ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/24103722

Collective Labor Supply and Welfare

Article *in* Journal of Political Economy · June 1992

DOI: 10.1086/261825 · Source: RePEc

CITATIONS
1,082

READS
560

1 author:

Pierre Andre Chiappori
Columbia University
169 PUBLICATIONS **9,132** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Family decision making View project

Behavioral finance and investment decision View project

All content following this page was uploaded by Pierre Andre Chiappori on 20 October 2014.

The user has requested enhancement of the downloaded file.

Collective Labor Supply and Welfare

Pierre-André Chiappori

Département et Laboratoire d'Économie Theorique et Appliquée

The paper develops a general, "collective" model of household labor supply in which agents are characterized by their own (possibly altruistic) preferences, and household decisions are only assumed to be Pareto efficient. An alternative interpretation is that there are two stages in the internal decision process: agents first share nonlabor income, according to some given sharing rule; then each one optimally chooses his or her own labor supply and consumption. This setting is shown to generate testable restrictions on labor supplies. Moreover, the observation of labor supply behavior is sufficient for recovering individual preferences and the sharing rule (up to a constant). Finally, the traditional tools of welfare analysis can be adapted to the new setting.

I. Introduction

The empirical analysis of joint labor supply decisions within a two-member household has received widespread attention in recent years. The theoretical models that underlie these works can be classified into two broad categories. Traditionally, the household, as a whole, is considered as the elementary decision unit; in particular, it is characterized by a *unique* utility function that is maximized under a budget constraint. On the other hand, several "collective" models have been proposed recently that are aimed at explicitly taking into account the individualistic elements of the situation; typically, they represent

This paper is part of a project supported by the Ministère de la Recherche (contract 86 J 0866). It has been presented in conferences in London and New York. I am indebted to Richard Blundell, François Bourguignon, Martin Browning, Christopher Flinn, Thierry Magnac, Robert Moffitt, Javier Ruiz-Castillo, Ian Walker, and the editor and two anonymous referees for useful comments. Errors are mine.

[*Journal of Political Economy*, 1992, vol. 100, no. 3]
© 1992 by The University of Chicago. All rights reserved. 0022-3808/92/0003-0003$01.50

Copyright © 1992. All rights reserved.

households by a pair of individual utility functions, together with a particular decision rule.

This alternative strand of the literature, however, does not seem to have yet developed a common framework; nor has it produced so far general restrictions on household behavior that would allow empirical tests to be performed (in the same way that Slutsky equations do in the traditional setting). The first purpose of this paper is to partially fill this gap. I propose a general approach for the collective line of investigation that relies on an axiomatic characterization of the decision rule. The main properties of the approach are derived for a standard labor supply model, with two labor supplies and a unique consumption good. The conditions imposed by this characterization are quite general. Indeed, it is only assumed that agents are either egoistic or "caring" in the Beckerian sense and that internal decision processes are cooperative, in the sense that they systematically lead to Pareto-efficient outcomes. This formalization encompasses not only some essential ideas of Becker's early contributions but also many of the collective approaches existing so far.

The basic result, from this point of view, is that it is possible to derive from the collective setting thus defined a set of testable restrictions on observable behavior (labor supply in our model) under the form of partial differential equations. These equations can be viewed as analogous, in the collective setting, to Slutsky relations in the traditional model. In particular, they can be used, for empirical purposes, as an alternative way of deriving additional structural conditions on functional forms for demand or labor supply functions. They can be translated into restrictions on parameters, which may in turn either be tested statistically or be used a priori for reducing the estimation task. Hence, all the advantages of the traditional setting, in terms of estimation of empirical behavior, are preserved in the new framework.

A second aim of the paper, which is closely related to the first, deals with the general question of *assignability*, namely, how families allocate their consumption across members. In most cases, few data are available on the intrahousehold allocation of goods. A natural idea, then, is to investigate whether information about nonobservable consumption could be deduced, with the help of adequate theoretical assumptions, from data on whatever good, the consumption of which can be unambiguously assigned to one of the members. For instance, in the Prais-Houthakker line, Deaton (1988) and Deaton, Ruiz-Castillo, and Thomas (1989) use data on consumption of adult goods to derive results about intrafamily allocation.

The present paper shows that the collective approach provides a

Copyright © 1992. All rights reserved.

very promising framework for the assignability problem. Specifically, it is shown that, in an economy with two (assignable) labor supplies and a unique (nonassignable) Hicksian consumption good, the collective setting developed here allows one to assign each member's consumption—and actually to recover the entire decision process—up to an additive constant; in particular, the variations of intrahousehold distribution of consumption with respect to wages or nonlabor income can be predicted exactly. This important fact, which has apparently been overlooked so far, has potentially major consequences. It suggests that, provided that one is ready to believe in the kind of collective rationality that has just been alluded to (and will be formally defined in the next section), there is very much to be learned about *internal* rules and allocation processes of the household from the sole observation of its *external* behavior (i.e., labor supply or aggregate consumption). From an economist's view, thus, the household need not be the kind of "black box" suggested by the traditional approach, even when the allocation of resources within the household cannot be directly recorded. A corollary of this result is also relevant for the traditional approach as well: specifically, it states that a sufficient condition for assignability is separability of the household utility with respect to individual consumption-leisure bundles, a result derived in a different context by Deaton et al. (1989).

Finally, the collective viewpoint has important normative implications. Indeed, it follows from the previous result that the observation of labor supplies is sufficient for welfare comparisons to be performed, *even when each member's utility is to be independently taken into account.* Hence, instead of exclusively concentrating on the distribution of wealth, consumption, or well-being *across* households, welfare analysis could—hence should—consider *intrahousehold* allocation as well. The last goal of the paper is precisely to illustrate how the traditional tools of public economics can be translated into the collective setting and to investigate the consequences of this shift of interest.

The paper is organized as follows. Section II briefly discusses the basic issues at stake, and Section III presents the model. Section IV gives the main results. In Section V, I investigate the consequences on welfare analysis; in particular, I show how individual indirect utilities can be deduced from the sole knowledge of labor supplies. Section VI considers two particular specifications of the model, namely the "collective neoclassical case" (which characterizes the links between the traditional framework and the collective setting) on the one hand and the Nash-bargained framework on the other hand. Section VII extends the model to "caring" agents. Conclusions are discussed in Section VIII.

Copyright © 1992. All rights reserved.

II. Traditional versus "Collective" Models

A. The Traditional Approach

Traditional analysis basically models the household as though it were a single individual. This "household utility" approach has been adopted in most recent papers on labor supply, taxation, and welfare measurement (see, e.g., Pollak and Wales 1981; Ray 1982; King 1983; Blundell et al. 1986; Blundell and Walker 1986). A major advantage, which may explain this popularity, is that it fits exactly within the familiar treatment of consumer choice. In particular, the usual tools of optimal taxation and tax-benefit analysis can be directly applied to this kind of model. Integrability theorems, for instance, allow one to recover preferences from the sole observation of market behavior, clearly a necessary first step for any normative analysis.

It has become increasingly clear, however, that the traditional formalization, attractive and convenient as it may seem, still raises a number of serious difficulties. Its first shortcoming—which, in my view, is also a major one—is methodological: such models simply fall short of meeting the basic rules of neoclassical microeconomic analysis. Micro approaches are grounded on methodological individualism, which basically requires individuals to be characterized by their own preferences rather than be aggregated within the ad hoc fiction of a collective decision unit. Modeling a *group* (even reduced to two participants) as though it were a single individual can be seen only as a mere holistic deviation.[1] On the contrary, it is my claim that individualism should be referred to even when one is modeling household behavior; that is, the latter should be explicitly recognized as a *collective* process involving (except for singles) more than one decision unit.

A second drawback of the traditional setting is that it describes a household as a black box: while its relationships with the outside economy can be characterized, nothing can be said about its *internal* decision processes. In particular, such issues as the allocation of the household's resources among its members are simply ignored. And, of course, little (if anything) can be said about household formation or dissolution: how a pair of single preferences aggregate into a unique common utility by marriage (or disaggregate through divorce) is a question that can hardly be addressed at all. A natural question, at

[1] In the holistic view, "social groups . . . are conceived as the empirical objects which the social sciences study, in the same way in which biology studies animals or plants. This view must be rejected as naïve, . . . and has to be replaced by the demand that social phenomena, including collectives, should be analyzed in terms of individuals and their actions and relations" (Popper 1969, p. 341).

Copyright © 1992. All rights reserved.

this point, is whether one can enter into the black box. But reaching this goal presumably requires new tools of investigation.

As a matter of fact, scientific curiosity is not the only motivation for analyzing intrahousehold decision processes; welfare considerations may also matter. When considering, for instance, policy issues involving individual welfare (such as optimal taxation or cost-benefit analysis), traditional models can be seriously inadequate and, in some cases, misleading. They rest on the idea that only the distribution of income *across* households matters.[2] The underlying, implicit assumption is that the allocation of consumption or welfare *within* the household is either irrelevant or systematically optimal relative to the policymaker's preferences. Of course, this is a purely ad hoc hypothesis, the realism of which is dubious. There is no rationale, as well as no evidence, for assuming that the internal distribution of resources is even in any sense. As Apps (1991) and Haddad and Kanbur (1992) rightly suggest, taking into account intrahousehold inequality might well significantly alter a number of normative recommendations provided by the traditional approach.

B. Becker's Contribution

The previous criticisms highlight the need for an alternative line of investigation, which emphasizes the individual (rather than the "family") as the basic decision maker. Its origins can be traced back to the seminal works of Becker (1973, 1974*a*, 1974*b*, 1981*a*, 1981*b*), which introduce several modeling innovations. Within Becker's framework, the household consists of two members, each member being characterized by his or her own preferences. The marriage decision generates a gain, which is shared between the members according to a predetermined rule. In Becker (1973), the rule depends on the state of the market for marriage; in later works, Becker introduces the idea of "caring" by assuming that the preferences of one of the members depend on the other person's utility function. A consequence is the well-known "rotten-kid" theorem: even if only one of the members is "altruistic" in that sense, everyone within the household will try to maximize the joint family income.

Becker asserts that caring solves the household distribution and allocation problem in the general case. It must be stressed, however, that this solution relies on two strong ad hoc hypotheses. Indeed, any

[2] As Lazear and Michael (1988, p. 1) put it, "the myth persists in economic modelling of well-being and in many social policy contexts that once we know the level of resources available to the household, that is all we need to know."

Copyright © 1992. All rights reserved.

collective model of household behavior faces (at least) two difficulties. One is linked with the general problem of preference aggregation: if individual utilities differ, how should one characterize collective decisions? Becker essentially avoids this problem by assuming that there exists a *unique* aggregate consumption good that is produced by the household and consumed by each member; each member simply tries to maximize his share of the total. This amounts to assuming that preferences are (ordinally) defined by the production function and, hence, are (ordinally) identical across individuals. How this framework can be extended to take into account, say, each member's trade-off between consumption and leisure is not clear.

A second problem relates to the bargaining issue itself ("who gets what?"). The answer provided by the "caring" solution is quite particular: as emphasized by Ben-Porath (1982, p. 54), "a condition for the theorem to hold is that the altruists must have the last word" since they must be able to freely modify their transfers in response to the other person's decisions. In the same way, Manser and Brown (1980, p. 32) point out very rightly that Becker introduced de facto a particular bargaining rule, namely, maximizing the altruistic member's utility.

