# Relative Fairness and Quasifairness

By Roger A. McCain

Drexel University, Philadelphia, PA

For presentation at the 2020 Conference of the Association for Social Economics, San Diego

> "In deciding principles for a just division of labor and a just division of the fruits of that labor, workers are to regard the economy as a system of cooperative, joint production. … the productivity of a worker in a specific role depends not only on her own efforts, but on other people performing their roles in the division of labor. … The comprehensiveness of the division of labor in a modern economy implies that no one produces everything, or indeed anything, they consume by their own efforts alone. In regarding the division of labor as a comprehensive system of joint production, workers and consumers regard themselves as collectively commissioning everyone else to perform their chosen role in the economy. In performing their role in an efficient division of labor, each worker is regarded as an agent for the people who consume their products and for the other workers who, in being thereby relieved from performing that role, become free to devote their talents to more productive activities."
>
> Elizabeth Anderson, 1999, pp. 321-322

To begin with the obvious, fairness has been an important topic of the philosophic literature in recent decades, with John Rawls' (1971 1993), ideas at the center. It is less well known that there is a literature on fairness (or equity) in neoclassical economics, which shows the influence, at least, of Rawls' difference principle, and that that literature has in turned influenced philosophical writing, principally through the work of Robert Dworkin (1981). Dworkin designates his view as "resource equalitarianism" and essentially adopts Varian's (1973) ideas from the neoclassical literature in order to define equal access to resources. A large literature has followed on this, much of it critical of Dworkin and focused on some troubling examples[1]. There are several criticisms of these ideas. First, the neoclassical literature uses the theory of preferences, and behavioral economics raises doubts about this. For brevity and to keep a more narrow focus, this essay will not discuss an alternative approach but will use the language

---

[1] Anderson's paper (quoted in the epigram) pointedly rejects this, and focuses particularly on the use of the term "envy" in the economics literature and Dworkin's writing (Anderson, 2000, p. 294 et seq.) Dworkin asserts that she has simply misunderstood this. (2002. p. 294, e.g.)

of preferences. Second, Rawls' difference principle may be questioned on the grounds that a change of rules that would improve the lives of people who are quite miserable – but not the most miserable – does not appear to be an improvement of fairness. Third, as Sen (2009) notes, these are all searches for a "transcendental institutional" ideal – a system entirely free of unfairness and (in the case of the neoclassical theory) of inefficiency. For many purposes, it would be useful to have a relative measure of fairness – one that would allow us to consider two social situations and say that one is more fair than the other, even though neither is wholly free of unfairness. Such a criterion would speak to both the second and third criticisms, and is the objective of this paper.

To begin, as a basis for contrast, the neoclassical literature may be quickly summarized as follows. Since this is neoclassical economics, the objects to be evaluated are market baskets of goods and services, often with job assignments and hours also specified, and the only basis of evaluation is the individual preferences over the objects, which may differ from one person to another. Suppose, then, that John prefers the market basket that Irving possesses to the one that John himself possess, then it is said that John "envies" Irving, and a fair or equitable allocation is one that is envy-free. For this study, the objects for comparison will be stable "social situations." A social situation[2] is essentially a set of coordinated roles as suggested in the epigram of the paper, together with the individual outcomes of the roles. These outcomes may include market baskets of goods and services, job assignments and hours, but also other conditions that make life worth living or onerous, such as quantities of public goods provided, the joys of family life, honors, respect and self-respect, being on the winning or the losing side in conflicts, and so on. Nevertheless individual preferences are assumed to be defined over the outcomes. The stability of a social situation corresponds to one or another game-theoretic solution concepts and will not be explicitly considered here. In what follows, the first section will outline concepts of relative fairness and quasifairness in the comparison of social situations; the second will offer an argument for their representation of fairness based on an adaptation of the veil of ignorance, and the third will argue that relative quasifairness, in particular, addresses what has been an unsolved problem: intergenerational fairness.

[2] This term is borrowed from Joseph Greenberg, The Theory of Social Situations (1990), a restatement of game theory; but is used in a somewhat different and less formal sense.

