =$

$1, 2, ...n\}$, so he has not determined the vNM preferences of the IO.[6] Furthermore, there is *no way* to derive these values from the information that Harsanyi has provided to the IO, and his axioms.

A moment's thought will show why this is so. The only information the IO has, consists in the profile of risk preferences of the individuals, and the total wealth to be allocated. But to decide whether it would rather become Alan with \$1000 or Barbara with \$3000, the IO *must be able to compare how well off Alan is with* \$1000 *with how well off Barbara is with* \$3000. (Or it must have some independent reason to prefer to be Alan, say.) There is no way the IO can make such comparisons with the information Harsanyi has given it. There is, in Harsanyi's specification of the problem, *absolutely no information* permitting interpersonal welfare comparisons. The vNM preferences of the individuals are purely ordinal preferences that measure 'utility' in a non-comparable way across persons.[7]

Clearly, *if* the IO were to possess vNM preferences on the lotteries on the space of extended prospects, such preferences would *imply* the existence of something that *looks like* an interpersonal welfare ordering, for the IO. For let such preferences exist and denote them by $\succsim$; then the statement "$(i, W) \succsim (j, W')$" which means "the IO would weakly prefer to be person i with wealth W to being person j with wealth W'" is *similar*, though surely not identical, to the statement "person i with wealth W is at least as well off as person j with wealth W'." Of course, the second statement presupposes that interpersonal comparisons of welfare are meaningful, while the

[6] Indeed, Harsanyi asserts that the weights of individual vNM utilities must be equal to reflect impartiality. We reject this view: if a preference relation is represented by a weighted sum of individual utilities, then there is a feasible manipulation of the individual utility functions that leads to a new representation of the preference relation as a weighted sum of individual utilities with equal weights.

[7] Many people are confused about this claim. VNM preferences are *ordinal* preferences on lotteries. There happens to be a very useful *cardinal* representation of those preferences, which allows us to calculate the utility of a lottery in a very simple way (by factoring out the probabilities). But the *preferences* are purely ordinal and non-comparable across persons. For further discussion, see Roemer (1996, chapter 4).

first statement does not: in the absence of the ability to make such comparisons the IO might have another decision procedure which leads it to its view about the relative merits of being person i or j with the associated wealths.

Our proposal is simple: we will amend Harsanyi's model by assuming that it *is* possible to perform interpersonal welfare comparisons. To this end, we assume that

There is a complete ***order on extended prospects****, denoted R, and for all i, j, W there exists W' such that $(i, W)I(j, W')$, where I denotes the symmetric part of the order R.*

The statement $(1, W^1)R(2, W^2)$ means 'person 1 with wealth W^1 is *at least as well off* as person 2 with wealth W^2'. This order is to be thought of as a fact about the world, a statement about how the persons experience life: it is not –*yet*– the subjective preference order of the IO.

We now append an axiom concerning the relationship between the IO's ordering of extended prospects and the interpersonal ordering R, that we name:

The Principle of Neutrality: $(i, W^i) \succsim (j, W^j) \Leftrightarrow (i, W^i)R(j, W^j)$.

In other words, the IO weakly prefers one extended prospect to another if and only if the person in the first extended prospect is at least as well-off (in terms of the interpersonally comparable coin called welfare or well-being) as the person in the second extended prospect. We call this 'neutrality' because it asserts that the IO brings no external considerations to bear concerning what person it would like to become: it only follows the dictates of the interpersonally comparable attribute called welfare or well-being, ignoring all other traits these individuals have (such as their sex, race, nationality, religious preference, or political views).[8]

[8] It is worth noting that our proposal, similar that might appear, does not follow the so-called *extended preference approach,* formally elaborated by Arrow (1963, 1977), Suppes (1966), Sen (1970), Kolm (1972), Suzumura (1983) and Mongin (2001) among others. This approach endows each member of the society with both an actual preference relation defined on the set of allocations (social states), and a relation defined on extended prospects.

We view the principle of neutrality as a principle of *impartiality* as well, for the reason stated above. Neutrality, that is, is a kind of impartiality. Of course, because information on the sex, race, religion, etc., of individuals is unavailable in our specification of the data of the problem, it is hard to see what *partiality* could mean on these environments –that is, given only the above data, the IO must necessarily ignore these other attributes of individuals in the world. Nevertheless, partiality could mean that the IO would prefer to be realized as one person i over another person j because he likes to have the risk preferences of i more than having the risk preferences of j. The principle of neutrality excludes *that* kind of partiality: for it requires the IO to choose between extended prospects *only* on the basis of the *welfare* he would experience in them.

2.2.1 An impossibility result

The data available to the IO are now $\{v^1, v^2, ..., v^n, \overline{W}, R\}$, and the *axioms* that relate his own preference order to the data are those of Acceptance and Neutrality. We will now show that these data and axioms together lead to an impossibility theorem.

To do so, we first introduce another concept. Let $\{W_a^1, W_a^2, ..., W_a^n\}$ be an *equal-welfare distribution of wealth*: that is a distribution such that

$$(i, W_a^i)I(j, W_a^j) \text{ for every pair } i, j,$$

where I is the symmetric part of the order R. Let there be two more equal-welfare distributions of wealth denoted $\{W_b^i\}_{i=1}^n$ and $\{W_c^i\}_{i=1}^n$, and suppose that these three distributions of wealth represent three welfare levels in increasing order of welfare, and so it follows that for each i, $W_a^i < W_b^i < W_c^i$, because we assume that welfare is strictly increasing in wealth.