The "collective" lines of research have been recently extended in several different directions. Apps (1981, 1982) and Apps and Jones (1986) have introduced "Walrasian" cooperative models; Ashworth and Ulph (1981), Bourguignon (1984), Ulph (1988), and Woolley (1988) refer to noncooperative game theory. An interesting approach, initiated by Manser and Brown (1980) and McElroy and Horney (1981) and illustrated by Haddad and Kanbur (1989, 1990), relies on equilibrium concepts borrowed from cooperative game theory; specifically, the household decision problem is placed into a bargaining model between two individuals.

C. *"Collective" Models: The Efficiency Approach*

I shall now describe the main features of my approach. First, households consist of several members, each of them being characterized by specific preferences. Agents are "egoistic" in the sense that their utility depends only on their own consumption and labor supply; however, the framework can readily be extended to the case in which agents are "caring" (see Sec. VII). Moreover, the *collective* nature of the household decision-making process is explicitly recognized. I assume, specifically, that the process is cooperative, that is, that household decisions are Pareto efficient. The main originality of the approach lies in the fact that no additional assumption is made about

Copyright © 1992. All rights reserved.

the process; in other words, *no restriction is imposed a priori on which point of the Pareto frontier will be chosen.*[3]

An alternative interpretation of the efficiency hypothesis that may help clarify its content is the following. Assume that the decision process is a two-stage budgeting one. Members first divide the total nonlabor income received by the household between them, according to some predetermined sharing rule. The way in which the rule emerges is outside the scope of this analysis; it may reflect the cultural environment, the weight of tradition, or, as in Becker, the state of the market for marriage. In any case, in the Beckerian tradition, I shall assume throughout the paper that the rule is a given characteristic of the marriage contract that is not directly observable.[4] Once income has been allocated, both members face an individual budget constraint; then they choose their own consumption and labor supply through constrained utility maximization. I first show that the two interpretations are equivalent: household decisions are efficient if and only if a sharing rule exists. In other words, efficiency essentially means that members' bundles maximize their utility *for some given level of total expenditures;* the distribution of expenditures across members, on the other hand, is defined by the sharing rule, on which no particular assumption is made.

The sharing rule interpretation is quite helpful in understanding

[3] In particular, this approach does not allow one to derive labor supplies and consumptions from preferences, since the knowledge of both utilities and the budget constraint generates only a continuum of Pareto-efficient outcomes. It must, however, be kept in mind that what economics is interested in is the *opposite* derivation. Typically, household behavior can be observed empirically. The role of a formal theory (besides generating testable restrictions) is to help recover some unobservable features—here, private consumptions and preferences—that are needed for interpreting the results and formulating normative judgments. The reason why additional assumptions are not needed in the collective approach is precisely that, as we shall see, Pareto efficiency alone is sufficient for this goal: it allows one to recover both individual preferences and the decision process. In other words, one does not need ad hoc assumptions aimed at characterizing the location of the household choice on the Pareto frontier since this location is already embedded in the household behavior and can be deduced from the form of labor supplies. In this case, Occam's razor (*essentia non sunt multiplicanda praeter necessitatem*) clearly suggests that one should accept the simplest theory that works. An additional point can be stressed. Since the outcome of any cooperative bargaining model must be efficient, at least under symmetric information, this formalization encompasses, in particular, the Nash-bargained framework mentioned above as well as the main features of Becker's models. However, the Pareto approach requires neither a cardinal representation of preferences nor interpersonal comparability of utilities: the set of Pareto-efficient outcomes is defined even if utility is ordinal and cannot be compared across individuals. This is in sharp contrast with Nash bargaining approaches, in which the problem of cardinal representation may generate important difficulties (see Chiappori 1991).

[4] In some cases (e.g., divorce), the sharing rule may become observable. Then conditions (1) of proposition 4 below could be directly tested.

Copyright © 1992. All rights reserved.

how the decision process can be recovered from the sole observation of labor supplies. The basic idea is that, for each "egoistic" member, changes either in household nonlabor income or in the spouse's wage can have only an income effect; specifically, they will affect the member's behavior only insofar as her share of nonlabor income, as defined by the sharing rule, is modified. This means that any simultaneous change in nonlabor income and spouse i's wage that leaves unchanged spouse j's labor supply must keep constant j's share as well. From this idea, it is possible to derive, from the knowledge of labor supply functions, the indifference surfaces of the sharing rule. Since both shares add up to one, the sharing rule itself can actually be recovered up to an additive constant. Finally, knowing the rule allows one to write down each member's actual budget constraint, and preferences can then be computed in the usual way.

Clearly, such tools will not be fully reliable until they have been tested empirically; this will be the subject of forthcoming research. It is worth emphasizing, however, that the developments suggested in this paper are totally in line with the neoclassical tradition. Individuals are rational in the sense that they maximize utility under constraints. Collective agreements are mutually beneficial. The innovation, in fact, essentially consists in deepening the individualistic foundations of consumer theory by claiming that the members of the household should be considered *independently* rather than altogether as maximizing agents. In that sense, these new results illustrate the power of the individualistic paradigm.[5]

III. The Model

A. *The Basic Framework*

Let us consider a two-member household. Member i ($i = 1, 2$) consumes leisure in quantity L^i and a private Hicksian composite consumption good in quantity C^i. Labor supplies $T - L^1$ and $T - L^2$ (where T denotes total available time) are observed, together with wages w_1 and w_2, nonlabor income y, and aggregate consumption $C = C^1 + C^2$; the price of the consumption good is set to one. Private consumptions C^1 and C^2, on the other hand, are not observed. A natural interpretation of this framework is that available data are *cross-sectional.* Then wages and income vary across households, and

[5] The term "neoclassical" has sometimes been used to denote the household utility function approach (see McElroy and Horney 1981). This appellation, however, does not seem fully adequate since the collective approach—at least in the general framework presented here—relies on such neoclassical hypotheses as utility maximization for each individual and Pareto efficiency for collective decisions.

Copyright © 1992. All rights reserved.

prices remain constant—hence the introduction of a Hicksian composite good.

Within a traditional framework, the first step would be to assume the existence of a unique utility function $U(L^1, L^2, C^1, C^2)$, which is maximized under the budget constraint. It is clear, however, that one cannot hope to recover consumptions from labor supplies within this framework. Indeed, since the price ($p = 1$) is common to C^1 and C^2, from Hicks's theorem, one can estimate only a reduced-form direct (indirect) utility, $\tilde{U}(L^1, L^2, C)$ ($\tilde{V}(w_1, w_2, y)$). Welfare comparisons are possible, but they essentially ignore the distribution of aggregate consumption C between C^1 and C^2. In particular, there is a continuum of "structural" utility functions $U(L^1, L^2, C^1, C^2)$ that would lead to the same reduced form, and each of them is associated with a particular pair (C^1, C^2) of individual consumption functions.[6]

As discussed above, I follow here a different path and assume that each member i is endowed with "direct" preferences on her own leisure and consumption that are represented by an "egoistic" utility $U^i(L^i, C^i)$ with the usual properties. The egoistic form of individual preferences is important; however, it will be shown in the last section that "caring" à la Becker would lead to identical results.

Let us now formally express the Pareto efficiency assumption. Household behavior, hence, must be a solution of the following program:

$$\max U^1(L^1, C^1)$$

$$\text{subject to } \mu\colon\ U^2(L^2, C^2) \geq \bar{u}_2, \qquad (\bar{P})$$

$$\lambda\colon\ w_1L^1 + w_2L^2 + C^1 + C^2 \leq (w_1 + w_2)T + y$$

for some utility level $\bar{u}_2$. It must be clear that, in general, $\bar{u}_2$ is a function of the environment (i.e., of w_1, w_2, and y). For any given wage/income combination, the set of efficient outcomes obtains as $\bar{u}_2$ varies within its domain. This leads to the following formal definition.

DEFINITION. A pair of labor supply functions $(L^1(w_1, w_2, y), L^2(w_1, w_2, y))$, together with an (aggregate) consumption function defined by the budget constraint, is said to be collectively rational if there exists a pair of individual consumption functions $(C^1(w_1, w_2, y), C^2(w_1, w_2, y))$ and some function $\bar{u}^2(w_1, w_2, y)$, such that, for all (w_1, w_2, y), (i) $C^1(w_1, w_2, y) + C^2(w_1, w_2, y) = C(w_1, w_2, y)$ and (ii) (L^1, L^2, C^1, C^2) is a solution of program $(\bar{P})$.

[6] To see why, let $V(w_1, w_2, p_1 y)$ be the general form of the indirect utility function associated with U, should the price p_1 of consumption good 1 be allowed to differ from the price of good 2, set to one. Then V is compatible with the reduced form $\tilde{V}$ if and only if $V(w_1, w_2, 1, y) \equiv \tilde{V}(w_1, w_2, y)$. Obviously, a continuum of functions will satisfy this equality; for any of them, Roy's identity allows one to derive C^1 and C^2.

Copyright © 1992. All rights reserved.

Remark.—The Lagrange multiplier μ of the first constraint in $(\overline{P})$ can be interpreted as the implicit weight of member 2's egoistic utility in the collective decision process; that is, $(\overline{P})$ is equivalent to the maximization of $U^1(L^1, C^1) + \mu U^2(L^2, C^2)$ under the budget constraint. It is important to note here that, in general, μ will be a function of w_1, w_2, and y. Incidentally, a particular case of $(\overline{P})$ obtains if one assumes the existence of a *fixed* household welfare function W, so that the household maximizes $W(U^1(L^1, C^1), U^2(L^2, C^2))$ subject to the budget constraint. Of course, since W is assumed to depend only on U^1 and U^2 and not on w_1, w_2, and y per se,[7] this hypothesis is extremely restrictive. In fact, this form is at the intersection of the neoclassical and collective frameworks. I shall consider this point in Section VI.

B. *The Sharing Rule Interpretation*

Let us now consider the alternative, "sharing rule," interpretation. That is, let us assume that nonlabor income y is shared between the members, and let $\varphi(w_1, w_2, y)$ be the amount received by member 1 and $y - \varphi(w_1, w_2, y)$ by member 2 (note that φ is allowed to depend on wages as well as on nonlabor income).[8] Moreover, φ may well be negative or greater than y (for instance, if y is low and wages are very different, one member may share labor income with the other). Now each member independently chooses consumption and labor supply, subject to the corresponding budget constraint. Member i's program can thus be written as

$$\begin{aligned} &\max U^i(L^i, C^i) \\ &\text{subject to } w_i L^i + C^i \leq w_i T + \varphi^i(w_1, w_2, y), \end{aligned} \tag{P_i}$$

where φ^1 stands for φ and φ^2 for $y - \varphi$.

The following result states that the income sharing rule interpretation is exactly equivalent to the initial setting, that is, that the existence of a sharing rule implies no more (and no less) than efficiency of the collective decision process.

PROPOSITION 1. Let $L^1(w_1, w_2, y)$ and $L^2(w_1, w_2, y)$ be arbitrary functions. There exists a function $\overline{u}_2(w_1, w_2, y)$ such that L^1 and L^2 are solutions of $(\overline{P})$ if and only if there exists a function $\varphi(w_1, w_2, y)$ such that L^i is a solution of (P_i) $(i = 1, 2)$.