## i. Fairness and Quasifairness

At a minimum, in order to discuss fairness, we must suppose that individual outcomes can be evaluated in terms that are correlated with individual well-being. For brevity and familiarity, the language of preference will be used, but no assumption is made that market choices reveal preferences in this relevant sense. Some terminology and notation are unavoidable, and the language of set theory will be used. We have a population of agents $\mathcal{A} = \{1,2,\dots,m\}$ who are both decision-makers and actors. It is assumed that $\mathcal{A}$ is countable and, unless stated otherwise, finite. For this study, the joint outcome of the joint action of the group is a vector, a set of individual outcomes, $\{q_1, q_2, \dots, q_m\}$. The individual outcomes $q_i$ are drawn from an outcome set $Q=\{q^1, q^2, \dots, q^r\}$, and for each agent i we posit a preference system by which each agent i can evaluate $q^1$ as preferable to $q^2$, $q^2$ as preferable to $q^1$, or the two as indifferent choices. We then may write $q^1 \mathcal{p}_i q^2$, $q^2 \mathcal{p}_i q^1$, or $\sim q^1 \mathcal{p}_i q^2$ and $\sim q^2 \mathcal{p}_i q^1$. Notice that here $\mathcal{p}_i$ denotes strict preference, "better than," whereas preference in neoclassical economics would be interpreted as "no worse than." Similarly, it will be sufficient to suppose that preference in this sense is quasitransitive, not necessarily transitive.

Fairness is a condition on a particular feasible outcome vector. If we follow the fairness literature in economics, we might identify the following scenario as unfair: given $i \epsilon \mathcal{A}, j \epsilon \mathcal{A}$, if $q_j \mathcal{p}_i q_i$, then the situation is unfair to the disadvantage of i. But this is consistent with $q_i \mathcal{p}_j q_j$, i.e. each person would choose the outcome the other has if it were possible. In some cases, where it is possible, a simple exchange may resolve this, but not always[3]. Accordingly, this study will identify unfairness only with a case in which the disadvantage is expressed in terms of the rational choices of both individuals. Accordingly, suppose that, given two individuals i and j and their individual outcomes $q_i$ and $q_j$, $q_j \mathcal{p}_i q_i$ and $q_j \mathcal{p}_j q_i$. Denote this relationship as $i\mathcal{L}j$. The letter $\mathcal{L}$ stands for "lags;" that is, i lags behind j in a choice hierarchy. Put otherwise, j is strictly better off than i according to the preferences of both of them. We may then say that Q is fair if it is feasible and for any two agents i and j neither $i\mathcal{L}j$ nor $j\mathcal{L}i$. In other words, Q is fair if it is free of lag relationships, that is, if no one agent lags another in a choice hierarchy. This is, of course, a very demanding concept of fairness. In the economics literature it is established that, because of

[3] For a simple example involving ants and grasshoppers, see McCain 2017 pp. 37-40.

cycles such as $i\mathcal{L}j\mathcal{L}k\mathcal{L}i$, fair outcomes may not exist. This agnostic result can be extended to fairness in the present sense: cyclical preferences (such as Arrow uses in his impossibility theorem) can produce a lag cycle. Another instance in which fair outcomes may not exist arises if $\mathcal{A}$ is countably infinite, as it might be if agents belong to a succession of generations.

An evaluation of relative fairness would compare two social situations $S^1$ and $S^2$ with their respective outcomes $q_i^1, q_j^1, q_i^2, q_j^2$. In what follows, if $i\mathcal{L}j$ at situation $S^\iota$, write $i\mathcal{L}^\iota j$. Following roughly the path blazed by Rawls, we consider those who are least advantaged in each case. Thus define a subset $\mathcal{N}^\iota \subset \mathcal{A}$ such that, at situation ι, the members of $\mathcal{N}^\iota$ lag some other agents but are not lagged. That is,

1. $\mathcal{N}^\iota = \{i \in \mathcal{A} \ni \forall j \in \mathcal{A}, j \neq i \Longrightarrow\ \sim j\mathcal{L}^\iota i \text{ and } \exists j \in \mathcal{A} \ni\ i\mathcal{L}^\iota j\ \}$.

Notice that, in a case of a lag cycle or an infinite succession of groups of agents, $\mathcal{N}^\iota$ may be a null set. Now consider social situations $S^1$ and $S^2$, and suppose that these situations satisfy the following three conditions:

2a. $i \in \mathcal{N}^1 \Longrightarrow \sim q_i^1 \wp_i q_i^2$

2b. $\exists i \in \mathcal{N}^1 \ni q_i^2 \wp_i q_i^1$

2c. $\forall i \in \mathcal{A}, \forall j \in \mathcal{A}, i\mathcal{L}^2 j \Longrightarrow\ i\mathcal{L}^1 j$

Of these three conditions, 2a. and 2b. specify that the transition from $S^1$ to $S^2$ is a Pareto-improvement for the members of $\mathcal{N}^1$, although it may not be a Pareto-improvement in the population as a whole. Condition 2c. specifies that the transition from $S^1$ to $S^2$ creates no new lag relations. For example, no member of $\mathcal{N}^1$ is so much enriched that they are lagged by someone in situation $S^2$. If these conditions are met, then we may say that $S^2$ is more fair than $S^1$. That is, the transition from $S^1$ to $S^2$ benefits the less favored, i.e. it benefits some of them and makes none worse off.