We again invoke the vNM theorem, which tells us that for each person i

In our approach, individuals posses preferences over wealth lotteries but not over extended prospects. The IO is the one possessing preferences over extended prospects.

there is a unique probability p^i such that:

$$v^i(W_b^i) = p^i \cdot v^i(W_a^i) + (1 - p^i) \cdot v^i(W_c^i) \tag{4}$$

In general, of course, the probabilities p^i will differ across individuals. The more risk averse an individual is, the lower will p^i be. We say that:

The individuals in the world $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ are ***risk isomorphic*** *if, for any choice of the three equal-welfare distributions $\{W_a^i\}, \{W_b^i\}$ and $\{W_c^i\}$, the numbers $\{p^i : i = 1, \ldots, n\}$ defined by (4) are identical for all $i = 1, ..., n$.*

What this means is that these individuals have identical risk preferences over *welfare lotteries*, where welfare is the coin of interpersonal comparability.

Risk isomorphism is a property of our environments: it requires as data both the profile of vNM preferences and the interpersonal ordering R. Clearly, it is a singular case, which will rarely if ever hold in 'real worlds.'

We have the following:

Theorem 1 *There is a vNM preference order on extended prospects (for the IO) that satisfies the principles of acceptance and neutrality if and only if individuals in the world $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ are risk isomorphic. If so, the order is unique and represented by the vNM utility function on extended prospects:*

$$U(i, W) = \frac{v^i(W) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)}, \tag{5}$$

where $\{W_a^i\}_{i=1}^n$ and $\{W_b^i\}_{i=1}^n$ are any two equal-welfare distributions of wealth such that $W_b^i > W_a^i$.[9]

Theorem 1 is basically an impossibility theorem. It says that, in what is the usual case (of risk non-isomorphism), the Harsanyi veil of ignorance,

[9] This is a correction of the stated theorem in Roemer (2002). I (Roemer) there incorrectly assumed something that implied that all environments were risk isomorphic, and so I claimed that the principles of neutrality and acceptance always characterized unique vNM preferences for the IO. Fortunately, the examples of that paper are correct, as they are all examples where risk isomorphism holds.

amended by information on interpersonal comparability and the principle of neutrality, is an incoherent thought experiment. In the singular case of risk-isomorphism, we uniquely determine the preferences of the IO (that is, we solve for the coefficients $\{a^i\}$ of equation (3).) In particular, if the environment is risk isomorphic, we know the *basic procedure* by which the IO selects the allocation of wealth. Formally,

Basic procedure: *Let $\{W_a^i\}_{i=1}^n$ and $\{W_b^i\}_{i=1}^n$ be any two equal-welfare distributions of wealth such that $W_b^i > W_a^i$. If individuals in the world $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ are risk isomorphic, then the IO selects an allocation $\omega^B = (W^1, W^2, ..., W^n)$ that maximizes*

$$\sum_{i=1}^{n} \frac{1}{n} \cdot \frac{v^i(W^i) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)}, \tag{6}$$

subject to the condition that $\sum_{i=1}^n W^i = \overline{W}$.

2.2.2 Deep imaginative empathy

Theorem 1 shows that in the risk non-isomorphism case, there are no vNM preferences for the IO. However, we can propose a procedure by which the IO might decide upon a distribution of wealth in any (risk-isomorphic or not) environment. This procedure uses an idea that Harsanyi (1977, page 51) refers to as 'imaginative empathy'.

Denote the individuals by $1, 2, \ldots, n$. The IO first 'steps in the shoes' of any person i, and chooses the wealth distribution i would choose, if i always converts wealth given to other people into the welfare-equivalent wealth for herself (i). We define this precisely as follows. For any pair (j, W) and any agent i define $\sigma_j^i(W)$ by $(j, W^j)I(i, \sigma_j^i(W))$. That is, $\sigma_j^i(W)$ is the wealth that i requires to reach the same level of welfare as j achieves with wealth W^j. If the distribution of wealth being contemplated is $(W^1, W^2, ..., W^n)$ then the IO, placing herself in i's shoes, would evaluate the birth lottery as having expected utility

$$\sum_{j=1}^{n} \frac{1}{n} \cdot v^i(\sigma_j^i(W^j)). \tag{7}$$

Thus the IO, using i's risk preferences, asks *how she would feel* as any person j, given the wealth j gets in the distribution: to do so, the IO must convert j's wealth to the welfare–equivalent wealth for i, since the IO is evaluating everything from i's perspective.

Harsanyi used the phrase *imaginative empathy* for the compassion the IO feels as it contemplates taking on the risk preferences of different people. But, since Harsanyi did not deal with interpersonal comparisons of welfare, his imaginative empathy referred to the IO's taking on only the risk preferences of different people, as modeled by his Principle of Acceptance. The formulation (7) is one we call *deep imaginative empathy*, because, when stepping in the shoes of person i, the IO takes on not only i's risk preferences, but imagines how i would feel *in terms of welfare* if he (i) were to be realized as any person j with a given wealth level W^j. Person i would experience j's wealth level W^j, which is equivalent to i's having the wealth level $\sigma_j^i(W^j)$.

