

Egalitarianism Against the Veil of Ignorance

By

John E. Roemer

September 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1328



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

<http://cowles.econ.yale.edu/>

“Egalitarianism against the veil of ignorance”

by

John E. Roemer¹

1. Introduction

The construct of a veil of ignorance has been used as a powerful tool of analysis in modern theories of distributive justice, in various forms, most prominently by three authors: Harsanyi (1953, 1977), Rawls (1971), and Dworkin (1981b). Of these, Rawls has claimed that the veil-of-ignorance argumentation leads to a kind of egalitarianism; Harsanyi, while proclaiming himself a utilitarian, believed that the kind of utilitarianism produced by his veil of ignorance would not be offensively inegalitarian; and Dworkin, although not using his veil of ignorance to justify resource egalitarianism, does use it to deduce or compute what the resource-egalitarian distribution *is*.

My aim, here, is to argue that veil-of-ignorance arguments, when properly done, contravene fundamental egalitarian principles. If I am right, and if the veil of ignorance is the right tool for deducing what distributive justice requires, then those of us who advocate egalitarianism must cease to do so. It will perhaps not surprise the reader to learn that I will, instead, indicate in the paper's conclusion why I believe the veil of ignorance is the wrong tool for deducing the requirements of distributive justice. That final argument is, however, more tentative in nature than what precedes it.

I have already written elsewhere (Roemer[1996, chapter 4]) that I think Harsanyi's 'impartial observer' argument, in which he deduces how a representative soul,

¹ Elizabeth S. and A. Varick Stout professor of political science and economics, Yale University. I am grateful to G.A. Cohen, Mathias Risse, and Joaquim Silvestre for their comments on an earlier draft.

behind an appropriate veil of ignorance, should solve the decision problem of allocating wealth among individuals in the world, is basically correct: what I mean is that, *given* the premise that the veil of ignorance is the right tool for the job, then Harsanyi has done the decision theory properly. Two comments are in order. First, Harsanyi put a false gloss on his conclusion -- namely, that the 'impartial observer (IO)' should be described as 'utilitarian.' This resulted from a confusion between the maximization of expected utility of one decision maker, and the maximization of the *average* utility of a set of individuals². The former makes sense with an (ordinal) preference relation over lotteries, and the latter makes sense only when utility is interpersonally comparable to a degree -- at least cardinally measurable and unit comparable (CUC). My claim is that Harsanyi's decision theory is correct, but he is wrong to call his soul a utilitarian: indeed, utilitarianism is not a coherent concept, given the information postulated by Harsanyi. It was unsurprising that Harsanyi erred in this way, as clarity on the problem of the measurability and comparability of utility was not achieved until the 1970s, two decades after Harsanyi wrote. (These later developments are surveyed in Roemer [1996, chapter 1].)

Secondly, Harsanyi does not, in fact, completely solve the decision problem of the IO behind the veil of ignorance; rather, he shows that the IO's optimization problem must be one of a continuum of possible optimization problems -- that is, that it must maximize *some* positive linear combination of the von Neumann -- Morgenstern utility functions of the individuals in question. But Harsanyi provides no (correct) argument for what the choice of those weights should be. That choice is indeterminate in his environment.

² The point was made initially by Sen (1977), and later elaborated upon by Weymark (1991).

I, and many others, have written that Rawls's difference principle does not flow from any conventional decision problem of a soul in his original position. The issue here is whether it is appropriate to take such a soul as being infinitely risk averse, for that seems to be the only sensible way of deducing the difference principle. But, apart from his exotic decision theory, Rawls's original position seems to be incorrectly specified for *his* purposes – at least I have argued this point (Roemer[1996, Chapter 5], and more recently, Roemer[in press]). The claim is that, in the original position, morally arbitrary aspects of persons' characteristics (such as the wealth they are assigned in the birth lottery) are treated identically to non-morally-arbitrary characteristics (principally, their plans of life), and such a move cannot reflect Rawls's views. I will not repeat that argument here.

Dworkin (1981b) constructs a 'thin' veil of ignorance, where the souls behind the veil know the preferences of those individuals whom they represent, but do not know those individuals' resource endowments. Unlike Harsanyi and Rawls, the activity behind the Dworkinian veil is not a decision problem of an individual soul, but is a market among many souls³. I have argued that, although Dworkin's own conception of how this insurance market should be modeled is wrong, it is a straight-forward exercise in general equilibrium theory to model properly the insurance market behind Dworkin's veil (Roemer [1985, 1996, in press]). I will review that exercise in section 4 below. It turns out that, when properly modeled, the (modified) Dworkinian insurance market can produce – and will in general produce-- some unpleasant results, that, I claim, disqualifies it as a mechanism for implementing equality of resources. Indeed, we will

³ Although Rawls claims the activity behind his veil is a 'social contract,' in fact, formally speaking, it is a decision problem of a single soul, as all souls in the Rawlsian original position are identical.

see that the problem with Dworkin's mechanism is related to the problem diagnosed in the generalization of Harsanyi's mechanism.