Suppose instead that we define a subset $\mathcal{M}^\iota \subset \mathcal{A}$ such that, at situation ι, the members of $\mathcal{M}^\iota$ lag some other agents but may or may not be lagged. In addition $\mathcal{M}^\iota$ will include isolates who neither lag nor are lagged by anybody. That is,

3. $\mathcal{M}^{\iota} = \{i \in \mathcal{A} \ni \exists j \in \mathcal{A}, j \neq i \ni \ i\mathcal{L}^{\iota}j\} \cup$

$$\{i \in \mathcal{A} \ni \forall j \in \mathcal{A}, j \neq i \Longrightarrow \ \sim j\mathcal{L}^{\iota}i \textit{ and} \sim i\mathcal{L}^{\iota}j\}.$$

Notice that this set might comprise the entire population in a case of a lag cycle or in a case of an infinite succession of groups of agents. Notice further that $\mathcal{N}^{\iota} \subset \mathcal{M}^{\iota}$. Again consider social situations $S^1$ and $S^2$ that satisfy the following three conditions:

4a. $i \in \mathcal{M}^1 \Longrightarrow \sim q_i^1 \wp_i q_i^2$

4b. $\exists i \in \mathcal{M}^1 \ni q_i^2 \wp_i q_i^1$

4c. $\forall i \in \mathcal{A}, \forall j \in \mathcal{A}, i\mathcal{L}^2 j \Longrightarrow \ i\mathcal{L}^1 j$

Again, conditions 4a. and 4b. specify that the shift from $S^1$ to $S^2$ is s Pareto-improvement among the members of $\mathcal{M}^1$, and 4c. is the same as 2c. above. Then we may say that situation $S^2$ is more quasifair than $S^1$. Quasifairness may seem to be the weaker condition, but in fact examples may be presented in which $S^1$ is more fair than $S^2$ but not more quasifair, and conversely.

It may seem that quasifairness is a very weak criterion, as a concept of fairness. After all, a change in the social situation that would make the second-richest person richer would be an increase in quasifairness if it did not reverse the ordering between the richest and second-richest and did not make anyone else worse off. “De minimus non curate lex:” is fairness at all concerned with relations among the very richest? To address this question it will be helpful to digress briefly from the main topic of this section and consider what a social situation might look like that is absolutely quasifair. For clarity, suppose that lump sum money transfers are possible. That is, there is an asset that has four properties: 1) it is a dimension of any outcome for any agent; 2) it is continuously divisible; 3) any agent, choosing between two outcomes that differ only in the magnitude of the asset holding will prefer the outcome with the larger asset holding; 4) a sufficiently small transfer of this asset from j to i will make i better off without creating any new lags. Then suppose $i\mathcal{L}^{\iota}j$. It follows that $i \in \mathcal{M}^{\iota}$. Thus, such a transfer from j to i is an increase in quasifairness. If we then envision a situation in which quasifairness cannot be improved, it would mean there are no i, j in the population such that $i\mathcal{L}^{\iota}j$. But this is precisely the condition for a perfectly fair situation. If then a situation is perfectly quasifair then it must also be perfectly fair.

The significance of the lump sum money transfers is that they assure us that a situation exists that is perfectly fair and quasifair and can be approached from any beginning point. Perhaps we might identify as *very quasifair* a situation from which there is no feasible shift to a situation that makes i better off, for any $i \in \mathcal{M}^{\iota}$, without creating new lags. A parallel definition would identify as *very fair* a situation from which there is no feasible shift to a situation that makes i better off, for any $i \in \mathcal{N}^{\iota}$, without creating new lags. Since $\mathcal{N}^{\iota} \subseteq \mathcal{M}^{\iota}$, a situation that is very quasifair must be very fair; but if $\mathcal{N}^{\iota} \subset \mathcal{M}^{\iota}$, , we might have a situation that is very fair but not very quasifair. In this context quasifairness is the stronger condition.

Returning to relative fairness, we might consider a hierarchical rule:

5a. If $S^2$ is more fair than $S^1$, do not choose $S^1$ when $S^2$ is available.

5b. If rule 5a. is not applicable, but $S^2$ is more quasifair than $S^1$, do not choose $S^1$ when $S^2$ is available.

5c. If neither 5a. or 5b. is applicable, but $S^2$ Pareto-dominates $S^1$ and moreover rule 2c-4c is fulfilled, do not choose $S^1$ when $S^2$ is available.