Denote by $\omega_i = (W_i^1, ..., W_i^n)$ a feasible distribution of wealth that maximizes expression (7) subject to the condition that $\sum_{k=1}^n W_i^k = \overline{W}$. Sequentially, the IO now performs this computation, taking on every person's viewpoint. This produces n wealth distributions $\omega_1, ..., \omega_n$. We propose that the IO takes the average of these distributions, $\frac{1}{n} \cdot \sum \omega_i$, as its recommended distribution.[10] Formally:

General procedure: *In the world* $\{v^1, v^2, ..., v^n, \overline{W}, R\}$*, the IO selects an allocation*

$$\omega^G = \frac{1}{n} \cdot \sum \omega_i, \tag{8}$$

where, for all $i = 1, \ldots, n$*,* $\omega_i = (W_i^1, ..., W_i^n)$ *is an allocation that maximizes expression (7) subject to the condition that* $\sum_{k=1}^n W_i^k = \overline{W}$*.*

Note that the general procedure just described provides a choice correspondence for the IO: to any environment $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ it associates a (not necessarily singleton) set of distributions of wealth. So does the basic procedure described in (6), when the environment is risk isomorphic. The

[10] Indeed, as we will see in the proof of Theorem 2, the IO can take any convex combination of these wealth distributions.

general procedure, however, can be performed for any environment, risk isomorphic or not. The next result shows that it is a *generalization* of the basic procedure: that is, if the environment is risk-isomorphic, the general procedure selects the allocations obtained maximizing the IO's vNM utility function provided in Theorem 1.

Theorem 2 *If individuals in the world* $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ *are risk isomorphic, then the following statements hold:*

(i) If the basic procedure yields a unique allocation of wealth then the general procedure yields the same allocation.

(ii) If the basic procedure yields more than one allocation of wealth, and all the v^i *functions are concave, then the general procedure yields the same set of allocations.*

2.3 The anti-prioritarian nature of the Harsanyi veil of ignorance

We now show the anti-prioritarian nature of the Harsanyi veil of ignorance by means of a simple example. There are two individuals, Andrea and Bob. They are each risk neutral. We may therefore take them to have the same linear vNM utility function, namely

$$v^A(W) = v^B(W) = W.$$

Let us suppose that the interpersonal welfare order is given by $(Andrea, W)I(Bob, 2W)$; that is, Bob always needs twice the wealth of Andrea to achieve the same welfare level as she. It is easy to see that this environment is risk isomorphic.

We now compute what the IO recommends for this example. Suppose that $\overline{W} = 1$, so a distribution of wealth is represented by $(W, 1 - W)$ where the first component goes to Andrea and the second to Bob. The IO must choose W. We know that $U(A, W) = U(B, 2W)$ by the principle of neutrality. Now the IO must choose W to

$$\text{maximize } \frac{1}{2}U(A, W) + \frac{1}{2}U(B, 1 - W).$$

By the formula just given we can write this as

$$\max \frac{1}{2}U(A, W) + \frac{1}{2}U(A, \frac{1-W}{2})$$

But by the principle of acceptance, this is equivalent to maximizing

$$\frac{1}{2}W + \frac{1}{2} \cdot \frac{1-W}{2} = \frac{1}{4} + \frac{W}{4} \tag{9}$$

which is achieved at $W = 1$: the IO would give all the wealth to Andrea.

Now in this environment, we consider Bob to be *disabled* with respect to Andrea: he requires more wealth than she to receive any given level of welfare. Thus, the Harsanyi veil of ignorance gives all the wealth to the able person. If we consider more than two individuals in this example, and they can be ordered with respect to 'ability', then the Harsanyi veil of ignorance assigns all the wealth to the most able individual(s).

What happens if we alter the risk preferences in the above Andrea-Bob example so that the individuals are risk averse? For small degrees of risk aversion, it continues to be the case that our amended Harsanyi veil delivers more wealth to the able agent, although it will deliver some wealth to both agents. Only for large degrees of risk aversion does the veil of ignorance assign more wealth to the disabled person. Therefore, our example has shown that the veil of ignorance, in general, violates priority. Note that, by Theorem 2, it follows that the general procedure is also anti-prioritarian, because in the special case of risk-isomorphism, we know it is anti-prioritarian.

Of course, the interpretation matters here. A situation where Bob requires twice Andrea's wealth to reach her level of welfare could also be due to Bob's having expensive tastes for which we hold him responsible, and in that case, we might not be so disturbed by the conclusion.[11] But we insist that that is not the problem we are here studying. We are discussing worlds where people differ in their ability to convert wealth into well-being, through no fault of their own.

[11] The issue of expensive tastes is focal in the contemporary literature on distributive justice: see, for the locus classicus, Dworkin (1981a).

We now assert that *priority requires that disabled individuals receive at least as much wealth as able ones.* That is our definition of priority for these worlds. In the environments under discussion, we have a clear way of deciding what it means to say 'i is worse off than j' in the sense of Parfit: it means that i is disabled relative to j, which means 'i requires more wealth than j to reach any given welfare level'.

We have now provided the argument that the veil of ignorance, completed from Harsanyi's first step by the addition of more information and a new axiom, is anti-prioritarian, in the sense that it fails in general to assign at least as much wealth to disabled agents as to able ones. Although Harsanyi's assumption that the IO must possess vNM preferences is, we believe, too strong –in the sense that it is inconsistent with the very reasonable axioms of acceptance and neutrality on our domain of problems, except in a singular case– we have produced a proposal for what the IO should do in the general case (of non-risk-isomorphism), and it also is anti-prioritarian.[12]

3 The Dworkin veil of ignorance

Ronald Dworkin (1981b) outlined a conception of the veil of ignorance that is coherent and can be formally modeled. In Dworkin's view, individuals are to be held responsible for their preferences, but not for their 'resources.' The veil of ignorance is supposed to shield people from knowledge of the characteristics they possess which are 'morally arbitrary' –here, their resources– but not from characteristics which they 'own' –here, their preferences. Thus, behind Dworkin's veil, the soul representing a person knows his person's vNM