In the next section, I will present what I think is a compelling solution to a representative soul's decision problem, in a particularly simple environment. I will build upon Harsanyi. To be specific, I will append one new postulate to Harsanyi's Impartial Observer argument, and thereby deduce the *unique* decision problem that the soul should solve. I will then study, in a series of examples, the distribution of wealth that follows from this solution, and argue that it contravenes an important egalitarian postulate. Thus, I will claim that a properly modeled decision problem behind the veil of ignorance cannot be used to justify egalitarianism.

Some might say this is no surprise -- they already knew. I disagree. Despite the qualifications that their authors have added over the years, it is, I think, unquestionable that serious thinkers with an egalitarian bias have often hoped to justify their ethos with veil-of-ignorance arguments. (In my view, this includes Dworkin, who, although claiming that his veil is used only to compute what egalitarianism requires, and not to argue for it, nevertheless hopes that the hypothetical insurance market he constructs behind a thin veil of ignorance will render his conception of resource egalitarianism appealing.) My intention is to show that such hopes are fruitless.

2. Harsanyi refined

There is a population of individuals, defined as a set of types, identified by a parameter $\omega \in \Omega$, where Ω is a sample space; types are distributed according to a probability measure F on Ω . Each type is characterized by two functions of wealth,

denoted $u(\cdot, \omega)$ and $v(\cdot, \omega)$. $u(\cdot, \omega)$ is a von Neumann-Morgenstern (vNM) utility function representing type ω 's preferences over wealth lotteries, and $\{v(\cdot, \omega) \mid \omega \in \Omega\}$ is a profile of functions defined on wealth which measures the welfare of individuals in an interpersonally comparable manner. Thus, $v(W^1, \omega^1) = v(W^2, \omega^2)$ means that a type ω^1 individual with wealth level W^1 and a type ω^2 individual with wealth level W^2 enjoy the same welfare level.

The only consistency requirement, with regard to the functions u and v is that, for each ω , they give the same orderings over wealth levels. Thus, one possible lottery is a sure thing, where an individual receives a given level of wealth with probability one, and the von Neumann-Morgenstern preferences of the individual must rank these sure-thing lotteries according to the individual's *welfare*. That is to say:

$$(\forall \omega \in \Omega)(\forall W, W')(u(W, \omega) \geq u(W', \omega) \Leftrightarrow v(W, \omega) \geq v(W', \omega)).$$

If we assume that u and v are strictly monotone increasing in W , then this requirement is trivially satisfied. (Of course, it would not be a trivial statement, were wealth a multi-dimensional vector.)

In the language of social-choice theory, we are given, for each individual, his ordinal preferences on wealth lotteries, which are assumed to obey the vNM axioms, and an interpersonally level-comparable measure of welfare. It is noteworthy that we lack sufficient information, in this environment, for utilitarianism to be coherent, for we have no *unit comparable* measure of welfare. (Of course, one could endow the profile v with unit comparability as well, and then utilitarianism would be coherent. But such information is unnecessary for that argument to follow.)

We will construct the vNM utility function of the soul that represents the generic member of Ω , that is, the soul whose job it will be to decide upon the distribution of wealth in the world, knowing only that it could become, in the birth lottery, any of the actual types in the population, according to the probability distribution F . Let a *prospect* be an ordered pair (W, ω) . We assume that the soul has a preference order over lotteries of prospects, which obeys the vNM axioms. Let U be the soul's vNM utility function on prospects, that is the vNM utility function that represents the soul's preferences over such lotteries. We assume:

A1. *Principle of acceptance* (Harsanyi[1953]). On the domain of wealth lotteries over prospects involving only a single type ω , $u(W, \omega)$ represents the soul's vNM preferences.

A1 is Harsanyi's principle of acceptance. It is justified by saying that, if the soul were asked to choose between two wealth lotteries in which only the wealth of a single individual, of some given type, is at issue, then it should choose the lottery that the individual in question would choose. This, at least, is required for the soul to be 'representative.'

It thus follows from the vNM representation theorem that there are functions $\alpha(\omega)$ and $\beta(\omega)$, with $\alpha(\omega) > 0$, such that:

$$\forall \omega \in \Omega \quad U(W, \omega) = \alpha(\omega)u(W, \omega) + \beta(\omega). \quad (2.1)$$

How does the soul choose between prospects involving the possibility of its becoming persons of two different types? We assert:

A2. *Principle of neutrality*. Let L be a lottery with over two prospects, (W, ω) and (W', ω') . The soul is indifferent between the two prospects if the types in those prospects enjoy equal welfare levels at the wealths in question; that is, if $v(W, \omega) = v(W', \omega')$.

This principle is one of neutrality, for it says that the soul imposes no external preferences over the nature of the lives led by individuals of different types: it is concerned only with the universal concept of well-being called welfare. Were the soul to base its choice over prospects involving different types on a view that some kinds of life were better than other kinds, for reasons not related to the welfare experienced by those types, it would, I think, be shirking its duty to represent all types. For by saying that two individuals enjoy the same level of welfare, I mean to say that they rate their lives as being equally good.