In this section, we have restated and modified a concept of fairness from "ordinalist" economics. Consistently with the "first-person or continuity test,"[4] that distributive judgments "should track people's own assessment of their relative standing," the relation between Irving and John is judged to be unfair to Irving if both Irving and John would prefer the outcome that John enjoys over that available to Irving. In such a case Irving is said to lag John, and in an absolutely fair situation, no one person would lag another. Relations of relative fairness and quasifairness are then defined. A transition between social situations increases fairness if it improves the situation of some of those who are least favored in the sense that they are lagged by no one, and makes none of them worse off. A transition between social situations increases quasifairness if it improves the situation of some of those who lag others, and makes none of them worse off. These rules may be combined in a hierarchical compound rule that applies the rule of quasifairness if fairness is not applicable and of Pareto-dominance modified by a preliminary fairness rule if the transition cannot be ranked either as increasing fairness or quasifairness. This hierarchical rule gives clear precedence to fairness over efficiency, is quasi-transitive[5], and can be justified by a veil of ignorance argument.

---

[4] Hansen and Midtgaard, 2011, p. 345. The term "continuity test" seems to originate with Williams, who in turn attributes the test to Dworkin: Williams, 2002a, b.

[5] Proof will be beyond the scope of this paper. The proposition has been proved in a slightly more general form.

ii. The Veil of Ignorance

We now reconsider the rules of relative fairness and efficiency developed in the previous section in a broadly Rawlsian way. Rawls relies on a simplification that will not be available here. For Rawls there is an objective, interpersonal criterion to determine the least advantaged group in the population. The objective criterion is access to primary goods, with liberty first among them. This essay, however, follows the ordinalist tradition in that it is not supposed that there is any objective or interpersonally comparable criterion of individual well-being. For agents in this discussion the criterion of individual well-being is subjective and might differ from person to person. Suppose, however, that j $\mathcal{L}^1$ i at $S^1$. Then i is aware that j will be less advantaged than i at $S^1$ according both to i's own subjective criteria and to j's.

Instead of an amorphous deliberation, for this essay the decision behind the veil of ignorance is the play of a metagame. The metagame is patterned after and extends the cake-cutting game of fair division (Dubins and Spanier 1961; Kuhn 1967). The outcome of the metagame is a choice among two or more outcome vectors corresponding to distinct social situations. In any case, the decision is made on behalf of a particular population, so that $\mathcal{A}$ is the same for each of the primitive games considered in the metagame, and the number of participants in the metagame is $|\mathcal{A}|$. At the first stage of the metagame, the deliberators each record their preferences between the outcome vectors under consideration. At the second stage, each member is randomly assigned a place in a sequential ordering[6]. At the third stage, in the randomly assigned order, the decision-makers choose the roles they will play in each of the various primitive games under consideration. At the last stage, the preference between outcome vectors previously recorded by the last decision-maker in the random order is recovered, and the outcome vector preferred by that decision-maker is the one chosen. Thus, anticipating that their

[6] In the case of an infinite population, the positions assigned would be last, second to last, third to last, etc, so that there is a last chooser but no first chooser. What is necessary for a defined result is that there be an agent who chooses last. For an infinite population, each agent would be aware that an infinite number of agents will have chosen before they do, so that in effect each agent is the last agent.

expressed preference will influence the collective decision only if they are last in the random ordering, the decision-makers will unanimously choose the outcome vector on the supposition that they will be last.

The unanimity at the first stage satisfies a condition for the Nash equilibrium to be considered a social contract, if the primitive games considered are alternative social systems. Since the decision-makers at the first stage are unaware of the roles they will play in the game chosen, the decision cannot rationally display any bias in favor of any one role or another. However, we might say that the decision is biased toward fairness. Suppose that the metagame were modified in one way: at the last stage, instead of taking the decision of the last-ranked player as decisive, we choose the decisive preference at random from the $|\mathcal{A}|$ expressed preferences with equal probabilities. Then the recording of preferences at the first stage would rationally be based on some probabilistic balancing of the possible outcomes and roles. If then we assume also that outcomes are numerical utility indices, we could expect a result more along the lines of Harsanyi's (1955) theory. Again, the decision at the first stage would be unanimous, since all would be "maximizing the same expected value" (to quote a colleague in an informal discussion.) But since we have assumed instead (in effect) that the last one chooses his slice after the others have cut, each of the others has an incentive to make the slices as equal as possible.

Since each expresses a preference at the first stage that best serves their own interest, the decision is a noncooperative solution of the metagame. Tentatively we will treat it as a Nash equilibrium of the metagame.