[7] Of course, writing W as a function of U^1, U^2 *and* w_1, w_2, y is equivalent to $(\overline{P})$. This point illustrates a formal difference between the collective model and traditional consumer analysis: the maximand in the former case will in general depend on prices.

[8] In general, φ can be thought of as a function of wages, income, and labor supplies. But here, labor supplies are given functions of wages and income, so that only the reduced form of the rule matters.

Copyright © 1992. All rights reserved.

Proof. Assume, first, that L^1 and L^2 together with two consumptions C^1 and C^2 are a solution of $(\bar{P})$ for some function $\bar{u}_2$. Define $\varphi(w_1, w_2, y) = w_1[T - L^1(w_1, w_2, y)] - C^1(w_1, w_2, y)$. Then L^1 is a solution of (P_1); otherwise, it would be possible to increase member 1's utility without changing member 2's expenditures, a contradiction. The same argument applies to L^2.

Conversely, assume that L^1 and L^2, together with C^1 and C^2, are solutions of (P_1) and (P_2) for some function φ. Define $\bar{u}_2(w_1, w_2, y) = U^2(L^2(w_1, w_2, y), C^2(w_1, w_2, y))$. Then L^1 and L^2 are solutions of $(\bar{P})$. Indeed, we know that $w_2L^2 + C^2 = e^2(w_2, \bar{u}_2)$, where e^2 is the expenditure function associated with U^2. Hence, the cost of any pair (L'^2, C'^2) providing the utility level $\bar{u}_2$ will be no less than $w_2T + y - \varphi(w_1, w_2, y)$. In particular, if a solution (L'^1, C'^1, L'^2, C'^2) of $(\bar{P})$ is such that $U^1(L'^1, C'^1) > U^1(L^1, C^1)$, the pair (L'^1, C'^1) costs no more than (L^1, C^1), a contradiction.

One can now characterize the set of labor supply functions that are consistent with the collective framework just described. From now on, I assume that L^1 and L^2 are three times continuously differentiable, and I impose the condition that φ is twice continuously differentiable.

IV. Characterization of the "Collective" Setting

Several questions can be raised, at this point, on the collective setting: (i) *Characterization:* Which necessary restrictions are imposed on L^1 and L^2 by the collective setting? (ii) *Integrability:* Are the restrictions sufficient; that is, is it possible, from any pair of labor supply functions satisfying them, to recover a sharing rule and a pair of individual preferences? (iii) *Uniqueness:* Are the sharing rule and the individual preferences uniquely determined?

An answer to question i appears in Chiappori (1988*b*); in order to make this paper self-contained, I shall briefly recall the principal result. In what follows, the notation X_z stands for the partial differential of function X with respect to variable z and $A = L^1_{w2}/L^1_y$ and $B = L^2_{w1}/L^2_y$ whenever $L^1_y \cdot L^2_y \neq 0$. Also, I introduce the following "regularity" assumption.

ASSUMPTION R. $L^1_y \neq 0$, $L^2_y \neq 0$, and $AB_y - B_{w2} \neq BA_y - A_{w1}$ for almost all (w_1, w_2, y) in $\mathbb{R}^2_+ \times \mathbb{R}$.

Note that assumption R is "generically" true, in the usual sense. An answer to question i is then given by the following result.

PROPOSITION 2. In the general case, the following conditions are necessary for any given pair (L^1, L^2) of demand for leisure functions to be solutions of (P_1) and (P_2) for some C^2 sharing rule φ: (*a*) $\alpha_y A + \alpha A_y - \alpha_{w2} = 0$, (*b*) $\beta_y B + \beta B_y - \beta_{w1} = 0$, (*c*) $L^1_{w1} - L^1_y[(T - L^1 -$

Copyright © 1992. All rights reserved.

$\beta B)/\alpha] \leq 0$, and (*d*) $L^2_{w2} - L^2_y[(T - L^2 - \alpha A)/\beta] \leq 0$, where

$$\alpha = \left(1 - \frac{BA_y - A_{w1}}{AB_y - B_{w2}}\right)^{-1} \quad \text{if } AB_y - B_{w2} \neq 0, \alpha = 0 \text{ otherwise,}$$

and

$$\beta = 1 - \alpha = \left(1 - \frac{AB_y - B_{w2}}{BA_y - A_{w1}}\right)^{-1}.$$

Conditions *a*, *b*, *c*, and *d* are analogous to Slutsky conditions in the sense that they provide a set of partial differential equations and inequalities that have to be satisfied by labor supply (or, here, demand for leisure) functions. In particular, it is possible, given any particular functional form for labor supplies, to translate the conditions into restrictions on parameters. The economic interpretation of α, β, A, and B will be discussed later; however, it is important to note here that they are totally defined (and can be computed immediately) from labor supplies.

We now come to the second and third questions, namely, integrability and uniqueness. Are conditions *a–d sufficient* for the existence of a sharing rule and a pair of utility functions from which L^1 and L^2 can be derived? And are the sharing rule and the preferences uniquely determined? The answer is given by the following results.

PROPOSITION 3. *Integrability.*—Let L^1 and L^2 be two C^3 functions satisfying assumption R and conditions *a–d*; let $\overline{w} = (\overline{w}_1, \overline{w}_2, \overline{y})$ be any point in $\mathbb{R}^2_{++} \times \mathbb{R}$. Then there exists a neighborhood $\mathcal{V}$ of $\overline{w}$ such that (i) there exists a sharing rule φ, defined over $\mathcal{V}$, and (ii) there exists a pair of utility functions (U^1, U^2) with the property that the solution of (P_i), at any point of $\mathcal{V}$, is the couple (L^i, C^i) for some $C^i \geq 0$.

PROPOSITION 4. *Uniqueness.*—Under the same hypothesis as proposition 3, (i) the sharing rule is defined up to an additive constant k; specifically, its partials are given by

$$\varphi_y = \alpha, \quad \varphi_{w2} = A\alpha, \quad \varphi_{w1} = B(\alpha - 1) = -\beta B; \tag{1}$$

(ii) for each choice of k, the preferences represented by U^1 and U^2 are uniquely defined; and (iii) the indifference curves corresponding to different values of k can be deduced from one another by translation.

The integrability result is local rather than global because of nonnegativity restrictions; that is, the sharing rule can in general be derived globally, but it must be checked that it does not lead to negative consumptions for particular values of (w_1, w_2, y). Also, the uniqueness

Copyright © 1992. All rights reserved.

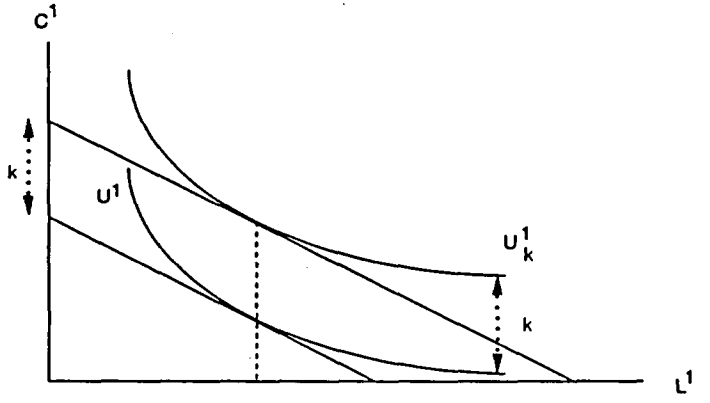


FIG. 1

result states that preferences are defined up to a translation and sharing rules up to an additive constant. This fact is an immediate consequence of the nonobservability of individual consumptions. The intuition goes as follows. Assume that a sharing rule φ and a pair of utilities U^1 and U^2 generate the observed demands for leisure. For any given k, an obvious alternative solution is defined by[9]

$$\begin{aligned} \varphi_k(w_1, w_2, y) &= \varphi(w_1, w_2, y) + k, \\ U_k^1(L, C) &= U^1(L, C - k), \\ U_k^2(L, C) &= U^2(L, C + k). \end{aligned} \tag{2}$$

Specifically, (2) simply says that the indifference curves of U_k^1 can be deduced from those of U^1 by vertical translation of magnitude k (see fig. 1). Of course, it is impossible to distinguish between (φ, U^1, U^2) and $(\varphi_k, U_k^1, U_k^2)$ from the sole observation of labor supplies. Thus it is clear that φ can be defined only up to a positive constant k, and once k has been chosen, the corresponding preferences can be recovered through (2).

The economic interpretation is clear. We can always assume that member i systematically receives k additional units of the consump-

[9] Indeed, with φ_k and U_k^1, program (P_1) can be written as

$$\begin{aligned} &\max U_k^1(L, C) = U^1(L, C - k) \\ &\text{subject to } w_1 L + C = w_1 T + \varphi_k(w_1, w_2, y) = w_1 T + \varphi(w_1, w_2, y) + k. \end{aligned} \tag{P_1'}$$

If $\overline{C} = C - k$, (P_1') becomes

$$\begin{aligned} &\max U^1(L, \overline{C}) \\ &\text{subject to } w_1 L + \overline{C} = w_1 T + \varphi(w_1, w_2, y), \end{aligned}$$

which is exactly (P_1) for φ and U^1.

Copyright © 1992. All rights reserved.

tion good (member j's share being reduced accordingly), provided that preferences are modified in such a way that member i, with the new utility U_k and the additional k units, is exactly as well off as (initially) with utility U_0 and the "normal" share φ. In figure 1, both the indifference curves and the budget constraint are translated vertically; this, of course, does not affect labor supply. The (strong) result in proposition 4, however, is that this is the *only* degree of freedom left in the choice of φ. As we shall see, though φ is indeterminate strictly speaking, this indeterminacy is not really harmful since it does not affect welfare comparisons.

Proof of propositions 2, 3, and 4.—A detailed proof can be found in Chiappori (1988*a*, 1989). A sketch of the argument follows. It is clear, from program (P_i) above, that we have $L^i(w_1, w_2, y) = l^i(w_i, \varphi^i(w_1, w_2, y))$, where l^i is i's Marshallian demand. Hence, $A = L^1_{w_2}/L^1_y = \varphi_{w_2}/\varphi_y$ and $B = L^2_{w_1}/L^2_y = -\varphi_{w_1}/(1 - \varphi_y)$.

With cross-derivative restrictions, this implies conditions a and b of proposition 2 plus relation (1) of proposition 4. In turn, the latter defines φ up to a positive constant, and integrability follows from traditional arguments. Q.E.D.

In addition, proposition 4 helps to provide a simple interpretation of the various parameters introduced so far. Here, $\varphi(w_1, w_2, y)$ is the share of nonlabor income y received by member 1, and α is the derivative of φ with respect to y. In words, α is the share of marginal nonlabor income received by member 1 (and, of course, $\beta = 1 - \alpha$ is the share received by member 2). Proposition 2 simply shows how α can be deduced from observed labor supplies.

V. Welfare Comparisons

Let us now consider the welfare properties of the collective framework. Assume that a reform changes the initial price-income bundle $\bar{w} = (\bar{w}_1, \bar{w}_2, \bar{y})$ to $w' = (w'_1, w'_2, y')$. How should the consequences on individual well-being be analyzed from the collective viewpoint?