Now let $W^1(\omega)$ be an allocation of wealth to types at which all types enjoy the same welfare, that is:

$$\forall \omega \in \Omega \quad v(W^1(\omega), \omega) = k^1,$$

and let $W^2(\omega)$ be another wealth allocation at which all types enjoy the same level of welfare, which is greater than they enjoy at the first allocation, that is:

$$\forall \omega \in \Omega \quad v(W^2(\omega), \omega) = k^2 > k^1.$$

(These two distributions of wealth need not be feasible for the society in question.) By A2, we have:

$$\begin{aligned} \forall \omega \in \Omega \quad \alpha(\omega)u(W^1(\omega), \omega) + \beta(\omega) &= k^{1'}, \\ \alpha(\omega)u(W^2(\omega), \omega) + \beta(\omega) &= k^{2'} \end{aligned} \tag{2.2}$$

for some constants $k^{2'} > k^{1'}$.

Now we are free to choose any vNM utility function from a positive affine family to represent type ω 's vNM preferences over wealth lotteries; we now suppose that we

have chosen the profile $u(\cdot, \omega)$ which renders $k^{1'} = 0$ and $k^{2'} = 1$. (This, indeed, fixes the function u .) Then we can solve (2.2) for the functions α and β :

$$\alpha(\omega) = \frac{1}{D(\omega)}, \quad \beta(\omega) = \frac{-u(W^1(\omega), \omega)}{D(\omega)}, \quad (2.3)$$

where

$$D(\omega) = u(W^2(\omega), \omega) - u(W^1(\omega), \omega),$$

and so we have:

$$U(W, \omega) = \frac{u(W, \omega)}{D(\omega)} - \frac{u(W^1(\omega), \omega)}{D(\omega)}. \quad (2.4)$$

We have thus completely determined the souls' vNM preferences on lotteries over prospects, for by the vNM axioms, it suffices to compute its utility function on the pure prospects themselves, which we display in (2.4). It is worth observing that the vNM preferences represented by U are independent of the particular profile u we chose to represent the preferences of individual types. If we replace each function $u(\cdot, \omega)$ with a positive affine transfer of itself, the function U is unchanged (note that the various constants factor out of both numerator and denominator in (2.4)). Thus, the function U depends only on the vNM *preferences* of the types, on the function v , and on the two wealth distributions chosen.

But this raises a final issue. The determination of U apparently depends upon the particular wealth distributions W^1 and W^2 chosen. It is, however, easy to verify that, if we chose two other wealth distributions at which welfare was equal across individuals, then the vNM utility function derived for the soul is equivalent to U . Thus, the preferences over prospects represented by U are in fact well-defined by axioms A1 and A2.

We now analyze the soul's decision problem, which is to choose a distribution of wealth among types. The soul chooses that feasible distribution of wealth $W(\omega)$ that maximizes its expected utility when facing the 'birth lottery' F . We will assume that average wealth in the society is given at \bar{W} , and is independent of how wealth is distributed. (This is not necessary, but will simplify our examples.) Then the soul's problem is:

$$\begin{aligned} \underset{W(\cdot)}{\text{Max}} \int U(W(\omega), \omega) dF(\omega) \\ \text{s.t.} \quad \int W(\omega) dF(\omega) = \bar{W}. \end{aligned} \quad (2.5)$$

If u is concave in W , then U is concave in W , and the solution to program (2.5) entails:

$$(C1) \quad \text{for some } \lambda > 0, \frac{u'(W(\omega), \omega)}{D(\omega)} \leq \lambda, \text{ with equality holding when } W(\omega) > 0,$$

$$\text{and} \quad (C2) \quad \int W(\omega) dF(\omega) = \bar{W},$$

where u' denotes the derivative of u with respect to W .

We thus assert that we have solved the problem of deciding upon the distribution of wealth using the veil-of-ignorance approach – at least in the simple environment postulated here (where all types are risk neutral or risk averse [thus concave vNM utilities], average wealth is independent of wealth's distribution, only welfare matters, etc.).

3. Some examples

We next compute the soul's choice of wealth distribution for some examples.

Example 1. All types are risk-neutral.

In this case, because we are at liberty to choose any vNM utility function from among a type's positive affine family, we choose:

$$\forall \omega \in \Omega \quad u(W, \omega) = W.$$

Then (C1) becomes

$$\frac{1}{W^2(\omega) - W^1(\omega)} \leq \lambda, \text{ with equality when } W^*(\omega) > 0.$$

(We denote by $W^*(\omega)$ the optimal solution to (2.5).) It follows that

$$\lambda = \text{Sup}_{\omega} \frac{1}{W^2(\omega) - W^1(\omega)},$$

and so

$$W^*(\omega) = 0, \text{ except when } \omega = \text{Arg Inf}_{\omega} (W^2(\omega) - W^1(\omega)).$$

Thus, the only types who receive any wealth at the optimal distribution are those for whom the wealth difference $(W^2(\omega) - W^1(\omega))$ is minimal.

It is reasonable to assert, I think, that:

A3a. If ω is disabled and ω' is able, but the two types are in other ways similar, then

$$W^2(\omega) - W^1(\omega) > W^2(\omega') - W^1(\omega').$$

That is, to effect a welfare saltus between two given welfare levels, for two individuals who are similar except for a disability, a larger wealth increment is needed for the disabled person.