Now consider a decision in which $S^2$ is more fair than $S^1$. Here is a possible reasoned response behind the veil of ignorance:

> I will not be instantiated as a member of $\mathcal{A}/\mathcal{N}^1$, since in that case my well-being could be reduced: that is, there would be j who lags me, and my well-being could be reduced if it were replaced by that of j, so that some earlier chooser will choose to be instantiated as i. Thus, I will be instantiated as a member of $\mathcal{N}^1$, and since the shift to $S^2$ is a Pareto-improvement for $\mathcal{N}^1$, I cannot be worse off for the shift. But, as last to choose, neither will I be better off at $S^2$. Thus, while I have no reasonable objection to the shift, I have no basis for choice between these two situations.

This is a bit discouraging! However, thinking of this as a Nash equilibrium, it is a weak equilibrium. That is, while our last chooser has nothing to gain by choosing $S^2$ over $S^1$, he has nothing to lose. We have, then, two weak Nash equilibria: one in which $S^1$ is chosen and one in which $S^2$ is chosen. Suppose, however, that our last chooser refines his decision by applying the "trembling hand" refinement. In effect, then, he reasons:

> Considering the specific $i \in \mathcal{N}^1$ who is made better off at $S^2$, it seems that, if others choose rationally, identity i will already be chosen by the time I make my choice; but there may be some small probability that others may choose irrationally, leaving identity i available to me when I choose. Taking this small probability into account, I am better off with $S^2$ than with $S^1$, so my choice is for $S^2$.

It seems, then, that a rule that requires us to choose the fairer of two social situations can be rationalized by reasoning behind a veil of ignorance, provided we accept this instance of "trembling hand stability" as rational. This is not clear beyond question. It requires us to admit probabilistic reasoning without such things as mathematical expectations, since outcomes in this context may be too complicated (or too simple!) to admit of multiplication and addition. And, conversely, the trembling hand criterion is usually expressed as a comparison of mathematical expectations of payoffs. Instead the form of the argument – "there are two contingencies, one in which I am better off and the other in which I am no worse off, and both have positive probabilities, so balancing the two I am better off" – is non-numerical. But the non-numerical argument is more fundamental. If we reject it as false, then the mathematical expectations, even when they can be computed, are meaningless.

The trembling hand refinement is well known in game theory as a means of resolving cases in which two or more weak equilibria exist. In effect it adds a failsafe clause to the common assumption in game theory that all players know that all other players will play rationally. In this refinement, we suppose that there is some probability p that another player will deviate from the rational strategy. Then p is allowed to approach zero in the limit, and if for sufficiently small p situation $S^1$ is always chosen, then $S^1$ is "trembling hand stable." For the argument made here it will generally be necessary that more than one person deviate from the rational choice – that in fact all members of $\mathcal{N}^1$ other than the last chooser fail to choose the advantaged position. If $|\mathcal{N}^1|$=N, then the compound probability is $p^{N-1}$; but so long as p>0, it follows that $p^{N-1}>0$, so the decision for $S^2$ is trembling-hand stable.

There is, however, a difficulty, and it is this: a precisely similar argument can be made to choose the situation that is more quasifair, or that is a Pareto-improvement even if it does not fall under the fairness rules. And we have seen that these can give conflicting answers.

Suppose, then, that the deliberators behind the veil of ignorance are asked to decide between Rules 1 and 2, in a case where they conflict. Note that $\mathcal{N}^1 \subseteq \mathcal{M}^1$, and if $\mathcal{N}^1$=$\mathcal{M}^1$, there can be no conflict. Therefore $\mathcal{N}^1 \subset \mathcal{M}^1$, and with $|\mathcal{N}^1|$=N, $|\mathcal{M}^1|$=M, M>N and, for any given probability p, $p^{N-1} > p^{M-1}$. Then the deliberator behind the veil of ignorance might reason as follows:

> Both rules 1 and 2 are applicable in this case, and I can have no objection to either of them, since neither will leave me worse off. Further, allowing for some probability of a failure of others to choose rationally, in each case there is some probability that I will benefit; but since $p^{N-1} > p^{M-1}$, the probability is always greater in case Rule 1 is applied, so my decision is for Rule 1.

A similar argument can be made for Rule 2 over Rule 3. This leads us to hierarchical rule. To summarize, a change according to the Pareto rule might be justified only if the change meets a fairness test, with fairness taking first priority and quasifairness taking second priority, and with Pareto-dominance third.