A. *Indirect Collective Utilities*

A first difficulty arises because of the collective nature of the decision process; the model is general enough for well-known paradoxes to appear. For instance, one might, in the Hicks-Kaldor-Scitovsky line, consider as globally beneficial for the household a reform that could *potentially* ameliorate both individuals' welfare (i.e., such that there is a point, on the new Pareto frontier, at which both members are better off than initially). However, it is well known (and could easily be illustrated in the model) that this criterion is neither complete nor

Copyright © 1992. All rights reserved.

acyclical. It can simultaneously accept a given reform (say, going from $\bar{w}$ to w') as well as the opposite move (from w' to $\bar{w}$). Also, a criterion of this kind may be particularly inadequate in view of the policy concern evoked above. The policymaker will probably be interested in *actual,* rather than *potential,* changes in welfare. If, because of the internal decision process, one of the members is worse off after the reform, this fact should be taken into account irrespective of whether another decision rule could have led to socially "better" outcomes.

An obvious advantage of the collective model, from this point of view, is the possibility of recovering private consumptions as well as individual welfare functions. That is, from the observation of labor supplies, one can deduce the consequences of the reform on each member. The fact that the sharing rule is defined only up to a constant does not raise any particular difficulty, as shown by the following result.

COROLLARY 1. Let φ, U^1, and U^2 be associated with a pair of given labor supply functions (in the sense of propositions 3 and 4). If U^1 (U^2) is increased by the reform, then for any k, U_k^1 (U_k^2) is also increased by the reform.

Proof. From proposition 4, replacing φ by φ_k and U^1 by U_k^1 does not change utility levels. Q.E.D.

One can now define a pair of collective indirect utility functions, v^1 and v^2. Intuitively, $v^i(w_1, w_2, y)$ is i's welfare when wages are w_1 and w_2 and *household* nonlabor income is y. In particular, v^i must be distinguished from the traditional indirect utility $V^i(w_i, Y)$ associated with U^i, which gives i's welfare when wage is w_i and i's *potential income* is Y. The difference is that, in the definition of v, the sharing rule is implicitly taken into account; that is, the relationship between those functions is simply

$$v^1(w_1, w_2, y) = V^1(w_1, w_1 T + \varphi(w_1, w_2, y)) \tag{3a}$$

$$v^2(w_1, w_2, y) = V^2(w_2, w_2 T + y - \varphi(w_1, w_2, y)). \tag{3b}$$

From relations (1) and (3), we can deduce that

$$v_y^1 = V_y^1 \cdot \varphi_y = V_y^1 \cdot \alpha, \tag{4a}$$

$$v_{w2}^1 = V_y^1 \cdot \varphi_{w2} = V_y^1 \cdot \alpha A, \tag{4b}$$

$$v_{w1}^1 = V_{w1}^1 + V_y^1(T + \varphi_{w1}) = V_{w1}^1 + V_y^1(T - \beta B). \tag{4c}$$

Since $L^1 = -(V_{w1}^1/V_y^1)$, equation (4c) gives

$$v_{w1}^1 = V_y^1(T - L^1 - \beta B). \tag{4d}$$

Also, note that V_y^1 (i.e., 1's marginal utility of his *own* income) is equal to λ, the Lagrange multiplier of the budget constraint in program

Copyright © 1992. All rights reserved.

($\bar{P}$). The economic interpretation is clear: the utility, for member 1, of an additional dollar received by the household is simply his marginal utility of income, multiplied by the share of the marginal dollar he will receive. In the same way, the marginal utility, for member 1, of member 2's wages is equal to his marginal utility of (own) income, multiplied by the increase in his income that results from the change in w_2. We can thus state the following formal result.

PROPOSITION 5. Let v^1 and v^2 be indirect utilities associated with L^1 and L^2. Then

$$v^1_{w1} = \lambda(T - L^1 - \beta B), \quad v^2_{w1} = \frac{\lambda}{\mu}\beta B, \tag{5a}$$

$$v^1_{w2} = \lambda\alpha A, \quad v^2_{w2} = \frac{\lambda}{\mu}(T - L^2 - \alpha A), \tag{5b}$$

$$v^1_y = \lambda\alpha, \quad v^2_y = \frac{\lambda}{\mu}\beta. \tag{5c}$$

Since v^1 and v^2 are defined up to composition by an increasing function, the partial derivatives are defined up to a multiplicative positive function (λ or λ/μ). Also, μ is the relative weight of members in the household decision process. In particular, the derivatives satisfy the following relationships:

$$\begin{aligned} v^1_y + \mu v^2_y &= \lambda, \\ v^1_{w1} + \mu v^2_{w1} &= \lambda(T - L^1), \\ v^1_{w2} + \mu v^2_{w2} &= \lambda(T - L^2). \end{aligned} \tag{6}$$

Relations (6) simply express the fact that, locally, program ($\bar{P}$) is equivalent to the maximization of $U^1 + \mu U^2$ under the budget constraint. As a first consequence, we can deduce a collective counterpart of traditional Roy identities:

$$\begin{aligned} T - L^1 &= \frac{v^1_{w1} + \mu v^2_{w1}}{v^1_y + \mu v^2_y}, \\ T - L^2 &= \frac{v^1_{w2} + \mu v^2_{w2}}{v^1_y + \mu v^2_y}. \end{aligned} \tag{7}$$

Second, it can be noted that $v^1_y + \mu v^2_y$ is always positive; that is, an increase in nonlabor income unambiguously ameliorates the collective maximand. However, v^1_y (v^2_y) is positive if and only if α is positive (α is lower than one). Though this may seem a natural assumption, it is by no means implied by the collective setting; specifically, depending on L^1 and L^2, α might perfectly take values outside (0, 1). To see

Copyright © 1992. All rights reserved.

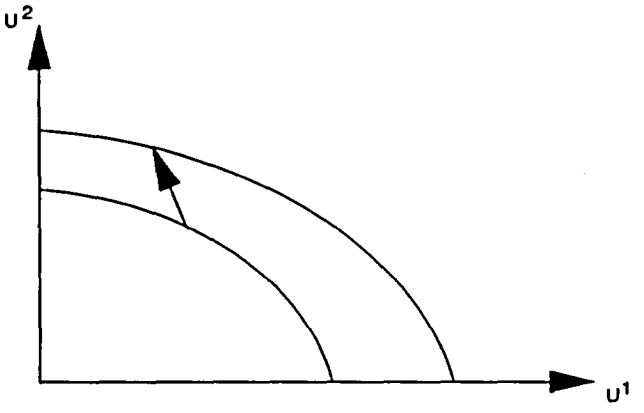


FIG. 2

why this is possible, consider figure 2, where axes represent utilities. Whenever y is increased, the Pareto frontier is translated to the northeast. It is then clear that, whatever collective index is maximized, the household will be better off in the sense defined by this index. This, however, does certainly not imply that *each* member will be better off. That is, the point chosen over the efficiency frontier may be moved as indicated in figure 2, where member 1's utility is *decreased* (of course, member 2's is increased). This illustrates the fact that an increase in nonlabor income (due to, say, higher benefits) might *lower* the welfare of one of the members. Also, the important point here is that one can directly check, from the sole knowledge of labor supplies, whether this is the case or not. Of course, the same argument exactly applies to wage increases as well.

A last consequence of equations (5) is that they allow one to recover indirect utilities from labor supplies. Since v^1 and v^2 are defined up to composition by an increasing function, we shall deduce the indifference surfaces of v^i from labor supplies. Briefly, consider the equation

$$v^1(w_1, w_2, y) = K \Leftrightarrow y = \psi^K(w_1, w_2), \tag{8}$$

which defines the generic indifference surface. Then

$$\psi^K_{w1} = -\frac{v^1_{w1}}{v^1_y}, \quad \psi^K_{w2} = -\frac{v^1_{w2}}{v^1_y};$$

hence ψ^K must satisfy

$$\psi^K_{w1} = -\frac{1}{\alpha}(T - L^1 - \beta B), \quad \psi^K_{w2} = -A. \tag{9}$$

Copyright © 1992. All rights reserved.

Here, L^1, α, β, A, and B can be expressed as functions of w_1, w_2, and $y = \psi^K(w_1, w_2)$; hence (9) provides a partial differential system, which can be integrated to give the ψ^K ($K \in \mathbb{R}$). Of course, K will be nothing else than the integration constant; the set of solutions of (9), indexed by this constant, is exactly the set of indifference surfaces. An example of a computation using a specific functional form is given below.

B. Welfare Effects of Tax Reforms

Let us go back to my initial question, that is, what are the effects of a reform that changes $\overline{w}$ into w'? There are four possible cases: (i) both v^1 and v^2 are increased, (ii) both v^1 and v^2 are decreased, (iii) v^1 is increased and v^2 is decreased, and (iv) v^1 is decreased and v^2 is increased.

The set of possible wage/income bundles, hence, will typically be partitioned into four areas corresponding to the various possible situations, in contrast to the traditional approach, characterized by two areas only. An illustration is given in figure 3 (under the assumption that only w_1 and w_2 are changed). These areas can be computed from labor supply functions, as indicated above; the frontiers are simply the indifference curves (or surfaces) of v^1 and v^2.

Finally, how can μ, the implicit weight of member 2, be recovered? First note that μ is defined only conditionally on a particular *cardinal* representation of preferences; that is, one must first choose two functions v^1 and v^2 (consistent with the ψ defined above). Now μ is simply defined by $\mu = (\beta/\alpha)(v^1_y/v^2_y)$. But, of course, this relation requires comparability of preferences across individuals; otherwise, μ is meaningless!

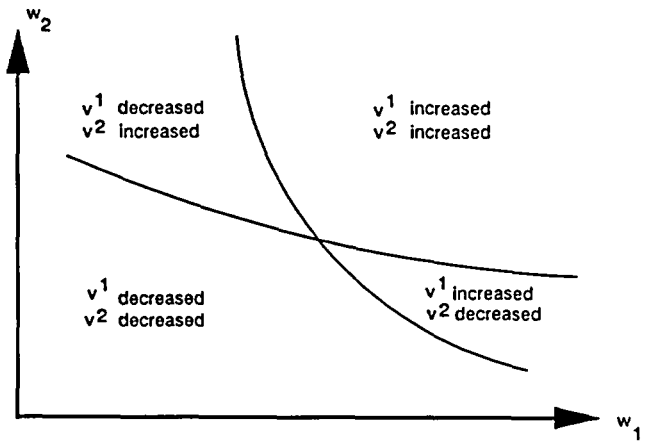


FIG. 3

Copyright © 1992. All rights reserved.

C. An Application

The previous results can be illustrated using a particular functional form. Consider the following demand for leisure:

$$\begin{aligned} L^1 &= a_1 + b_1 y + c_1 y \log y + d_1^1 w_1 + d_1^2 w_2, \\ L^2 &= a_2 + b_2 y + c_2 y \log y + d_2^1 w_1 + d_2^2 w_2. \end{aligned} \tag{10}$$

These demands are linear, except for the Workin-Leser additional term $y \log y$, which allows for flexibility in income (i.e., nonlinear Engel curves). Since $L^1_{w^2} = d_1^2$ and $L^1_y = (b_1 + c_1) + c_1 \log y$, we get

$$A = d_1^2(b_1 + c_1 + c_1 \log y)^{-1}, \quad B = d_2^1(b_2 + c_2 + c_2 \log y)^{-1} \tag{11}$$

and

$$\begin{aligned} \alpha &= \frac{c_2}{c_2 b_1 - b_2 c_1}(b_1 + c_1 + c_1 \log y), \\ \beta &= \frac{c_1}{c_1 b_2 - b_1 c_2}(b_2 + c_2 + c_2 \log y). \end{aligned} \tag{12}$$

The signs of α and β depend not only on the signs but also on the respective magnitudes of the income terms b_i and c_i. It can be noted that, for y high enough, either α or β must be negative; a typical graph is presented in figure 4.