[12] One might try to defend Harsanyi's veil of ignorance *and* prioritarianism by saying that, when such monumental issues are at stake as one's wealth for a lifetime, rational individuals would be highly risk averse, thus excluding from the domain of possible worlds, profiles of risk preferences which generate the conflict with priority. We are skeptical. Real people frequently take life-threatening risks that indicate that they do not have excessively high degrees of risk aversion. It is unappealing to say that the only rational persons are the ones who are extremely risk averse.

preferences, but does not know the worldly wealth or the ability his person has. Dworkin's innovation was to propose that souls behind this veil of ignorance enter into insurance contracts with each other, to seek indemnity against bad luck in the birth lottery. In this section, we propose a model of this insurance market.

Here we again present a two-person example, and our model of Dworkin's thought experiment. Suppose we again have Andrea and Bob, and Bob is disabled with respect to Andrea–to wit, he requires $2W$ in wealth to reach the same welfare level as Andrea reaches with W. For the sake of variety, we will now suppose that Andrea and Bob have the same risk preferences over wealth and their vNM utility function is given by

$$v(W) = \sqrt{W}.$$

This time, Andrea and Bob are risk averse.

As we said, Dworkin wishes to hold persons responsible for their risk preferences, but not for their talents. In this case, *talent* is the ability to convert wealth into welfare. Thus, behind the veil of ignorance he constructs, the soul representing a person knows its person's vNM utility function, but does not know its person's talent.

Behind the veil of ignorance, there are two souls–call them α and β–who represent Andrea and Bob, respectively. Each soul knows the welfare producing capacities of Andrea and Bob, and each believes that it will become Andrea or Bob with equal probability (or, to paraphrase, that it will acquire Andrea's and Bob's talent with equal probability).

Thus there are two states of the world, from the viewpoint behind the veil, as follows:

State	α becomes	β becomes
1	Andrea	Bob
2	Bob	Andrea

In state 1, soul α becomes Andrea and soul β becomes Bob; in state 2, the assignments of souls to persons are the other way around. *We* know that state 1 will occur, but the souls behind the veil assign a probability of one-half to each state's occurring.

We assume that, in the real world, Andrea has an endowment W^A of wealth and Bob has an endowment of W^B.

Behind the veil, the souls purchase insurance against bad luck in the birth lottery. We assume (after Dworkin) that the souls have equal purchasing power for insurance. This is where equality enters importantly into Dworkin's view. It does not matter how much purchasing power they each have: we shall say each has zero initially. This means that the only way to purchase insurance for indemnity in one state is to sell insurance for the other's indemnity in the other state.

We model the insurance market as follows. There are two commodities: the first is a contract which will deliver \$1 to the holder should state 1 occur, and the second is a contract which will deliver \$1 to the holder should state 2 occur. Let us denote the prices (behind the veil) for these two commodities by p^1 and p^2. Note that these commodities are purchased behind the veil, using the currency that exists there, which we call clamshell currency, to follow Dworkin's usage.

Denote by x_1^α and x_2^α the amount of commodity 1 and commodity 2, respectively, that soul α purchases. If x is positive, that means she purchases contracts that will deliver to her x dollars if the state of the subscript occurs; if x is negative, that means she will deliver x dollars to someone else should that state occur. The budget constraint for soul α is

$$p^1 x_1^\alpha + p^2 x_2^\alpha = 0,$$

which means that the amount of commodity 1 she can purchase must cost exactly the income she generates by selling commodity 2 (or, the other way around). This constraint derives from the fact that her endowment of 'clamshells' behind the veil is zero. If the soul faces prices (p^1, p^2) then her optimization problem is as follows: choose x_1^a and x_2^a to maximize

$$\frac{1}{2}\sqrt{W^A + x_1^\alpha} + \frac{1}{2}\sqrt{\frac{W^B + x_2^\alpha}{2}} \text{ subject to } p^1 x_1^\alpha + p^2 x_2^\alpha = 0. \tag{10}$$

The objective she maximizes is her expected utility, but to understand it, we must again invoke the notion of deep imaginative empathy. The expression under the first radical is clear: this is what her wealth will be if she becomes Andrea (state 1). The expression under the second radical is trickier. In state 2, she (the soul α) becomes Bob; the wealth she would then have is $W^B + x_2^\alpha$. However, she must evaluate this wealth from Andrea's viewpoint–and by hypothesis the welfare this amount of wealth generates for Bob is exactly the welfare that one-half this amount generates for Andrea. So deep imaginative empathy gives us the objective in (10).

In other words, soul α uses Andrea's vNM utility function to evaluate lotteries over wealth, and she converts wealth that she would experience as Bob into welfare-equivalent wealth, for Andrea. The similarity to our general procedure in the last section should be clear.

In like manner, the optimization problem for soul β is to choose x_1^β and x_2^β to maximize

$$\frac{1}{2}\sqrt{W^B + x_1^\beta} + \frac{1}{2}\sqrt{2(W^A + x_2^\beta)} \text{ subject to } p^1 x_1^\beta + p^2 x_2^\beta = 0.$$

Note that, if soul β becomes Andrea, he must evaluate her wealth in terms of the welfare-equivalent wealth for Bob.

An *equilibrium* in the insurance market consists in:

(1) a pair of prices p^1 and p^2, and

(2) commodity demands $(x_1^\alpha, x_2^\alpha, x_1^\beta, x_1^\beta)$ such that the markets for both commodities clear, that is: $x_1^\alpha + x_1^\beta = 0 = x_2^\alpha + x_2^\beta$.