### iii. Intergenerational Fairness

In *A Theory of Justice*, Rawls has only a brief passage on fair relations among generations, about what he calls "just saving," but it is nevertheless something of a struggle. The difficulty is a conflict with the difference principle. In a progressive society, earlier generations will be less advantaged than later ones. On the one hand, it is not possible for the present generation to make transfers to earlier generations; on the other hand, whatever is saved and invested today is transferred to generations who are likely, on the whole, to be better off than the generation that makes the transfer. Rawls treats the deliberators behind the veil of ignorance as the current generation and trusts their regard for their posterity to determine what is to be set aside for the benefit of future generations. In *Political Liberalism*, however, he rejects that discussion as "defective" (p. 20 fn. 22.) Instead, the deliberators behind the veil of ignorance "who are assumed to be contemporaries, do not know the present state of society. They have no information about the natural resources or productive assets, or the level of technology. … all questions of justice are dealt with by constraints that apply to contemporaries. Consider the case of just savings: … the correct principle is that which the members of any generation … would adopt as the principle their generation would want preceding generations to follow (and later generations to follow…)" (pp. 273-4).

A literature along these lines has long existed in economics. This literature stems from Phelps, 1961. However, this "golden rule of saving" literature is utilitarian. Indeed, it is hard to see how Rawls' later principle of just saving can be applied without something like a cardinal utilitarian measure of wellbeing, since the comparison of wellbeing of different generations seems needed for the agents behind the veil of ignorance to judge among different saving plans. Moreover, Rawls' assumption that all deliberators are of the same generation (A Theory of Justice p. 140, Political Liberalism p. 273, note also Paden 1997, p. 38, Finneron-Burns 2017 pp. 808-9) – although they do not know which – seems to treat generations as separate. But generations overlap; so, if the agents behind the veil of ignorance are representative of the whole population (as they must be if we are investigating *fairness*) they will be of more than one generation. In economics there is a large literature on the implications of overlapping generations, some of which may be surprising. See Economist (2017) for a brief ordinary-language discussion and note also Samuelson 1958, Weil 2008, Barrell and Weal 2010, among many others. Further, on the one hand, overlapping generations present behind the veil of ignorance provide a motivation for transfers to a coming generation, which are otherwise problematic. They also partly relax the constraint that transfers may not be made to earlier generations. Rather, old-age pensions become possible, and may (or may not) be regarded as fair. Overlooking this possibility causes the theory of justice as fairness to stand mute on a crucial issue of modern economics and policy.

Rawls' just saving, in both versions, has been extensively criticized in the literature of philosophy. An issue common to many critics is the seeming contradiction between just saving and Rawls' difference principle. (See e.g. Wall, 2003.) In Rawls' earlier formulation, regard for posterity is an other-regarding preference that is at odds with the rational egoism Rawls assumes of agents in the original position. Rawls supposes that rational egoists behind the veil of ignorance would save as little as possible (Finneron-Burns, pp. 806-810.) Substantially, Rawls here (Justice as Fairness, pp. 140, 291-2) is describing a Nash equilibrium among the contemporary deliberators at the original position. Since deceased generations cannot agree to anything, in *A Theory of Justice*, Rawls sees the Nash equilibrium as unavoidable among purely self-interested agents. In *Political Liberalism*, however, he adopts what Wall calls prioritarianism in intergenerational allocation (Wall, 2003, pp. 88-90) or what MacClellan (p. 74) calls the universalizability principle. Rawls' latter approach is generally considered less problematic, but critics regard it as somewhat ad-hoc (MacClellan p. 787, Wall 90-94, Finneron-Burns 814-5, Paden, 1997, p. 41.) Further, the difference principle and the just saving principle might conflict (Paden, p. 29, Wall, p. 80). If the deliberators at the original position choose a prioritarian rule for saving, why not also (as Wall argues) for distribution among contemporaries? The previous section has put forward the view that (something like) the difference principle can be identified as a principle of rational fairness in that such a principle extends the cake-cutting models from game theory. (Note also Nelson, 1980, p. 504, Kohlberg, 1973, p. 642, Goodman, 1991, p.17.) The principle of just saving, as we have it in *Political Liberalism*, does not share the same ground to identify it with fairness[7]. If both intergenerational and distributive judgements are based on prioritarian reasoning, we would obtain something closer to Harsanyi's utilitarianism.

[7] Note Rawls, Political Liberalism, pp. 71-72. Here Rawls contrasts his view with cake-cutting models on the argument that the cake-cutting models presuppose "an independent and already given criterion" of what is just (or fair), while his "rational autonomy" does not. But this applies only to the cake-cutting example in its simplest form, with a uniform cake that defines a one-commodity world. In a one-commodity world, equal division seems prima facie fair. But consider instead a cake with raisins that are not uniformly distributed, and two children who differ in their taste for raisins. Then equal division of the cake is not so clearly fair – it does not respect the preference of one cake-eater for raisins. Thus "whatever principles the parties select … are accepted as just." (p. 72) The fairness of equal division in a one-commodity case is not presupposed but verified by the cake-cutting model. On the other hand, of course, Rawls makes different use of fairness than this work does. Rawls' concern is with the nature of justice, interpreting "justice as fairness," to quote a title he used more than once. Our concern here is with fairness as a complement or constraint on efficiency. Either issue requires some "model" of fairness, and it is here that Rawls' ideas are no less important for the purposes of this work.