Conditions a and b of proposition 2 are always satisfied since α_{w^2} is zero and αA is constant. Condition c gives

$$d_1^1 + \frac{c_1}{c_2} d_2^1 - \frac{D}{c_2}(T - L^1) \leq 0,$$

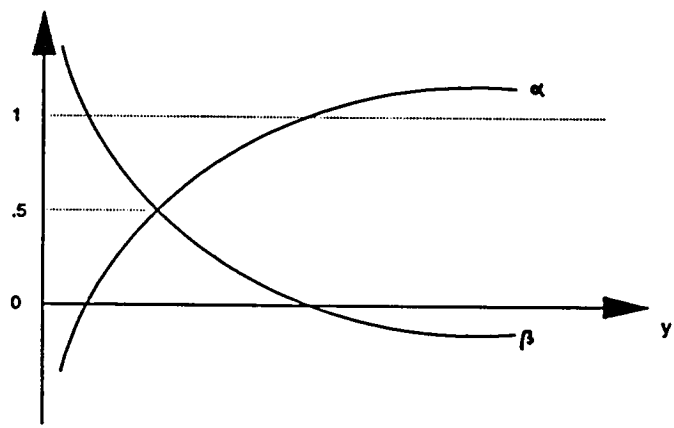


FIG. 4

Copyright © 1992. All rights reserved.

and, in the same way, condition d gives

$$d_2^2 + \frac{c_2}{c_1} d_1^2 - \frac{D}{c_1}(T - L^2) \leq 0,$$

where $D = c_1 b_2 - b_1 c_2$.

Assuming that these conditions are satisfied, we can, first, derive the sharing rule. It is characterized by

$$\varphi_y = \alpha = \frac{c_2}{D}(b_1 + c_1 + c_1 \log y),$$

$$\varphi_{w2} = A\alpha = \frac{d_1^2 c_2}{D}, \tag{13}$$

$$\varphi_{w1} = -\beta B = -\frac{d_2^1 c_1}{D},$$

which gives

$$\varphi = \frac{c_2}{D}(b_1 y + c_1 y \log y) + \frac{d_1^2 c_2}{D} w_2 + \frac{d_2^1 c_1}{D} w_1 + k. \tag{14}$$

Also, we can recover the collective indirect utility v^1; indeed, with previous notation, the generic indifference curve ψ satisfies

$$\psi_{w1} = -\frac{1}{\alpha}(T - L^1 - \beta B), \quad \psi_{w2} = -A.$$

Defining $\theta = b_1 \psi + c_1 \psi \log \psi$, we get

$$\theta_{w2} = -d_1^2$$

and

$$\theta_{w1} = \frac{D}{c_2}(-T + L^1 + \beta B)$$

$$= \frac{D}{c_2}\left(-T + a_1 + \theta + d_1^1 w_1 + d_1^2 w_2 - \frac{c_1 d_2^1}{D}\right).$$

Hence

$$\theta = K e^{Dw_1/c_2} - d_1^1 w_1 - d_1^2 w_2 + \gamma_1,$$

where K is the integration constant and

$$\gamma_1 = \frac{d_1^1 c_2 + d_2^1 c_1}{D} + T - a_1.$$

Copyright © 1992. All rights reserved.

This leads to

$$v_1(w_1, w_2, y) = e^{-Dw_1/c_2}(-\gamma_1 + b_1 y + c_1 y \log y + d_1^1 w_1 + d_1^2 w_2) \quad (15)$$

and, by symmetry,

$$v_2(w_1, w_2, y) = e^{-Dw_2/c_1}(-\gamma_2 + b_2 y + c_2 y \log y + d_2^1 w_1 + d_2^2 w_2).$$

Finally, for these particular indirect utilities, the reader may check that

$$\mu = \frac{\alpha}{\beta} \cdot \frac{v^1}{v^2} = \left(\frac{c_1}{c_2}\right)^2 \exp D\left(\frac{w_2}{c_1} - \frac{w_1}{c_2}\right). \quad (16)$$

Note that μ is always nonnegative.

In words, the labor supply function defined by equation (14) can be derived from the maximization of the (household) welfare index

$$U^1(L^1, C^1) + \left(\frac{c_1}{c_2}\right)^2 \exp D\left(\frac{w_2}{c_1} - \frac{w_1}{c_2}\right) U^2(L^2, C^2),$$

where U^i is the corresponding direct utility. Also, *traditional* indirect utilities can easily be derived from the previous results. Just note that

$$v^1(w_1, w_2, y) = e^{-Dw_1/c_2} \frac{D}{c_2}\left[(\varphi + w_1 T) - w_1 T + \gamma\left(1 - \frac{c_2}{D}\right) - T + a_1\right].$$

Hence, for $Y = w_1 T + \varphi$,

$$V^1(w_1, Y) = D_1 e^{-Dw_1/c_2}(Y - w_1 T + \gamma_1'), \quad (17)$$

with obvious notation. Direct utilities can readily be recovered from (17).

VI. Further Specifications of the Decision Process

A. *Collective Neoclassical Labor Supply*

Though the collective approach provides an alternative framework, it is by no means incompatible with the traditional setting. In fact, an interesting situation obtains when both representations are simultaneously fulfilled. Assume, for instance, that the household behavior can be represented by the maximization, under the budget constraint, of a (unique) utility function of the form $W[U^1(L^1, C^1), U^2(L^2, C^2)]$. Of course, the neoclassical hypotheses are satisfied since a unique index is maximized. On the other hand, it is well known that, for any given pair of individual utility functions (U^1, U^2), the maximization

Copyright © 1992. All rights reserved.

of a Bergsonian index $W[U^1, U^2]$ will always lead to Pareto-efficient decisions. Hence the collective assumptions are verified as well. How can the corresponding labor supply functions be characterized? And how can preferences be recovered?

It should be noted that the answers to these questions cannot be immediately derived from traditional results. On the one hand, the collective index W must be separable in (L^1, C^1) and (L^2, C^2). Should private consumptions be observable, this would simply imply, in addition to Slutsky conditions, that Gorman's separability conditions are satisfied. But here, C^1 and C^2 are not observed; moreover, the two corresponding prices are always equal. For these two reasons, Gorman's conditions cannot be used directly. Hence it is not clear, from a traditional point of view, which conditions on L^1 and L^2 *alone* are necessary or sufficient for the existence of such a "collective neoclassical utility."

A complete answer can nevertheless be deduced from the previous results. Let Π denote the following program:

$$\begin{aligned} &\max W[U^1(L^1, C^1), U^2(L^2, C^2)] \\ &\text{subject to } w_1L^1 + w_2L^2 + C^1 + C^2 = (w_1 + w_2)T + y. \end{aligned} \tag{Π}$$

PROPOSITION 6. Let L^1 and L^2 be arbitrary C^2 functions of w_1, w_2, and y. The following conditions are necessary and sufficient for the existence of functions C^1, C^2, U^1, U^2, and W such that L^1 and L^2 are solutions of (Π): (i) conditions *a–d* of proposition 2 and (ii) the Slutsky conditions

$$L^1_{w2} - (T - L^2)L^1_y = L^2_{w1} - (T - L^1)L^2_y,$$

$$L^1_{w1} - (T - L^1)L^1_y \leq 0,$$

$$L^2_{w2} - (T - L^2)L^2_y \leq 0.$$

Proof. See Chiappori (1989).

Proposition 6 clearly shows the key role of the separability assumption that is implicit in the collective setting. We have seen in Section I that the neoclassical framework is not sufficient for the derivation of C^1 and C^2. However, under the additional restriction that household preferences should be separable in individual welfares, C^1 and C^2 can be identified up to a constant, and individual as well as household utilities can be recovered. In particular, we get, as a corollary of this proposition, a result that has been known for some time within the "assignability" literature (see Deaton et al. 1989), namely, that individual consumptions can be recovered provided that preferences exhibit some adequate separability properties.

Copyright © 1992. All rights reserved.

COROLLARY. Assume that the household maximizes a unique utility function that is separable in individual consumption/labor supply bundles. Then individual consumptions can be assigned up to an additive constant.

Also, it should be noted that this "collective utility" case is highly particular within the collective framework developed above. The existence of a (fixed) household utility function is by no means an innocuous assumption; on the contrary, it entails strong additional restrictions on the form of labor supply. For example, it is easy to exhibit functional forms of labor supply that are compatible with the collective setting, but not with the Slutsky conditions. It can thus be argued that, while the traditional model is not restrictive enough, that model plus separability may be unnecessarily restrictive (when compared with the collective approach). Slutsky restrictions are not necessary, whether for recovering income sharing rules or for welfare purposes.

B. Bargaining Approaches

A second particular case of the general framework developed above is the cooperative game theory approach, initiated by Manser and Brown (1980) and McElroy and Horney (1981). The basic idea is to place the household decision problem into a bargaining framework and then to use some cooperative equilibrium concept (e.g., Nash bargaining). Specifically, one must first define, for each agent i, a "threat point" $\overline{U}^i$, corresponding to the minimum level of welfare the agent can obtain "if no collective agreement is reached." Then, for any given wage/income combination, the outcome of the decision process is supposed to maximize the product $(U^1 - \overline{U}^1)(U^2 - \overline{U}^2)$ under the budget constraint; hence, collective labor supply will be the solution of the Nash bargaining program:

$$\begin{aligned} &\max\,[U^1(L^1, C^1) - \overline{U}^1][U^2(L^2, C^2) - \overline{U}^2] \\ &\text{subject to } w_1 L^1 + w_2 L^2 + C^1 + C^2 \leq (w_1 + w_2)T + y. \end{aligned} \tag{NB}$$

Of course, any solution of this program is Pareto efficient; hence, it is a particular case of $(\overline{P})$. An interesting question, however, is whether (NB) provides restrictions on labor supplies that go beyond those given in proposition 2, that is, whether the Nash bargaining assumption brings additional structure to the general setting we have just developed. The answer is not immediate, because a program like (NB) entails several degrees of freedom, described as follows.

1. We must define the threat points $\overline{U}^i$. The question here is what is exactly meant by the expression "if no collective agreement is

Copyright © 1992. All rights reserved.

reached." Some authors (Ulph 1988; Woolley 1988) have suggested that threat points should be identified with the (assumed unique) *noncooperative* Nash equilibrium of the game; intuitively, if the players cannot agree, they will skip to a noncooperative kind of behavior, with Nash equilibria as natural outcomes. This idea leads to a kind of two-stage process: the household first computes the noncooperative utility levels and then uses them as status quo points for deriving the Nash-bargained outcomes. Whether this setting leads to restrictions on labor supplies, however, is an open question.