There is a unique equilibrium[13] in this market. It is:

$$\begin{aligned} p^1 &= p^2 = 1 \\ x_1^\alpha &= \frac{2W^B - W^A}{3}, \; x_2^\alpha = -x_1^\alpha \\ x_1^\beta &= \frac{W^A - 2W^B}{3}, \; x_2^\beta = -x_1^\beta. \end{aligned}$$

[13] To be precise, the demands and supplies are uniquely determined. The prices can be any pair of equal positive numbers.

As we said, we know that, in the event, state 1 occurs; this means that the final wealth levels (under the Dworkinian tax scheme) must be

$$\begin{aligned} W^{A,final} &= W^A + x_1^{\alpha} = \frac{2}{3} \cdot (W^B + W^A) \\ W^{B,final} &= W^B + x_1^{\beta} = \frac{1}{3} \cdot (W^B + W^A) \end{aligned}$$

Thus, disabled Bob ends up with one-third of the total wealth, and able Andrea ends up with two-thirds of the total wealth.

In other words, the Dworkinian insurance market is in general anti-prioritarian. It does *not* (in general) assign at least as much wealth to the disabled individual as to the able individual.

This section and the last one do not prove that all veil-of-ignorance models are necessarily anti-prioritarian: we have established, however, that the two most coherent proposals in the philosophical literature for conceptualizing the veil of ignorance are so, at least under what we consider to be a reasonable addition to those models, namely, information on interpersonal comparability of welfare. Without such comparability, the notion of priority cannot be defined.

4 Discussion

We are concerned with the following syllogism:

A. Justice requires impartiality;

B. Impartiality, as far as justice is concerned, is properly modeled by veil-of-ignorance thought experiments;

C. Veil-of-ignorance thought experiments in general recommend anti-prioritarian allocations.

Therefore,

D. Justice is not prioritarian.

A has a long intellectual history; it is an axiom we do not challenge. C is, so far as we can tell, a fact about veil-of-ignorance models. We have tried to convince the reader of its validity in this paper. We believe that the ethical statement D is invalid: we hold that justice is prioritarian. We therefore must reject the premise B.

Those who reject D could avail themselves of alternatives to rejecting B, such as:

1. Constructing a model of the veil of ignorance that does not conflict with prioritarianism, thus negating C. Perhaps this can be done. Our approach has been to formalize two of the best models of the veil of ignorance offered in the last half century and to show they are anti-prioritarian.[14] But this is not a *proof* that C is true.

2. Refining the definition of impartiality to exclude the veil of ignorance. Perhaps this can be done. We take this to be the strategy of Brian Barry (1989, 1995)–how else could he claim that justice is (or as) impartiality, and also believe that justice is prioritarian or more? Perhaps this is also Scanlon's (1998) strategy: we leave this for others to judge.

There is also the possibility of:

3. Admitting that a second principle, besides impartiality, is required to characterize justice –examples are solidarity, fraternity or reciprocity– and then to argue that impartiality, in conjunction with the new principle, excludes anti-prioritarian allocation rules, like the veil of ignorance.

Approach #3 would be one way of elaborating the rejection of premise B above.

In part 2 of this research project, we do not carry out either strategy 2 or 3 above, desirable as they may be, but rather do something else, which we view as important: we explore what allocation rules are *consistent* with impartiality, priority, and a third principle, solidarity, on a space of important allocation problems.[15] Indeed, we characterize such rules.

[14] We have not studied Binmore's (1994) formulation of the veil. Rawls' (1971) proposal has many problems, which we discuss elsewhere (e.g., Roemer, 1996; forthcoming).

[15] Moreno-Ternero and Roemer (2005)

We conclude by briefly relating our analysis of the Harsanyi veil of ignorance to the recent literature that we cited in the Introduction. We reiterate our central point: we have completed the Harsanyi veil-of-ignorance model by appending two things: information on interpersonal comparability –to be thought of as a fact about the world– and an axiom, relating the IO's vNM preference order on lotteries on extended prospects to this information. Our proposal might appear to be similar to the proposal of Karni (1998) which, besides accepting the principle of acceptance, formulates an axiom of impartiality on the IO preferences. Karni claims this axiom 'renders meaningful interpersonal comparisons of variation in ordinal welfare'.[16] Karni (1988) obtains a representation theorem for the IO preferences, in a more general framework than ours, whose utility function is similar in form to what we obtain in Theorem 1. It is:[17]

$$U^*(i, W) = \frac{v^i(W) - v^i(0)}{v^i(\overline{W}) - v^i(0)}.$$

The preferences on extended prospects represented by U^* are the same as the preferences represented by U in our Theorem 1 if and only if:

(a) all individuals derive equal welfare from zero wealth, and

(b) all individuals derive equal welfare from receiving the entire wealth of the society.

In our environments, this is clearly a singular case. Now U^* clearly satisfies the principle of acceptance. Therefore, by our Theorem 1, it is the generically true that U^* violates the principle of neutrality, on our environments. That is, when interpersonal comparability is meaningful (i.e., the order R exists) Karni's impartial observer will not generally rank two extended prospects as equally desirable if the individuals in them receive equal welfare. We consider this an undesirable feature of Karni's approach. From our viewpoint, Karni's IO is 'partial' in the sense that he must generally bring other

[16] See also Karni (2003).