Indeed, the posture that we implicitly adopt in A3a is that we know what disability means, and A3a asserts a relationship between disability and the production of welfare. But we could stand differently, and take A3a as the *definition* of disability.

In like manner, we might assert:

A3b. If ω has expensive tastes and ω' has cheap tastes, but the two types are in other ways similar, then $W^2(\omega) - W^1(\omega) > W^2(\omega') - W^1(\omega')$.

If we accept A3a and A3b, either as reasonable properties of disability and expensive tastes, or as definitions of those properties, then, in example 1, the only types who receive any wealth at the optimal distribution are those who are able and/or have cheap tastes.

Example 2. There is a fixed number p , $0 < p < 1$, such that $\forall \omega \in \Omega \quad u(W, \omega) = \frac{W^{1-p}}{1-p}$.

This is the constant-relative-risk-aversion (CRRA) utility function. All types are equally risk averse in this example. We take $\Omega = [0,1]$, $W^2(\omega) = k^2\omega$, $W^1(\omega) = k^1\omega$, $k^2 > k^1$.

Assuming, for the moment, an interior solution where everyone receives positive wealth, (C1) reduces to

$$\frac{W^*(\omega)^{-p}(1-p)}{D_1(\omega)} = \lambda,$$

where $D_1(\omega) = W^2(\omega)^{1-p} - W^1(\omega)^{1-p}$, and so

$$W^*(\omega) = \left(\frac{1-p}{\lambda} \right)^{\frac{1}{p}} D_1(\omega)^{-\frac{1}{p}}; \quad (2.6)$$

λ may now be chosen to render (C2) true, and so (2.6) is indeed the optimal solution.

Note that $D_1(\omega)^{-\frac{1}{p}}$ is monotone decreasing in $D_1(\omega)$. Since $W^2(\omega) - W^1(\omega) = (k^2 - k^1)\omega$ is increasing in ω , we identify larger ω with more disability. But

$D_1(\omega) = \omega^{1-p}(k^{2^{1-p}} - k^{1^{1-p}})$ is also increasing in ω . Thus we have, in this example, that the able receive more wealth at the chosen distribution than the disabled.

Example 3. $\Omega = [0, 1]$, F is the uniform distribution, and $u(W, \omega) = \frac{W^{1-\omega}}{1-\omega}$. In this

example, types have different but constant relative degrees of risk aversion, *and* a person's type is completely specified by his risk preferences. (That is, all individuals with the same risk preferences have the same welfare function $v(\cdot, \omega)$.)

To enable us to compute a solution in this example, we need to specialize further.

We show:

Proposition. *In example 3, if*

$$W^2(\omega) = \omega^{\frac{2\omega}{1-\omega}} \text{ and } W^1(\omega) = \omega^{\frac{2\omega+1}{1-\omega}}, \quad (2.7)$$

then

$$W^*(\omega) = \frac{1}{\omega^2} \left(\frac{1}{\lambda} \right)^{\frac{1}{\omega}},$$

where λ is the solution of the equation $\lambda \text{Log} \lambda = \bar{W}^{-1}$.

Proof:

1. (C1) implies

$$W^*(\omega) = \left(\frac{1-\omega}{\lambda D_2(\omega)} \right)^{\frac{1}{\omega}},$$

where $D_2(\omega) = W^2(\omega)^{1-\omega} - W^1(\omega)^{1-\omega}$.

Using (2.7) we have:

$$D_2(\omega) = (1 - \omega)\omega^{2\omega},$$

and we hence compute

$$W^*(\omega) = \lambda^{\frac{1}{\omega}} \frac{1}{\omega^2}.$$

Now (C2) in this example becomes:

$$\int_0^1 \lambda^{\frac{1}{\omega}} \frac{1}{\omega^2} d\omega = \bar{W}. \quad (2.8)$$

Let $q = -\frac{1}{\omega}$ be a change of variable; since $\frac{dq}{d\omega} = \frac{1}{\omega^2}$, we may transform (2.8) to:

$$\int_{-\infty}^{-1} \lambda^q dq = \bar{W}.$$

Integrating gives:

$$\frac{\lambda^q}{\text{Log}\lambda} \Big|_{-\infty}^{-1} = \bar{W}; \quad (2.9)$$

if $\lambda > 1$, (2.9) evaluates to:

$$\frac{1}{\lambda \text{Log}\lambda} = \bar{W},$$

as claimed in the proposition's statement. Now $\bar{W} > 0$ tells us that indeed $\lambda > 1$, and the proposition is proved.

In figure 1, we plot the functions W^2 , W^1 , and W^* of the Proposition in the (ω , wealth) plane. The higher and lower thin lines are the graphs of the functions W^2 and W^1 , respectively. Each heavy line is the solution W^* for a particular value of \bar{W} , with higher average wealths associated with the higher graphs. We see that the low ω types

require a large wealth increment to make the saltus between the two welfare levels, while the high ω types require a much smaller wealth increment. We therefore identify low ω 's with the disabled. The optimal distribution is non-monotonic, but it is generally true at all levels of average wealth that the most able receive greater wealth than the most disabled. The biggest wealths, however, go to types in the mid-range of 'ability.'