On the contrary, McElroy and Horney (1981) argue that threat points must be understood as individual utility levels when divorce is involved. Specifically, if agent i's wage (nonlabor income) when divorced is w_i (y_i), then $\overline{U}^i = V^i(w_i, y_i)$, where V^i is i's indirect utility function. Two ambiguous points, however, remain. First, should we take V^i as the indirect utility function that corresponds to i's direct, "egoistic" utility when married, U^i? In this case, we must assume that preferences (say, the marginal rate of substitution between leisure and consumption) are not affected by the marital status, a rather strong hypothesis. Conversely, if preferences are allowed to depend on the marital status, then the *simultaneous* estimation of the U's and the V's may be difficult. Second, we must be able to observe the way in which household nonlabor income would be split between the members in case of divorce; again, this will turn out to be a difficult task and may require ad hoc assumptions on the divorce procedure.

2. We must choose a particular *cardinal* representation of preferences. It must be remembered, indeed, that Nash bargaining requires cardinality, since composing utility functions, threat points, or both by an arbitrary monotonic mapping modifies the solution of (NB). This point may, in particular, raise some problems when one is trying to independently estimate preferences or threat points (see n. 10 below).

Of course, the choice between these alternative options will crucially affect the structure of the model. The simplest way, here, is probably to assume that preferences do not depend on marital status. In that case, the Nash-bargained decision must be a solution of the following program:

$$\max [U^1(L^1, C^1) - V^1(w_1, y_1)][U^2(L^2, C^2) - V^2(w_2, y_2)]$$

$$\text{subject to } w_1 L^1 = w_2 L^2 + C^1 + C^2 \leq (w_1 + w_2)T + y_1 + y_2 \quad \text{(NB')}$$

for some particular cardinal representation of preferences for the agents. This, in turn, leads to the following formally well-defined problem: Given any pair (L^1, L^2) of labor supply functions that satisfy the conditions of proposition 2, is it possible to find a particular

Copyright © 1992. All rights reserved.

cardinal representation of preferences, such that $(L^1(w_1, w_2, y), L^2(w_1, w_2, y))$ is a solution of (NB') for all w_1, w_2, and y (where $y = y_1 + y_2$)? Should the answer be negative, then the Nash bargaining framework (or, more precisely, the simplified version defined by the assumptions above) would actually introduce additional structure within the general model presented here. My conjecture is that this is the case. However, no formal proof has been provided so far; this must be the topic of further research. Also, the additional conditions, even if they exist, may be extremely difficult to derive formally, as suggested by the failure of most previous attempts.

Finally, we may broaden the set of assumptions by allowing for altruistic preferences or for dependence of preferences to marital status. This line has been followed, for instance, by McElroy and Horney's (1981) model; however, this (very general) framework does not seem restrictive enough to provide tractable restrictions, at least in the simple framework investigated here.[10]

Even if *formal* conditions may be difficult to derive, the bargaining approach may, however, suggest several intuitive conclusions that, in some cases, lead to empirical tests. As an illustration, consider the way in which the sharing rule φ is affected by a change in the wage of one of the members. Assume, for instance, that member 2's wage is increased. How should member 1's share φ of nonlabor income be modified? One may, at this point, interpret the sharing process in two opposite ways. A first interpretation would emphasize the "redistribution" purpose: transfers occur to compensate inequalities in wage incomes within the household. In that case, an increase in w_2 ameliorates member 2's situation and hence reduces the need for a compensating transfer in his favor; a consequence is that φ should *increase*. The alternative interpretation, based on the general idea of a bargaining process, leads to an inverse conclusion. An increase in

[10] From McElroy and Horney's (1981) approach, it is simply not possible to derive falsifiable restrictions on demand functions; a detailed analysis appears in Chiappori (1988*a*) (see also Chiappori [1990] for a more formal argument). It has been recently suggested, in particular by McElroy (1992) and McElroy and Horney (1992), that restrictions could be deduced from a bargaining model with "egoistic" agents (plus a collective good) by the following trick: Estimate indirect utilities when divorced from a distinct sample of divorced individuals. Then compute accordingly threat points for married couples and estimate the whole bargaining model. However, a number of difficulties still have to be solved. Selectivity bias can be difficult to correct and is likely to require an explicit model of household formation. Also, whether preferences are allowed to depend on marital status clearly becomes an essential issue (if they are, preferences when married may be difficult to recover). Finally, the Nash bargaining equilibrium concept may turn out to be inadequate for this approach. The reason is that it requires a *cardinal* representation of preferences that is difficult to obtain by independent estimation on a sample of divorced individuals (such an estimation will typically provide [at best] an ordinal representation).

Copyright © 1992. All rights reserved.

w_2 ameliorates member 2's bargaining strength since he would probably be better off than before in case of divorce; technically, member 2's threat point, whatever its precise definition, is likely to increase with w_2. In this case, member 2 will be able to recover a larger share of nonlabor income, and φ should consequently *decrease*. In other words, the "compensating transfer" story suggests the following properties for φ: $\varphi_{w_2} = \alpha A > 0$ and $\varphi_{w_1} = -\beta B < 0$, whereas bargaining ideas lead to $\varphi_{w_2} = \alpha A < 0$ and $\varphi_{w_1} = -\beta B > 0$.

Again, it can be stressed that both hypotheses can be tested from the knowledge of labor supplies. In particular, the choice of the appropriate theoretical structure (e.g., Nash bargaining vs. redistribution) could be guided by a preliminary estimation of the general Pareto model, which would allow one to check whether the bargaining conditions above are indeed satisfied. In the same line, the Pareto model can provide a useful framework for testing more specific approaches (such as bargaining models) since the former encompasses the latter.

VII. Extensions

A. *Caring*

Agents have been modeled so far as egoistic, in the sense that each agent's utility depends only on his own consumption and labor supply. This assumption, however, is quite restrictive; in particular, the egoistic model provides little rationale for household formation (or dissolution). It is thus important to stress that *egoistic preferences are not necessary for the previous results to hold true.* Specifically, caring (à la Becker) can be introduced without fundamentally altering the conclusions of the model. To see why, assume that member i actually maximizes some "altruistic" index $W^i[U^1(L^1, C^1), U^2(L^2, C^2)]$ that depends on both his own direct, "egoistic" utility and his spouse's; W^i is continuous, increasing, and quasi-concave. Now, a fundamental remark is that *any decision that is Pareto efficient within this new setting would be Pareto efficient as well, were the agents egoistic.* Indeed, any change that ameliorates U^1 without decreasing U^2 would strictly increase both W^1 and W^2; hence such a change is not possible if the starting point is already efficient for the W. It can readily be shown that the locus of Pareto-efficient decisions for altruistic agents is a connected subset of the Pareto frontier $\mathcal{P}$ derived from egoistic preferences. Specifically, the egoistically efficient outcomes that are efficient for the W as well lie on $\mathcal{P}$ between the members' best choices under caring (fig. 5). In other words, the basic conclusions above (with the possible exception of corner solutions) still apply when caring is introduced. Note that

Copyright © 1992. All rights reserved.

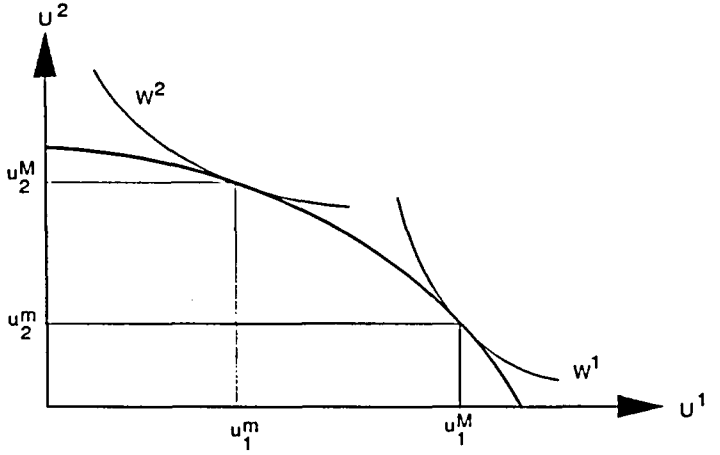


FIG. 5

this property is due to the Pareto hypothesis; specifically, it holds true because we consider the whole Pareto frontier (rather than any particular point on it). It would not obtain, for instance, within any of the bargaining frameworks discussed in Section VI.

This conclusion is not really surprising. As it may have become clear from the remarks above, the important property of this setting, from which most of the results derive, is the separability of the welfare indexes in (L^1, C^1) on the one hand and (L^2, C^2) on the other hand. However, though this property does not seem exceedingly restrictive, it is certainly not an innocuous assumption. Typically, my marginal rate of substitution between leisure and consumption may depend on my wife's free time. Hence, the most general form for individual utilities should be $U^i(L^1, C^1, L^2, C^2)$. The problem, however, is that such a structure might well not be restrictive at all.[11] And, of course, no uniqueness conclusion can be expected to hold in this case. So the cost of introducing this general form of altruism may be very high, in terms of predictive power of the model.

B. *Many Consumption Goods, Public Consumption, and So Forth*

A natural extension of the model would be to assume that there are n consumption goods in the economy; the more general hypothesis is

[11] In fact, it is possible to derive nonparametric restrictions characterizing labor supplies in that case (Chiappori 1988*b*). Conversely, it can be shown that, whenever one of the spouses is assumed not to work, any finite set of data on labor supply is compatible with this setting (Chiappori 1990).

Copyright © 1992. All rights reserved.

that, for each of them, only aggregate consumption can be observed. Sticking to the cross-sectional interpretation mentioned above, we would be interested, in that case, in deriving the way in which *each member's* consumption of each good depends on wages and nonlabor income. This will be the topic of further research; I conjecture, however, that such a derivation is possible in general.

An analogous, but in a sense opposite, extension would be to distinguish between various kinds of income; this will be the case, for instance, if a part y_1 of nonlabor income y is directly given to member 1, who can spend it at his own will (we may think of family benefits received by the wife, personal wealth, etc.). For instance, several contributions (e.g., McElroy 1990; Thomas 1992) have tested the "income pooling" hypothesis (only total nonlabor income may matter) that is implied by the traditional approach. In the framework of this paper, the sharing rule φ will depend on four arguments, namely, w_1, w_2, y_1, and $y_1' = y - y_1$. It can be shown that this additional component will generate additional testable conditions on labor supplies; again, a detailed exposition appears in Chiappori (1989). In other words, the model presented here can perfectly encompass such empirical facts as differential effects of each spouse's own income on behavior. Moreover, it allows one to derive additional conditions that characterize this situation and leads to immediate econometric tests. This is done in particular by Bourguignon et al. (1992).

Finally, the framework has to be extended to take into account the existence of collective consumption within the household. This task may require specific assumptions, for example, that each member's preferences are additively separable with respect to the collective goods. This aspect is of special interest since it would allow one to introduce children's expenditures within the model, provided that one is ready to assume that children's consumption can be modeled as a public good for parents' preferences.

VIII. Conclusions

The main conclusions of the paper can be summarized as follows. First, if we model the household as a pair of individuals characterized by a particular utility function of their leisure and consumption or, alternatively, an altruistic index of the "caring" type, then Pareto efficiency alone generates a set of testable restrictions on labor supplies, which are independent of (though not incompatible with) the traditional conditions. Second, in sharp contrast to the traditional approach, this collective setting allows one to "assign" private consumptions, as well as to recover individual welfare functions. In particular, it is possible to deduce, from the shape of labor supply func-

Copyright © 1992. All rights reserved.

tions, the income sharing rule within the household. Hence, the field open to normative judgment is no longer limited to the interhousehold distribution of welfare; the issue of intrahousehold allocation can be considered as well. The collective approach seems especially adequate for analyzing the effects of particular policies (e.g., tax-benefit systems) on *individual* poverty or inequality. Of course, the value judgments implied by such concepts cannot stop one from facing the difficulties due to the collective nature of the situation. For instance, a reform may increase a member's well-being at the expense of the other's, so that the consequence on social welfare is not straightforward. The collective approach, however, enables one to weight individual utilities differently from what is implicit in the household decision process instead of assuming that the latter is always socially optimal.