Agents behind the veil of ignorance could choose a plan that is Pareto-preferable over others, despite its unfairness. (Wall, p. 82.)

Some other grounds for criticism of Rawls need not detain us much here. A large literature in economics analyzes saving decisions as an intergenerational max min problem (Arrow 1973, e.g.), but since Rawls never adopts that position, it will not be discussed in detail here. (Arrow admits that he is ignoring the “richness” of Rawls’ writing, p. 323.) It is also argued that decisions on intergenerational allocations could not be made by the deliberators behind the veil of ignorance because past decisions made them what they are, so that choosing a different plan for earlier generations would be choosing non-existence, and future generations cannot participate because their identity depends on decisions to be made (Finneran-Burns, pp. 815-16.) If I understand it, this view misunderstands the character of deliberators at the original position as *representative* agents of their population, or in Rawls’ terminology from Political Liberalism, of the initial position as a device of representation.

Dismissing these latter points, and despite Rawls’ shift of position, it seems fair to observe that intergenerational fairness is an unsolved problem, at least so long as we interpret “fairness” in a way consistent with the difference principle. In part, this may reflect the transcendental-institutionalist character of the answers sought. Those who are born earlier in a progressive society are, on the whole, less well off: this is an unfairness that cannot be eliminated. The oldest living generation might be subsidized by our living juniors to the extent that we are not unavoidably worse off than they; but even if this is done, Rawls is right: there is nothing we can do for the dead. Thus, along with Rawls, we might suppose that fairness is inappropriate to intergenerational allocation. However, quasifairness can be applied.

With Rawls we undertake to treat the ongoing progressive society as “a system of cooperation over generations over time.” (Political Liberalism, p. 274.) We may consider what would be a cooperative solution for all generations. Of course, no such coalition could literally be formed. Still, the coalition of all generations is no less literally possible than Rawls’ initial position is, and as with the original position, we can ask what conclusion would be drawn if we reason *as if* finding such a cooperative solution. But aside from the impossibility of assembling a coalition among all those living, dead and not yet born, a problem of summation arises. The most common approach in economics would be to maximize the discounted present value of the sum of future utilities. This Bernoulli sum may be finite. But discounting the utilities of future generations is a questionable procedure. Time preference may correspond to a rate of discount for future periods of the life of one generation, but the first-person principle demands that the well-being of each generation be evaluated in terms of its own preference, not those of earlier (or later) generations. Further, this Bernoulli sum requires a starting point, an "original" generation.

Suppose, instead, that a criterion of fairness is applied to the intergenerational sequence. First, since for every generation j, j-1$\mathcal{L}$j, the set $\mathcal{N}$ is null. It follows that the criterion of relative fairness, as defined above, cannot be applied. Instead, consider quasifairness. By the same token, the set $\mathcal{M}$ comprises the entire population of all generations. Further, in a representative agent model the only lags are intergenerational, and these remain unchanged in the comparison of one growth path to another. Thus, a Pareto improvement for the entire population is at the same time an increase in quasifairness. The Pareto-optimum cannot be improved on with respect either to fairness or quasifairness. Far from a conflict, in this application, the conditions of quasifairness and efficiency coincide.

## iv. Concluding Summary

To say that a social arrangement, allocation or system of institutions is "fair" is prima facie to say that it is not biased in the results experienced by those who participate in it. Rawls' "representative device," the veil of ignorance, provides us with a representation of that absence of bias. However, concepts of fairness that arise from this approach tend to be "transcendental institutional" or absolute criteria, and the difference principle, in particular, seems to fail if a least favored group cannot be identified. This leaves intergenerational fairness, in particular, mysterious. This paper first defines two concepts of relative fairness and relative quasifairness, building on the neoclassical model of equitable allocations. A hierarchical rule for choosing between alternative social situations that gives fairness priority over quasifairness and quasifairness over efficiency. In order to assess the fairness of this hierarchical rule, a metagame is outlined that 1) extends the cake-cutting model of fairness, 2) requires agents to choose among rules behind a relevant veil of ignorance, and 3) treats the trembling-hand stable Nash equilibrium of the metagame as the rational solution behind the veil of ignorance. A max min criterion is that rational solution (though a slightly different metagame would give rise to a utilitarian solution as in Harsanyi's writing.) The hierarchical rule is supported as a trembling hand solution to the metagame and thus objectively fair. In application to intergenerational fairness, we find that the criterion of relative fairness cannot be applied, for the familiar reason that a least favored group cannot be identified. However, relative quasifairness can be applied, and in this case there is no conflict between quasifairness and efficiency. The concept of quasifairness seems to be new, and in resolving the unsolved problem of intergenerational fairness, useful.