In this example, low ω is associated both with disability, and with a low degree of risk aversion ($\omega=0$ is risk neutral). Thus, with an equal wealth distribution as benchmark, wealth is redistributed from disabled types who are almost risk neutral to able types who are highly risk averse.

We cannot say precisely what utilitarianism would recommend in any of these examples, because, as I wrote, we do not have sufficient information to define utilitarianism. Utilitarianism would distribute wealth to maximize the average value of v , but that average value is not well-defined without unit comparability.

4. Dworkin's insurance mechanism

Dworkin, unlike Rawls or Harsanyi, does not employ a veil of ignorance to argue for his political view, but rather, as he says, to define what it consists in. In brief, Dworkin believes that equality-of-resources is a state in which the transferable resources (principally, money) that individuals have, provide them appropriate compensation for the shortfalls they sustain in their endowments of non-transferable resources. Dworkin believes that individuals are responsible for their preferences, but not for their resource

endowments, and what distribution of money is ‘appropriate’, in the above sense, must rest upon this fact.

In Dworkin’s scheme, individuals are represented by souls behind a veil of ignorance who know the preferences (in particular, the preferences over lotteries) of the persons they represent, but do not know what allocation of resources – internal or external—their persons will receive in the birth lottery. We will identify external resources with wealth and call internal resources ‘non-transferable.’ Each soul is given the same amount of purchasing power with which to purchase insurance against a bad draw in the birth lottery, which allocates those resources. There will be an equilibrium in this insurance market, which will redistribute money wealth among actual persons, contingent upon the realization of the birth lottery (when insurance contracts pay out). Dworkin recommends a tax scheme in the world that implements the contracts that the souls sign behind his veil. This scheme models Dworkin’s view that individuals should be responsible for their preferences (as the souls employ those preferences when choosing contracts), but not for their resources (as all souls have the same purchasing power behind the veil).

There is a canonical way of modeling this scheme using the notion of a market for contingent claims. Imagine that each soul begins with *zero* income behind the veil. (In particular, each has an equal amount of purchasing power.) The probability distribution of resources is known to all souls. A *state of the world* is a particular allocation of resources to souls (that is to say, to preference orders). Let s be a state of the world. The commodity X_s is the promise by the insurance company to deliver \$1000 to the purchaser of the commodity, should state s occur. If I, a soul, purchase x_s units of X_s ,

then, if state s occurs, the person I represent receives a check for $1000x_s$ dollars from the insurance company. If I sell x_s units of this commodity, then, should state s occur, my person writes a check for $1000x_s$ to the insurance company. Feasibility of these contracts requires that all promises to pay can be met with the wealths available to the persons involved in the state in question.

A price vector is a vector (p_1, \dots, p_s, \dots) , one for each state, where p_s is the price of one unit of X_s . It is important to note that these commodities are bought and sold before the veil is lifted: the market takes place behind the veil of ignorance, and among souls, not persons. An equilibrium is a price vector and a matrix of demands $\{x_{is}\}$, where x_{is} is the demand by soul i for commodity X_s , at which the total demand for the commodity X_s equals the total supply of this commodity, for every s . (Supplies are negative numbers.) It is important to note that the only way a soul acquires the purchasing power to purchase delivery of dollars in some states is by selling the promise to deliver dollars in other states. Thus, any soul who wishes to buy insurance in some states must sell insurance in other states. Typically, we imagine a soul's selling contracts to deliver dollars in states in which its person is wealthy and well-resourced, and purchasing promises to deliver dollars in states in which its person is poor and/or poorly resourced.

Formally, we define an equilibrium as follows. A prospect is an ordered pair (W, s) -- the enjoyment of transferable wealth W in state s . In different states, the souls are embodied as persons with different endowments of non-transferable resources and wealth -- but a soul always has the preferences of one individual. Let W_{is} be the money wealth of soul i in state s , and let E_{is} be the endowment of non-transferable resources of soul i in state s . Then soul i 's vNM utility function is defined on pairs (W, E) . Let π_s be

the probability that state s occurs, and let the states be enumerated $1, 2, \dots, S$. An equilibrium is a price vector $p = (p_1, \dots, p_S)$ and a matrix of demands $\bar{x} = \{\bar{x}_{is}\}$ such that:

(1) (utility maximization) For each i , the vector of demands $(\bar{x}_{i1}, \dots, \bar{x}_{iS})$

maximizes $\sum_s \pi_s u_i(W_{is} + x_s, E_{is})$ subject to the budget constraint $\sum_s p_s x_s \leq 0$ and the

feasibility constraints $W_{is} + x_s \geq 0$, for all s ; and

(2) (market clearing) for each s , $\sum_i x_{is} = 0$.

The budget constraint in (1) states that souls must finance the purchase of deliveries of wealth in some states by selling contracts promising to deliver wealth in other states, and the other constraints in (1) say that there is no state in which a soul can promise to deliver more wealth to the insurance company than its person is endowed with, in that state.