The basic framework can be extended in several ways, such as several goods or several sources of income. Of course, numerous questions deserve further work. Collective conditions must be empirically tested against neoclassical ones; this will be the topic of forthcoming research. From a purely theoretical viewpoint, the problems linked with the multiplicity of consumption goods can be investigated. Of special interest is the introduction of collective goods or, on the contrary, of specific commodities that can be consumed only by one of the members. Also, "corner" solutions, such as nonparticipation, need to be analyzed, and domestic labor supply should be considered as well. Finally, the collective approach should not be limited to labor supply. In the spirit of the model above, Bourguignon et al. (1992) investigate the differential effects of individual incomes on household consumption. They find that, while the traditional approach (and specifically the "income pooling" property) is strongly rejected, the restrictions implied by the collective framework are compatible with the data (a general survey of this line of research can be found in Bourguignon and Chiappori [1992]). Hence, the collective approach should be viewed as a general research program, in which most of the work still has to be done.

References

Apps, Patricia F. *A Theory of Inequality and Taxation.* Cambridge: Cambridge Univ. Press, 1981.

———. "Institutional Inequality and Tax Incidence." *J. Public Econ.* 18 (July 1982): 217–42.

———. "Modeling Female Labour Supply and Tax Reform: Welfare Effects of Work at Home and Intra-Family Inequality." Manuscript. Sydney: Univ. Sydney, 1991.

Copyright © 1992. All rights reserved.

Apps, Patricia F., and Jones, Glenn S. "Selective Taxation of Couples." *Zeitschrift für Nationalökonomie*, suppl. 5 (1986), pp. 1–15.

Ashworth, J. S., and Ulph, D. T. "Household Models." In *Taxation and Labour Supply*, edited by C. V. Brown. London: Allen & Unwin, 1981.

Becker, Gary S. "A Theory of Marriage: Part I." *J.P.E.* 81 (July/August 1973): 813–46.

———. "A Theory of Marriage: Part II." *J.P.E.* 82, no. 2, pt. 2 (March/April 1974): S11–S26. (*a*)

———. "A Theory of Social Interactions." *J.P.E.* 82 (November/December 1974): 1063–93. (*b*)

———. "Altruism in the Family and Selfishness in the Market Place." *Economica* 48 (February 1981): 1–15. (*a*)

———. *A Treatise on the Family*. Cambridge, Mass.: Harvard Univ. Press, 1981. (*b*)

Ben-Porath, Yoram. "Economics and the Family—Match or Mismatch? A Review of Becker's *A Treatise on the Family*." *J. Econ. Literature* 20 (March 1982): 52–64.

Blundell, Richard W.; Meghir, Costas; Symons, Elizabeth; and Walker, Ian. "A Labour Supply Model for the Simulation of Tax and Benefit Reforms." In *Unemployment, Search and Labour Supply*, edited by Richard W. Blundell and Ian Walker. Cambridge: Cambridge Univ. Press, 1986.

Blundell, Richard W., and Walker, Ian. "A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data." *Rev. Econ. Studies* 53 (August 1986): 539–58.

Bourguignon, François. "Rationalité individuelle ou rationalité stratégique: le cas de l'offre familiale de travail." *Rev. Economique* 35 (January 1984): 147–62.

Bourguignon, François; Browning, Martin; Chiappori, Pierre-André; and Lechene, Valerie. "Intrahousehold Allocation of Consumption: Some Evidence from French Data." *Ann. d'Economie et de Statistique* (1992), in press.

Bourguignon, François, and Chiappori, Pierre-André. "Collective Models of Household Behaviour: An Introduction." *European Econ. Rev.* (1992), in press.

Chiappori, Pierre-André. "Nash-bargained Household Decisions: A Comment." *Internat. Econ. Rev.* 29 (November 1988): 791–96. (*a*)

———. "Rational Household Labor Supply." *Econometrica* 56 (January 1988): 63–90. (*b*)

———. "Modelling Collective Household Decisions: Welfare Implications." Manuscript. Paris: DELTA, 1989.

———. "La fonction de demande de biens collectifs: théorie et application." *Ann. d'Economie et de Statistique*, no. 19 (July–September 1990), pp. 27–42.

———. "Nash-bargained Household Decisions: A Rejoinder." *Internat. Econ. Rev.* 32 (August 1991): 761–62.

Deaton, Angus S. "The Allocation of Goods within the Household: Adults, Children, and Gender." Working Paper no. 39. Washington: World Bank, 1988.

Deaton, Angus S.; Ruiz-Castillo, Javier; and Thomas, Duncan. "The Influence of Household Composition on Household Expenditure Patterns: Theory and Spanish Evidence." *J.P.E.* 97 (February 1989): 179–200.

Haddad, Lawrence, and Kanbur, Ravi. "How Serious Is the Neglect of Intra-Household Inequality?" Discussion Paper no. 95. Coventry: Univ. Warwick, 1989.

Copyright © 1992. All rights reserved.

———. "Is There an Intra-Household Kuznets Curve?" Discussion Paper no. 101. Coventry: Univ. Warwick, 1990.

———. "Intra Household Resource Allocation and the Theory of Targeting." *European Econ. Rev.* (1992), in press.

King, Mervyn A. "Welfare Analysis of Tax Reforms Using Household Data." *J. Public Econ.* 21 (July 1983): 183–214.

Lazear, Edward P., and Michael, Robert T. *Allocation of Income within the Household.* Chicago: Univ. Chicago Press, 1988.

McElroy, Marjorie B. "The Empirical Content of Nash-bargained Household Behavior." *J. Human Resources* 25 (Fall 1990): 559–83.

McElroy, Marjorie B., and Horney, Mary Jean. "Nash-bargained Household Decisions: Toward a Generalization of the Theory of Demand." *Internat. Econ. Rev.* 22 (June 1981): 333–49.

———. "Nash-bargained Household Decisions: Reply." *Internat. Econ. Rev.* 31 (February 1990): 237–42.

Manser, Marilyn, and Brown, Murray. "Marriage and Household Decision-making: A Bargaining Analysis." *Internat. Econ. Rev.* 21 (February 1980): 31–44.

Pollak, Robert A., and Wales, Terence J. "Demographic Variables in Demand Analysis." *Econometrica* 49 (November 1981): 1533–51.

Popper, Sir Karl R. *Conjectures and Refutations: The Growth of Scientific Knowledge.* 3d ed. London: Routledge and Kegan Paul, 1969.

Ray, Ranjan. "Estimating Utility Consistent Labour Supply Functions: Some Results in Pooled Budget Data." *Econ. Letters* 9, no. 4 (1982): 389–95.

Thomas, Duncan. "The Distribution of Income and Expenditures within the Household." *Ann. d'Economie et de Statistique* (1992), in press.

Ulph, D. T. "A General Non-cooperative Nash Model of Household Consumption Behaviour." Working paper. Bristol: Univ. Bristol, 1988.

Woolley, F. "A Non-cooperative Model of Family Decision Making." Working Paper no. 125. London: London School Econ., 1988.

Copyright © 1992. All rights reserved.

Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment

Peter A. Diamond

The Journal of Political Economy, Volume 75, Issue 5 (Oct., 1967), 765-766.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at http://www.jstor.ac.uk/about/terms.html, by contacting JSTOR at jstor@midas.ac.uk, or by calling JSTOR at 0161 275 7919 or (FAX) 0161 275 6040. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The Journal of Political Economy is published by University of Chicago. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.ac.uk/journals/ucpress.html.

The Journal of Political Economy
©1967 University of Chicago

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor@midas.ac.uk.

©1998 JSTOR

http://www.jstor.ac.uk/
Tue Nov 3 16:53:24 GMT 1998

CARDINAL WELFARE, INDIVIDUALISTIC ETHICS, AND INTERPERSONAL COMPARISON OF UTILITY: COMMENT

PETER A. DIAMOND
Massachusetts Institute of Technology

NOT very recently, Professor John Harsanyi (1955) presented in this *Journal* three appealing axioms for social choice under uncertainty which lead to the conclusion of a social welfare function which is additive in individual utilities. While not directly addressing his defense of these axioms, I wish to argue that one of them is not consistent with notions of justice held by some individuals.[1] Since this is an ethical discussion, the argument will take the form of an example which suggests the problem inherent in the axiom and some comments on the nature of the example.

Harsanyi's three axioms are: (1) individual decision making satisfies the axioms for expected utility maximization; (2) social welfare can be written as an increasing function of individual expected utilities; and (3) social choice satisfies the axioms for expected utility maximization. It is the third axiom with which I wish to quarrel.

In mathematical terms, we can express the first axiom as individual choice conforms to the maximization of expected utility, v_i, where

$$v_i = \int u_i [c_i(\theta)] \, dF(\theta) \, , \qquad (1)$$

with u_i being the utility function of the ith individual; $c_i(\theta)$, his consumption in state θ; and $F(\theta)$ the probability distribution of the states of nature.

The second axiom is that social choice should conform to the maximization of welfare, w, which can be written as a function of individual expected utilities:

$$w_1 = f_1 (v_1, v_2, \ldots, v_n) \, . \qquad (2)$$

This axiom implies that social choice can be expressed as a choice among vectors of expected utilities, which are determinate, not random, and thus social choice under uncertainty need not be considered.

The third axiom says that the social objective function can be written as expected welfare (with welfare a function of individual utilities):

$$w_2 = \int f_2 \{u_1 [c_1(\theta)] \, , \; u_2 [c_2(\theta)] \, , \ldots , \, u_n [c_n(\theta)]\} \, dF(\theta) \, . \qquad (3)$$

These three axioms imply that welfare can be written additively,

$$w = \sum_{i=1}^{n} \lambda_i \int u_i [c_i(\theta)] dF(\theta). \qquad (4)$$

(In the presence of differing individual subjective probabilities, a case not considered by Harsanyi, these three axioms are inconsistent.)

As an example, let us consider a society composed of two identical individuals, A and B, facing a choice between two alternatives, α and β, with two possible and equally probable states of nature, θ_1 and θ_2. Let us further assume that social choice, in addition to satisfying the first two axioms above, is symmetric in its treatment of the two individuals. It is assumed that under alternative α, the utility of A is 1 and that of B is zero, independent of the state of nature; while under β, these are the utility levels if θ_1 occurs, but they are reversed if θ_2 occurs. In tabular form, we have

[1] This comment is also relevant for part of the Robert H. Strotz paper (1958). For further discussion of these matters, see also Franklin M. Fisher and Jerome Rothenberg (1961, 1962) and Strotz (1961).

	if θ_1 occurs	if θ_2 occurs
Alternative α:	$u_A = 1$, $u_B = 0$,	$u_A = 1$, $u_B = 0$,
Alternative β:	$u_A = 1$, $u_B = 0$,	$u_A = 0$ $u_B = 1$.