References


Anderson, Elizabeth S (1999), What Is the Point of Equality? Ethics v. 109, no. 2 (Jan) pp. 287-337.

Arrow, Kenneth J (1973), "Rawls's Principle of Just Saving," The Swedish Journal of Economics v. 75, no. 4 (Dec) pp. 323-335.

Barrell, Ray and Martin Weale (2010), "Fiscal policy, fairness between generations, and national saving," Oxford Review of Economic Policy v. 26, no. 1 (Spring) pp. 87-116.

Dubins, L. E. and E. H. Spanier (1961), "How to Cut a Cake Fairly," American Mathematical Monthly v. 68, (Jan) pp. 1-17.

Dworkin, Ronald (1981), "What is Equality? Part 2: Equality of Resources," Philosophy & Public Affairs v. 10, no. 4 (Autumn) pp. 283-345.

Dworkin, Ronald (2002), Sovereign Virtue Revisited, Ethics: Symposium on Ronald Dworkin's Sovereign Virtue - Introduction v. 113, no. 1 (Oct) pp. 106-143.

Economist, (2017), "Economics Brief: Overlapping generations: Kicking the can down an endless road." (Aug 31)

Finneron-Burns, Elizabeth (2017), "The Intergenerational Original Position," Social Theory and Practice v. 43, no. 4 (Oct) pp. 805-823.

Goodman, L. E. (1991), On Justice: An Essay in Jewish Philosophy (Yale University Press).

Greenberg, Joseph (1990), The Theory of Social Situations: An Alternative Game-Theoretic Approach (Cambridge University Press).

Hansen, R. S and S. F. Midtgaard, (2011) "Sinking Cohen's Flagship — or Why People with Expensive Tastes Should not be Compensated," Journal of Applied Philosophy, v. 28, no. 4, (Nov.) pp. 341-354.

Harsanyi, John (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," Journal of Political Economy v. 63, (Aug) pp. 309-321.

Kohlberg, Lawrence (1973), "The Claim to Moral Adequacy of a Highest Stage of Moral Judgment," The Journal of Philosophy, v. 70, no. 18 (Oct) pp. 630-646.

Kuhn, H. W. (1967), "On Games of Fair Division," Essays in Mathematical Economics in Honor of Oskar Morgenstern edited by M. Shubik (Princeton) pp. 29-37.

MacClellan, Joel (2013), "How (Not) to Defend a Rawlsian Approach to Intergenerational Ethics," Ethics and the Environment v. 18, no. 1 (Spring) pp. 67-85.

McCain, Roger A. (2017), Approaching Equality: What Can Be Done About Wealth Inequality (Cheltenham: Edward Elgar).

Nelson, William (1980), "The Very Idea of Pure Procedural Justice," Ethics v. 90, no. 4 (Jul) pp. 502-511.

Paden, Roger (1997), "Rawls's Just Savings Principle and the Sense of Justice," Social Theory and Practice, v. 23, no. 1 (Spring) pp. 27-51.

Paden, Roger (1996), "Reciprocity and Intergenerational Justice," Public Affairs Quarterly v. 10, no. 3 (Jul) pp. 249-266.

Phelps, Edmund (1961), "The Golden Rule of Accumulation: A Fable for Growthmen," The American Economic Review v. 51, no. 4 (Sept) pp. 638-643.

Rawls, John (1971), A Theory of Justice (Cambridge: Belknap Press).

Rawls, John (1993), Political Liberalism (Columbia University Press).

Samuelson, Paul A. (1958), "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," Journal of Political Economy v. 66, no. 6 (Dec.) pp. 467-482.

Sen, Amartya (2009), The Idea of Justice (Cambridge, Mass: The Belknap Press).

Varian, H. R. (1975), "Distributive Justice, Welfare Economics, and the Theory of Fairness," Philosophy and Public Affairs v. 4, pp. 223-247.

Wall, Steven (2003), "Just Savings and the Difference Principle," Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition v. 116, no. 1 (Oct) pp. 79-102.

Williams, Andrew, (2002a) "Equality for the Ambitious: review of Sovereign Virtue: The Theory and Practice of Equality by R. M.Dworkin," The Philosophical Quarterly, Vol. 52, No. 208 (July), pp. 377-389.

Williams, Andrew, (2002b) “Dworkin on Capability,” Ethics, Vol. 113, No. 1, Symposium on Ronald Dworkin's Sovereign Virtue (October), pp. 23-39.