The post-transfer wealth of the person whom soul i becomes in state s is $W_{is} + \bar{x}_{is}$

(remember that if a person i pays out in state s then $\bar{x}_{is} < 0$).

It is a consequence of the Arrow-Debreu existence theorem that, if the utility functions u_i are quasi-concave (which, informally, we may think of as happening if agents are risk neutral or risk averse), then such an equilibrium exists⁴. It is perhaps important to reiterate that all *economic* activity -- that is, the buying and selling of commodities, which are *contingent claims*-- happens behind the veil of ignorance, and using the unit of exchange that is recognized there, not real-world dollars. After the veil

⁴ I don't say equilibrium exists only under this condition.

is lifted, and a state of the world is realized, then deliveries of dollars are made among real persons.

Now the actual world we live in is one of these states, call it s^* . At equilibrium, souls hold contracts either to deliver dollars to the insurance company, to receive dollars from the insurance company, or to do neither, should state s^* occur. When these transactions occur, a transfer of dollars among citizens will have transpired. Dworkin's tax system would mimic this transfer.

What I have shown (Roemer [1985, 1996, 2001]) is that this tax system will generally behave 'pathologically,' from a resource-egalitarian viewpoint. The simplest example is the following. There are two persons, Andrea and Bob, and there is one internal resource that we may call 'ability.' The money wealth endowment of the each, in the real world, is \$10,000. In the real world, Andrea is able and Bob is disabled -- to wit, she possess four units of 'ability' and he, only one unit -- and they both are born with half of the wealth, \$10,000 each. Andrea and Bob have the *same* von Neumann–Morgenstern utility function over ability and wealth, namely:

$$u(W, E) = E^{\frac{1}{2}} W^{\frac{1}{2}},$$

where E is units of ability and W is units of money wealth.

Because they have the same preferences, it turns out that the equilibrium in the market for contingent claims behind Dworkin's veil, where each soul assumes that it will receive an allotment of ability of one (four) with probability one-half (one-half), is characterized by the solution to the following optimization problem:

$$\text{Max}_x \frac{1}{2} u(10 + x, 1) + \frac{1}{2} u(10 - x, 4).$$

Let x^* be the solution to this program. Then the Dworkin tax scheme is that in which Andrea will transfer x^* thousand dollars to Bob.

The solution to this program is $x^* = -6$; in other words, *disabled Bob must transfer \$6000 to able Andrea.*

Thus the Dworkin insurance market can produce the same pathology as afflicted the allocation chosen by the generalized Harsanyi IO in the last section. I contend that no resource-egalitarian would endorse this recommendation.

Unlike the situation in sections 2 and 3, there is no interpersonal comparison of welfare required here. The utility function u represents a von Neumann – Morgenstern preference order over lotteries on prizes (W, E) . Bob and Andrea have the same preference order. Nowhere need we posit that the welfare of Bob and Andrea is interpersonally comparable. Thus, to speak of ‘equalizing the welfares’ of Bob and Andrea is meaningless, unless we postulate, in addition, some way of comparing their welfare levels. Thus, the example is perfectly consistent with Dworkin’s view, enunciated in Dworkin (1981a), that interpersonal comparisons of welfare are not available.

One might protest that my example is cooked to produce the unpleasant⁵ result. On the contrary, I contend that, were we to model preferences in our own world, complex as they are, we would surely encounter instances of this kind of transfer in the equilibrium of the market for contingent claims behind the Dworkinian veil of ignorance. The example is not bizarre; the ‘pathology’ that it exhibits is, for all practical purposes, unavoidable.

⁵ When I thus describe the result, I mean, from an egalitarian’s viewpoint. Those who are persuaded that Dworkin’s insurance mechanism is a compelling ethical procedure will not find the result unpleasant.

Without level-comparable information on welfare across persons, we cannot speak of the ‘prioritarian’ view, which says that we should give priority (in the distribution of wealth) to the worse-off. (On prioritarianism, see Parfit[1997].) We do not have such information in the Dworkin insurance scheme. But because the insurance mechanism ignores such level-comparable information, the allocation it recommends cannot generally favor the worse off. In particular, suppose we *do* have level-comparable information about welfare, and we believe that Bob with \$10,000 is worse off than Andrea with \$10,000 in the above example -- not an outrageous view. Then clearly the insurance mechanism behaves in an anti-prioritarian manner. Indeed, it is precisely a prioritarian intuition which lies behind my claim that distributive justice should render disabled persons more wealthy than able ones, other things equal.⁶

5. Conclusion

We have extended Harsanyi’s impartial observer argument by enriching the environment postulated by Harsanyi to include information on welfare levels of individuals. Thus, individuals are equipped both with preferences over wealth lotteries, and a conception of well-being which is interpersonally level comparable. We suggested a natural axiom, the principle of neutrality; together with Harsanyi’s principle of acceptance, we are able to determine completely the preferences over lotteries over prospects of a representative impartial observer. This enabled us to compute exactly what distribution of wealth would be recommended, from behind the veil of ignorance, by the impartial observer for the society that it represents.