[17] Karni works on a more general space of states of the world: we have here specialized his result to our environments, where the states are wealth distributions, and individual vNM utility functions on wealth are assumed to be strictly increasing.

considerations than ‘welfare’ to bear in deciding between extended prospects.

Dhillon and Mertens (1999) and Segal (2000) also derive preferences for the IO which, in our environment, specialize to the representation U^* above. They do not, however, claim that interpersonal comparisons can be made in their worlds, and so we cannot levy against their models the criticism we have just made of Karni's.

As our concern in this paper has been to explore the relationship of the veil of ignorance to prioritarianism, and since the concept of priority assumes that interpersonal comparisons are possible, those comparisons must be possible in our formulation of the problem. We did not, however, argue that adding such information is the only way to complete Harsanyi's approach, although, it need hardly be said, we find it a compelling one.

5 Appendix

We provide in this Appendix the proofs of the theorems in Section 2.

Proof of the “if part” in Theorem 1

Suppose first that individuals in the world $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ are risk isomorphic. Let U be the vNM utility function on extended prospects defined as in (5). Clearly, acceptance holds: for each i, $U(i, \cdot)$ is a positive affine transformation of v^i. We also show that neutrality holds.

Let $(i, W)I(j, W')$. We show that $U(i, W) = U(j, W')$. To do so, let $\{W_a^i\}_{i=1}^n$ and $\{W_b^i\}_{i=1}^n$ be two equal-welfare distributions of wealth such that $W_b^i > W_a^i$. We distinguish three cases.

Case 1: $W_a^i \leq W \leq W_b^i$

By risk isomorphism, there exists $p \in [0, 1]$ such that

$$v^i(W) = p \cdot v^i(W_a^i) + (1-p) \cdot v^i(W_b^i) \text{ and } v^j(W') = p \cdot v^j(W_a^j) + (1-p) \cdot v^j(W_b^j).$$

Thus,

$$U(i, W) = \frac{v^i(W) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)} = 1 - p = \frac{v^j(W') - v^j(W_a^j)}{v^i(W_b^j) - v^i(W_a^j)} = U(j, W').$$

Case 2: $W < W_a^i < W_b^i$

By risk isomorphism, there exists $p \in [0,1]$ such that

$$v^i(W_a^i) = p \cdot v^i(W) + (1-p) \cdot v^i(W_b^i) \text{ and } v^j(W_a^j) = p \cdot v^j(W') + (1-p) \cdot v^j(W_b^j).$$

Thus,

$$U(i,W) = \frac{v^i(W) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)} = \frac{-1+p}{p} = \frac{v^j(W') - v^j(W_a^j)}{v^i(W_b^j) - v^i(W_a^j)} = U(j,W').$$

Case 3: $W_a^i < W_b^i < W$

By risk isomorphism, there exists $p \in [0,1]$ such that

$$v^i(W_b^i) = p \cdot v^i(W_a^i) + (1-p) \cdot v^i(W) \text{ and } v^j(W_b^j) = p \cdot v^j(W_a^j) + (1-p) \cdot v^j(W').$$

Thus,

$$U(i,W) = \frac{v^i(W) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)} = \frac{1}{1-p} = \frac{v^j(W') - v^j(W_a^j)}{v^i(W_b^j) - v^i(W_a^j)} = U(j,W').$$

Suppose now that $(i,W)\widehat{R}(j,W')$, where $\widehat{R}$ is the strict part of the order R. Define W^* by $(i,W^*)I(j,W')$, i.e., $W^* = \sigma_j^i(W')$. We now know that $U(i,W^*) = U(j,W')$. But since $v^i(W) > v^i(W^*)$, substitution into the definition of $U(i,W)$ immediately shows that $U(i,W) > U(j,W')$ which shows that neutrality holds.

If we take two other equal-welfare wealth distributions from the ones chosen here, call them $\{\widehat{W_a^i}\}$ and $\{\widehat{W_b^i}\}$, the new function, call it $\widehat{U}$, thereby defined, is an affine transformation of the function U. To see this, six different cases need to be analyzed. We only show one. The remaining cases are analogous.

Case 1: $W_a^i < \widehat{W_a^i} < \widehat{W_b^i} < W_b^i$

By risk isomorphism, there exists $p, q \in [0,1]$ such that

$$v^i(\widehat{W_a^i}) = p \cdot v^i(W_a^i) + (1-p) \cdot v^i(W_b^i) \text{ and } v^i(\widehat{W_b^i}) = q \cdot v^i(W_a^i) + (1-q) \cdot v^i(W_b^i).$$

Then,

$$\widehat{U}(i, W) = k \cdot U(i, W) - c,$$

where $k = \frac{1}{p-q}$ and $c = \frac{1-p}{p-q}$.

Thus, we have shown that there is a vNM preference ordering on extended prospects for the IO that is well-defined, independent of the choice of equal-welfare wealth distributions, and that satisfies acceptance and neutrality.

We show now that the order is unique. Assume there exists another order that satisfies the principles of acceptance and neutrality. By acceptance, this new order should be represented by a vNM utility function $\widehat{U}$ satisfying $\widehat{U}(i, W) = \alpha_i \cdot v^i(W) + \beta_i$ for all i and W, and for some $\alpha_i > 0$ and β_i.