⁶ It is noteworthy that such an intuition need not be based on a welfarist view; it could be that we believe for reasons other than welfare, disabled people should have more material resources than able ones. This

We showed, in a series of examples with standard sorts of preferences, that the recommended distribution contravenes what most egalitarians would recommend, namely, that *ceteris paribus*, disabled individuals should receive more transferable resources (wealth) than able ones. Here, disability means, simply, the requirement of a greater wealth increment than others to achieve a given saltus in welfare.

Although the ‘egalitarian principle’ here contravened may also be contravened by utilitarianism, it is important to note that the present argument is not utilitarian. As I noted earlier, utilitarianism is not even a well-defined concept, given the information postulated here. Just because the veil of ignorance and utilitarianism may engender similar results does not mean they are the same thing.

Neither does the environment we have specified in section 2 permit us to study some of the more interesting questions that recent political philosophy has focused upon, such as how responsibility, circumstances, and effort impinge upon the requirements of justice. We are to think of the environment here postulated as one in which these issues do not arise. If we cannot solve the problem of justice in this simple environment, how are we ever to solve it in more complex environments? (However, in section 4, where we study Dworkin’s mechanism, responsibility does come into play, as we hold persons responsible for their preferences.)

Thus, in sections 2 and 3, if a person is ‘disabled,’ it is here assumed to be through no fault of his own, or if a person has expensive tastes, it is, likewise, a characteristic for which we do not hold her responsible. Consequently, under these assumptions, most egalitarians would recommend that the ‘disabled,’ and persons with

would follow, for example, from Sen’s (1985) capability approach; see also Cohen (1989).

expensive tastes, receive more resources than the ‘able,’ and persons with cheap tastes. We have therefore illustrated what appears to be a fundamental inconsistency between egalitarianism and the recommendation following from a veil-of-ignorance construction.

Perhaps the conclusion should be that egalitarianism -- or, more weakly, the egalitarian principle that I have invoked-- is insupportable. To the contrary, I wish to propose that the veil-of-ignorance thought-experiment is not a good one.

There is a cost and a benefit to using the veil of ignorance. We often mention the benefit, seldom the cost. Truth be told, it would be better to make decisions *ex post*, that is, after we know which preference orders and welfare functions are associated with which social positions (that is, after the birth lottery has occurred). The problem with making distributional decisions *ex post* is one of maintaining objectivity: how can we be sure that the decision makers, if they are drawn from the society in question, are not simply making recommendations from self-interest? The *benefit* of the veil-of-ignorance construct is that it forces objectivity, or impartiality. But the *cost* is that we must make decisions with a great handicap – we have discarded massively important information that is available to us in the real world, namely, what the actual joint distribution of resources (here, wealth) and types *is*. The veil-of-ignorance approach asks us how we would allocate resources *if we did not know that actual distribution*. But would it not be better to think about the problem of distribution (now, redistribution) knowing what the actual distribution is, *if we could otherwise maintain impartiality*?

The answer is surely yes, because we, or the decision maker in question, would have much more information available. Maximizing expected utility, *ex ante*, is a *pis-aller*, a technique a decision maker must use, should she be required to take an action

before the state of the world has been revealed. But if that decision maker can afford to postpone her decision until the world's state has been revealed, *so much the better*. How, then, can we maintain the impartiality achieved, coercively, by the veil-of-ignorance thought experiment, but without incurring the substantial cost that accompanies it?

To be more specific, by forcing the decision maker to decide in the *ex ante* posture, we admit an element into the inquiry that has no obvious relevance to the question of distributive justice, namely, the element of a decision maker's preferences under uncertainty. I say such preferences have no obvious relevance to the question at hand, because they emerge as central *only* because of our method, that is, of constructing the veil of ignorance. It is certainly not *a priori* obvious why any individual's preferences over lotteries should influence what the ethically correct distribution of resources is. Allowing risk preferences to influence our decision about what distributive justice requires is a *cost* of the veil-of-ignorance method of inquiry – a cost, I say, we should seek to avoid, and perhaps can avoid, to bear. (In contrast, it is surely salient to allow a decision maker's preferences over lotteries influence our evaluation of whether insurance markets are efficient. For uncertainty is of the essence when insurance is at issue.)

What I have argued is that that cost is manifest in precluding us from supporting a principle – that disabled individuals should, *ceteris paribus*, receive more resources than able ones. Faced with this preclusion, we should at least question the method that led to it.

We may state an analogy with the treatment of a disease. Perhaps there is a powerful medicine that will cure the disease, but at the cost of inducing a debilitating side

effect. We may well decide not to use the medicine, and to seek another solution. I am saying that the veil of ignorance is a powerful medicine, but it requires us to sustain the side effect of employing preferences under uncertainty.

Although Dworkin's veil of ignorance is not as thick as Harsanyi's or Rawls's, his insurance mechanism is victim to the same problem that we have identified with the thick veil, namely, the possible (and likely) transfer of resources from disabled to able persons. Indeed, exactly the same critique applies to Dworkin's insurance scheme: it achieves impartiality at too high a cost.