Let $\{W_a^i\}_{i=1}^n$ and $\{W_b^i\}_{i=1}^n$ be two equal-welfare distributions of wealth. Then, by neutrality, there exist two numbers K_a and K_b such that

$$K_a = \alpha_i \cdot v^i(W_a^i) + \beta_i \text{ and } K_b = \alpha_i \cdot v^i(W_b^i) + \beta_i, \text{ for all } i = 1, ..., n.$$

Thus,

$$\alpha_i = \frac{K_b - K_a}{v^i(W_b^i) - v^i(W_a^i)} \text{ and } \beta_i = K_a - \alpha_i \cdot v^i(W_a^i).$$

Consequently,

$$\widehat{U}(i, W) = \alpha_i \cdot v^i(W) + \beta_i = (K_b - K_a) \cdot U(i, W) + K_a,$$

which says that $\widehat{U}$ an affine transformation of the function U. ■

Proof of the "only if part" in Theorem 1

Suppose now that risk isomorphism does not hold. Then, there exist three equal-welfare distributions of wealth denoted $\{W_a^i\}_{i=1}^n$, $\{W_b^i\}_{i=1}^n$ and $\{W_c^i\}_{i=1}^n$ such that $W_c^i > W_b^i > W_a^i$ for all $i = 1, ..., n$, and two individuals i, j such that:

$$v^i(W_b^i) = p^i \cdot v^i(W_a^i) + (1 - p^i) \cdot v^i(W_c^i),$$

and

$$v^j(W_b^j) = p^j \cdot v^j(W_a^j) + (1 - p^j) \cdot v^j(W_c^j),$$

with $p^i \neq p^j$. Thus,

$$p^i = \frac{v^i(W_b^i) - v^i(W_a^i)}{v^i(W_c^i) - v^i(W_a^i)} \neq \frac{v^j(W_b^j) - v^j(W_a^j)}{v^j(W_c^j) - v^j(W_a^j)} = p^j \tag{11}$$

Now, if a vNM preference order on lotteries on extended prospects satisfying acceptance and neutrality exists, and U is a vNM utility function representing it, then

$$U(i, W_k^i) = U(j, W_k^j) \text{ for all } i, j = 1, ..., n \text{ and for all } k = a, b, c.$$

By acceptance, there exist positive numbers α^i and numbers β^i and numbers K_a, K_b, K_c such that:

$$K_a = \alpha^i v^i(W_a^i) + \beta^i,\ K_b = \alpha^i v^i(W_b^i) + \beta^i \text{ and } K_c = \alpha^i v^i(W_c^i) + \beta^i \text{ for all } i.$$

We immediately have by subtracting these equations from each other:

$$\frac{K_b - K_a}{K_c - K_a} = \frac{v^i(W_b^i) - v^i(W_a^i)}{v^i(W_c^i) - v^i(W_a^i)} \text{ for all } i,$$

which contradicts (11). ■

Proof of Theorem 2

Let $\{v^1, v^2, ..., v^n, \overline{W}, R\}$ be a risk-isomorphic world. Let $\{W_a^i\}_{i=1}^n$ and $\{W_b^i\}_{i=1}^n$ be two equal-welfare distributions of wealth such that $W_b^i > W_a^i$. Let $S = \{\omega = (W^1, ..., W^n) : \sum_{i=1}^n W^i = \overline{W}\}$. We denote by Ω the set of allocations selected by the basic procedure (6), i.e.,

$$\Omega = \arg\max_{\omega \in S} \left\{ \sum_{i=1}^n \frac{1}{n} \cdot \frac{v^i(W^i) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)} : \omega = (W^1, ..., W^n) \right\}$$

Fix $k \in \{1, ..., n\}$ and consider the set:

$$\Omega_k = \arg\max_{\omega \in S} \left\{ \sum_{i=1}^n v^k(\sigma_i^k(W^i)) : \omega = (W^1, ..., W^n) \right\},$$

where $\sigma_i^k(W^i)$ denotes the wealth that k requires to reach the same level of welfare as i achieves with wealth W^i. Denote by $\widehat{\Omega}$ the set of allocations selected by the general procedure (8). Then, $\widehat{\Omega}$ is average of the sets Ω_k. Formally,

$$\widehat{\Omega} = \left\{ \sum_{k=1}^{n} \frac{1}{n} \cdot \omega_k : \omega_k \in \Omega_k \right\},$$

Claim. $\Omega = \Omega_k$ for all $k \in \{1, ..., n\}$.

Proof of the claim. Denote generically $\omega = (W^1, ..., W^n)$ and fix $k \in \{1, ..., n\}$. By definition, $(i, W^i)I(k, \sigma_i^k(W^i))$ for all $i = \{1, ..., n\}$. Now, by the definition of risk-isomorphism, it follows that

$$\frac{v^i(W^i) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)} = \frac{v^k(\sigma_i^k(W^i)) - v^k(W_a^k)}{v^k(W_b^k) - v^k(W_a^k)}, \text{ for all } i = \{1, ..., n\}.$$

Thus,

$$\sum_{i=1}^{n} \frac{v^i(W^i) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)} = \sum_{i=1}^{n} \frac{v^k(\sigma_i^k(W^i)) - v^k(W_a^k)}{v^k(W_b^k) - v^k(W_a^k)} = \lambda \cdot \sum_{i=1}^{n} v^k(\sigma_i^k(W^i)) - \mu,$$

where

$$\lambda = \frac{1}{v^k(W_b^k) - v^k(W_a^k)} > 0 \text{ and } \mu = \frac{v^k(W_a^k)}{v^k(W_b^k) - v^k(W_a^k)} > 0.$$

Therefore, an allocation $\omega \in S$ maximizes $\sum_{i=1}^{n} \frac{1}{n} \cdot \frac{v^i(W^i) - v^i(W_a^i)}{v^i(W_b^i) - v^i(W_a^i)}$, if and only if ω maximizes $\sum_{i=1}^{n} v^k(\sigma_i^k(W^i))$. In other words, $\omega \in \Omega \iff \omega \in \Omega_k$. This proves the claim.