I do not have an algorithmic answer to the question posed five paragraphs above, but the vague answer is, we must exercise our independent capacity for impartiality. As Serge Kolm (1996, p.20) has written, in a critique of original-position arguments: " ... justice is not blind-folded egoism, but open-eyed and informed objectivity." Brian Barry (1995) makes the distinction between contractarian arguments and arguments from impartiality, and I think he has the right idea. Thomas Scanlon (1998) has the right idea, when he tries to construct criteria for the admissibility of arguments that citizens can make to each other (we might call these criteria of impartiality). These authors are attempting to argue for varieties of egalitarianism without the tool of the veil of ignorance, by appealing directly to impartiality. What I hope to have shown is that this is the only way to do so.

References

- Barry, B. 1995. *Justice as Impartiality*, Oxford: Oxford University Press
- Cohen, G.A. 1989. "On the currency of egalitarian justice," *Ethics* 99, 906-944
- Dworkin, R. 1981a. "What is equality? Part 1: Equality of welfare," *Philosophy & Public Affairs* 10, 185-246
- 1981b. "What is equality? Part 2: Equality of resources," *Philosophy & Public Affairs* 10, 283-345
- Harsanyi, J. 1953. "Cardinal utility in welfare economics and the theory of risk-taking," *Journal of Political Economy* 61, 434-435
- , 1977. *Rational behavior and bargaining equilibrium in games and social situations*, New York: Cambridge University Press
- Kolm, S.C. 1996. *Modern theories of justice*, Cambridge, MA: MIT Press
- Parfit, D. 1997. "Equality and priority," *Ratio-New Series* 10, 202-221
- Rawls, J. 1971. *A theory of justice*, Cambridge, Mass.: Harvard University Press
- Roemer, J.E. 1985. "Equality of talent", *Economics and Philosophy* 1, 151-181
- ,1996. *Theories of distributive justice*, Cambridge, Mass.: Harvard University Press
- , In press. "Three egalitarian views and American law," *Law and Philosophy*
- Scanlon, T. 1998. *What we owe to each other*, Cambridge, Mass.: Harvard University Press

Sen, A. 1977. "Non-linear social welfare functions: A reply to Professor Harsanyi," in R. Butts and J. Hintikka (eds.) , *Foundational problems in the social sciences* , Dordrecht: D. Reidel

Sen, A. 1985. *Commodities and capabilities*, Amsterdam: North-Holland

Weymark, J. 1991. "A reconsideration of the Harsanyi-Sen debate on utilitarianism," in J. Elster and J. Roemer (eds.), *Interpersonal comparisons of well-being*, New York: Cambridge University Press

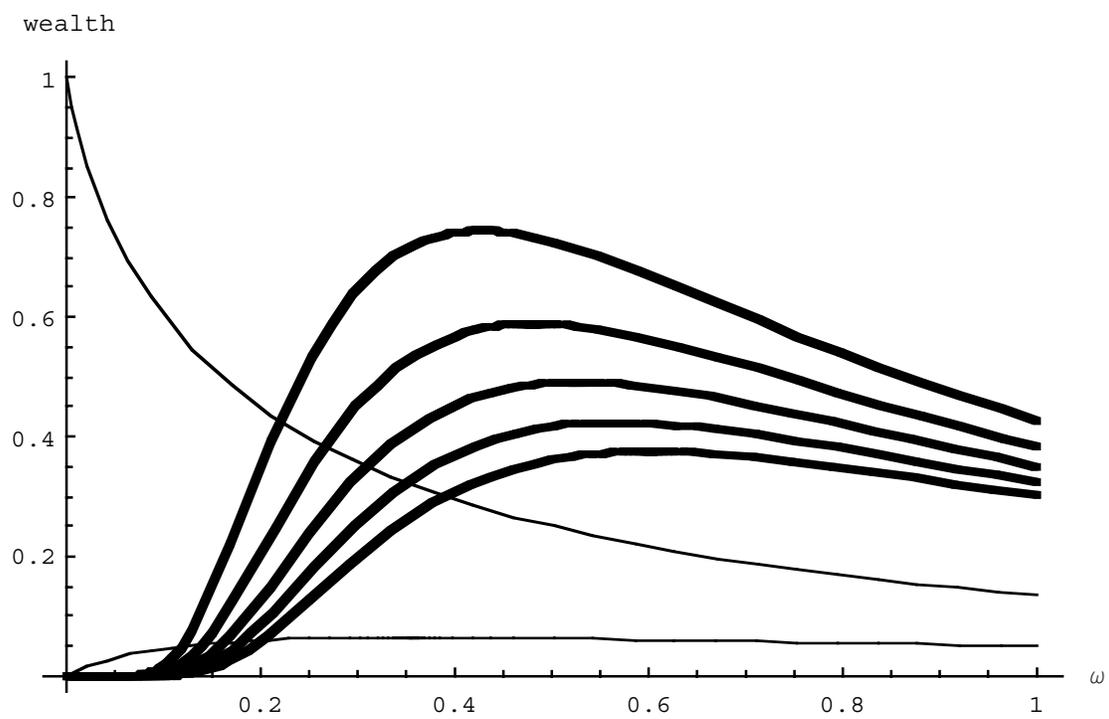


Figure 1 : Graphs of the functions of the Proposition