Case 1: Ω is a singleton set.

If $\Omega = \{\omega\}$ then, by the claim, $\Omega = \{\omega\} = \Omega_k$ for all k, and therefore $\widehat{\Omega} = \{\omega\} = \Omega$. In other words, the general procedure described in (8) would select a unique allocation of wealth that would coincide with the allocation selected by maximizing the utility function in (5).

Case 2: Ω is not a singleton set.

It is straightforward to show that $\Omega \subseteq \widehat{\Omega}$.[18] The converse inclusion is also true, provided that we assume all individual vNM utility functions are

[18] Note that if $\omega \in \Omega$ then, by the claim, $\omega \in \Omega_k$ for all $k = \{1, ..., n\}$ and $\omega = \frac{1}{n} \cdot \sum_{k=1}^{n} \omega$.

concave. Formally, let $\omega \in \widehat{\Omega}$. Then, $\omega = \frac{1}{n} \cdot \sum_{k=1}^{n} \omega_k$, for some $\omega_k \in \Omega_k$. Since $\Omega = \Omega_k$ for all $k = 1, ..., n$, then $\omega_k \in \Omega$ for all $k = 1, ..., n$. Now, if all individual vNM utility functions are concave, it follows that Ω is a convex set and therefore $\omega \in \Omega$.[19] ■

[19] Note that, had we defined the general procedure (8) by selecting any convex combination (not necessarily with equal weights).of the wealth distributions maximizing (7), the theorem would also hold.

References

[1] Arrow, K. J. (1963), **Social Choice and Individual Values**. Yale University Press

[2] Arrow, K. J. (1977), Extended sympathy and the possibility of social choice. **American Economic Review**, Papers and Proceedings 67, 219-229.

[3] Barry, B. (1989), **Theories of Justice**, Vol. 1. Berkeley. University of California Press.

[4] Barry, B. (1995), **Justice as impartiality**, Oxford: Clarendon Press.

[5] Binmore, K. (1994), **Playing fair**, MIT Press. Cambridge, MA.

[6] Cohen, G.A. (1992), "Incentives, inequality, and community," in Grethe Peterson, ed., The Tanner Lectures on Human Values, Salt Lake City: University of Utah Press

[7] Dhillon, A and Mertens, J.F. (1999), Relative utilitarianism: An improved axiomatization. **Econometrica** 67, 471-498.

[8] Dworkin, R. (1981a), What is equality? Part 1: Equality of welfare. **Philosophy & Public Affairs** 10, 185-246.

[9] Dworkin, R. (1981b), What is equality? Part 2: Equality of resources. **Philosophy & Public Affairs** 10, 283-345.

[10] Harsanyi, J. (1953), Cardinal utility in welfare economics and in the theory of risk-taking, **Journal of Political Economy** 61, 434-435.

[11] Harsanyi, J. (1977), **Rational behavior and bargaining equilibrium in games and social situations**, Cambridge University Press.

[12] Karni, E. (1998), Impartiality: definition and representation, **Econometrica** 66, 1405-1415.

[13] Karni, E. (2003), Impartiality and interpersonal comparisons of variations in well-being, **Social Choice and Welfare** 21, 95-111.

[14] Kolm, S. C. (1972). **Justice et équité**. Editions du Centre National de la Recherche Scientifique. English trans. Justice and Equity. MIT Press

[15] Moreno-Ternero, J.D., and Roemer, J.E. (2005), Impartiality and priority. Part 2: a characterization with solidarity. **Mimeo**. Yale University

[16] Mongin, P (2001). The impartial observer theorem of social ethics. **Economics and Philosophy** 17, 147-179

[17] Parfit, D. (1997), **Equality and Priority**. Ratio: New Series 10, 202-221. Blackwell Publishers Ltd. Oxford.

[18] Rawls, J. (1971), **A Theory of Justice.** Harvard University Press. Cambridge, MA.

[19] Roemer, J.E. (1996), **Theories of Distributive Justice**, Harvard University Press, Cambridge MA.

[20] Roemer, J.E. (2002), Egalitarianism against the veil of ignorance, **Journal of Philosophy** 99, 167-184.

[21] Roemer, J.E. (2004), Eclectic distributional ethics, **Politics, Philosophy, and Economics** 3, 267-281.

[22] Roemer, J.E. (forthcoming), Solidarity, priority, and distributive justice, in C. Sypnowich (ed.), [a festschrift for G.A. Cohen]. Oxford University Press (in press).

[23] Scanlon, T. (1998), **What we owe to each other**, Harvard University Press. Cambridge, MA.

[24] Segal, U. (2000), Let's agree that all dictatorships are equally bad. **Journal of Political Economy** 108, 569-589.

[25] Sen, A. (1970), **Collective Choice and Social Welfare**. North Holland

[26] Suppes, P. (1966), Some formal models of grading principles. **Synthese** 6, 284-306

[27] Suzumura, K. (1983), **Rational Choice, Collective Decisions and Social Welfare**. Cambridge University Press

[28] Temkin, L.S. (1993), **Inequality**, New York: Oxford University Press.

[29] Weymark, J. (1991), A reconsideration of the Harsanyi-Sen debate on utilitarianism, in **Interpersonal Comparisons of Well-Being**, ed. by J. Elster and J. Roemer. Cambridge University Press, Cambridge, U.K.