Harsanyi's third axiom, in combination with the other assumptions, leaves society indifferent between the two alternatives. However, β seems strictly preferable to me, since it gives B a fair shake while α does not. (In terms of expected utilities, under α we have $v_A = 1$ and $v_B = 0$ while under β, $v_A = v_B = \frac{1}{2}$.)

I am willing to accept the sure-thing principle for individual choice but not for social choice, since it seems reasonable for the individual to be concerned solely with final states while society is also interested in the process of choice.

REFERENCES

Fisher, Franklin M., and Rothenberg, Jerome. "How Income Ought To Be Distributed: Paradox Lost," *J.P.E.*, LXIX (April, 1961), 162–80.

———. "How Income Ought To Be Distributed: Paradox Enow," *ibid.*, LXX (February, 1962), 88–93.

Harsanyi, John. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *J.P.E.*, LXIII (August, 1955), 320.

Strotz, Robert H. "How Income Ought To Be Distributed: A Paradox in Distributive Ethics," *J.P.E.*, LXVI (June, 1958), 189–205.

———. "How Income Ought To Be Distributed: Paradox Regained," *ibid.*, LXIX (June, 1961), 271–78.

Soc Choice Welf (2013) 40:411–437
DOI 10.1007/s00355-011-0611-9

ORIGINAL PAPER

Collective risk aversion

Elyès Jouini · Clotilde Napp · Diego Nocetti

Received: 8 July 2011 / Accepted: 8 October 2011 / Published online: 1 November 2011
© Springer-Verlag 2011

Abstract In this article we analyze the risk attitude of a group of heterogeneous agents and we develop a theory of comparative collective risk tolerance. In particular, we characterize how shifts in the distribution of individual levels of risk tolerance affect the group's attitude towards risk. In a model with efficient risk-sharing and two agents an increase in the level of risk tolerance of one or of both agents might have an ambiguous impact on the collective level of risk tolerance; the latter increases for some levels of aggregate wealth while it decreases for other levels of aggregate wealth. For more general populations we characterize the effect of first-order like shifts (individual levels of risk tolerance more concentrated on high values) and second-order like shifts (more dispersion on individual levels of risk tolerance) on the collective level of risk tolerance. We also evaluate how shifts in the distribution of individual levels of risk tolerance impact the collective level of risk tolerance in a framework with exogenous egalitarian sharing rules. Our results permit to better characterize differences in risk taking behavior between groups and individuals and among groups with different distributions of risk preferences.

1 Introduction

Many decisions to undertake risks are made by groups. A priori, one would expect that the theory of comparative risk aversion developed by Pratt (1964) and Arrow (1971),

E. Jouini
Université Paris-Dauphine, Ceremade, 75016 Paris, France
e-mail: jouini@ceremade.dauphine.fr

C. Napp
CNRS & Université Paris-Dauphine, DRM, 75016 Paris, France

D. Nocetti (✉)
School of Business, Clarkson University, Potsdam, NY, USA
e-mail: dnocetti@clarkson.edu

which characterizes the proclivity of individuals to undertake risks, would easily translate into a theory of group risk taking. Consider, for instance, three individuals, *A, B,* and *C*. Suppose that *C* is more risk averse than *B* and *B* is more risk averse than *A*. Intuition strongly suggests that, when acting together, *A* and *C* would be less willing to undertake risks than *A* and *B*. Paradoxically, Mazzocco (2004) showed that such intuition is not always correct. For some levels of wealth an increase in the degree of risk aversion of the most risk averse individual in a group may decrease the collective level of risk aversion.[1] Mazzocco (2004) presented this paradoxical result through a numerical example with two individuals and isoelastic preferences. Our objective in this article is to extend this line of inquiry by establishing precisely the conditions for this phenomenon to occur and, more generally, by evaluating how changes in the distribution of individual preferences affect a group's attitudes towards risk.

To be perfectly clear, the Arrow–Pratt theory of comparative risk aversion does apply to utility functions of groups. So, for example, if a group is more risk averse than another in the Arrow–Pratt sense then this group will also require a larger risk premium to eliminate a fair risk. The interpretation of such comparative statics result, however, is clouded by the following fact. The degree of risk aversion of the group depends upon both the distribution of preferences among the agents and their optimal allocations. So a change in the distribution of preferences impacts the collective level of risk aversion through two channels. A direct one as well as an indirect one due to the fact that changes in the distribution of preferences lead, in turn, to changes in the efficient allocation of wealth. Therefore, if risk is shared efficiently, collective risk aversion has to be determined endogenously.

There is one special case in which the problem greatly simplifies: given an efficient allocation of wealth, if all individuals in the group have a constant and common absolute cautiousness (the derivative of the reciprocal of absolute risk aversion), e.g., under CARA or CRRA utility functions with a common level of relative risk aversion, the group has the same absolute cautiousness (Wilson 1968). Comparative statics of risk aversion at the aggregate level is then not different from comparative statics at the individual level. The assumption of homogeneity in individual preferences, however, does not have empirical support (e.g., Barsky et al. 1997) and, in fact, defeats the purpose of Arrow–Pratt's theory of comparative risk aversion. Therefore, in this article we tackle the problem of comparing attitudes towards risk among groups composed by individuals with heterogeneous risk preferences.

We show, in the setting of Mazzocco's (2004) paper, that the collective level of risk tolerance is a wealth share weighted average of the individual levels of risk tolerance and increasing the risk tolerance level of one agent has two effects: an increase of one of the terms of the average but a possible decrease of its relative weight in the average. As a result, there are two possible shapes for the collective risk tolerance as a function of the risk tolerance level of one of the agents: increasing curve or increasing then decreasing curve.

[1] The fact that efficient groups may behave in a complex manner is well known. Pratt and Zeckhauser (1989) showed, for example, that a group may be willing to accept a gamble which combines two individually unacceptable lotteries.

In fact, we establish the possibility of an even more perplexing situation: An increase in the degree of risk tolerance of *both* members of a couple may decrease their collective degree of risk tolerance.[2] We clearly characterize these different situations in terms of the size of the aggregate endowment relative to the endowment that corresponds to the fair efficient allocation. We also characterize, for the two-agent case and for more general populations, first-order like shifts (individual levels of risk tolerance more concentrated on high values) that have an unambiguous impact on the collective level of risk tolerance.

Since the key aspect of our analysis is preference heterogeneity we also evaluate how more dispersion on the individual levels of risk tolerance (second-order shifts) affects the collective risk preferences. We show that, for high levels of wealth (relative to the level that corresponds to the fair efficient allocation), more heterogeneity tends to increase collective risk tolerance, while the opposite is true for low levels of wealth.

Finally, we extend our analysis to a framework in which all members of a group receive the same endowment (egalitarian groups). This setup is appropriate to analyze situations in which the members of a group derive utility from a public good and situations in which a private good is simultaneously consumed by many individuals. For example, many goods within a household are simultaneously consumed by all the members of a family. Within this framework, and under very general individual preferences, we establish the impact of first- and second-order shifts on the collective level of risk tolerance.

In addition to the work of Mazzocco (2004), our article is closely related to the work of Hara et al. (2007), who studied the properties of collective preferences for a given distribution of individual risk preferences. We extend their analysis by evaluating how changes in the distribution of individual preferences affect the collective attitudes towards risk. In this way, our analysis also complements the work of Gollier (2001, 2007), who explored how heterogeneity in the initial endowment of wealth and how heterogeneity in beliefs affect a group's attitude towards risk. At a more general level, we believe that our results may shed light into the empirical literature on 'choice shifts', which compares decisions made by groups relative to decisions made by the members of the group in situations of uncertainty (e.g., Baker et al. 2008; Shupp and Williams 2008; Masclet et al. 2009), a topic which we further discuss in Sect. 6.[3]

The article proceeds as follows: In Sect. 2 we present the model with efficient risk sharing and we establish a number of useful results about the efficient allocations of endowments and the collective risk preferences. In Sect. 3 we briefly evaluate the case of CARA preferences, which serves as a useful benchmark. In Sect. 4 we analyze the case of isoelastic heterogeneous preferences. After presenting general properties of collective preferences we evaluate shifts in the distribution of individual preferences, first in the case of two agents and then under more general populations. In Sect. 5

[2] We also show, however, that a uniform increase in the degree of risk tolerance of both individuals unambiguously increases the risk tolerance of the group.

[3] In this literature the objective is to elicit the risk attitude of groups as compared to the members of the group. Another strand of related empirical literature evaluates whether, under uncertainty, groups behave in a more consistent manner than individuals (see e.g., Bone et al. 1999; Charness et al. 2007).

we evaluate collective risk preferences for the case of exogenous egalitarian sharing rules, while Sect. 6 concludes. All the proofs are provided in Appendix.

2 The model

We consider a standard static model in which a group of heterogeneous agents consume a single good. The endowment per person in the consumption good is defined by a random variable x on the probability space (Ω, F, P). Agents have a common belief over the probability space. In order to take into account finite as well as infinite sets of agents, the agent space is described by (I, ι, Q), where $I = [0, \infty)$ and Q is a probability measure on I. Individuals are indexed by $i \in I$ and we denote by E^Q the expectation with respect to Q.

We consider a 'consensus' group *à la* Samuelson (1956). That is, the group acts as if there was a social planner who wants to reach a Pareto efficient allocation of risks and solves the following maximization program

$$U(x) = \max_{\int x_i \mathrm{d}Q(i) = x} \int \lambda_i u_i(x_i) \mathrm{d}Q(i). \tag{1}$$

where u_i is the utility function of agent i, x_i is the consumption of agent i, and λ_i is the weight (e.g., decision power) granted to agent i. The utility function $U(x)$ corresponds to the highest social utility level among all possible endowment distributions across agents.

Throughout the article, we make the following assumption on the utility functions.

Assumption *(U)* For all i, the utility function $u_i : [d_i, \infty) \to \mathbb{R} \cup \{-\infty\}$ is assumed to be infinitely differentiable on (d_i, ∞) with $u_i' > 0$ and $u_i'' < 0$ and satisfies Inada's conditions, i.e., $\lim_{x \to d_i} u_i'(x) = \infty$ and $\lim_{x \to \infty} u_i'(x) = 0$.

For a given agent i and a given consumption level x, the absolute (resp. relative) risk aversion $A_i(x)$ (resp. $R_i(x)$), the absolute (resp. relative) risk tolerance $t_i(x)$ (resp. $s_i(x)$) are given by

$$A_i(x) = -\frac{u_i''(x)}{u_i'(x)}, \qquad R_i(x) = -x\frac{u_i''(x)}{u_i'(x)} = xA_i(x)$$
$$t_i(x) = -\frac{u_i'(x)}{u_i''(x)} = \frac{1}{A_i(x)}, \qquad s_i(x) = -\frac{u_i'(x)}{xu_i''(x)} = \frac{1}{R_i(x)} = \frac{t_i(x)}{x}.$$

Note that CARA and CRRA utility funtions clearly satisfy Assumption (U).

If we denote by v the function defined by $v(x, i) = u_i'(x)$, we will also make the following assumption.

Assumption *(LSPM)* The function v is log-supermodular in (x, i), i.e., $\frac{\partial \log v}{\partial x}(x, i)$ is nondecreasing in i.

Remark that the log-supermodularity of $v(x,i)$ means that $A(x,i) = A_i(x)$ is non-increasing in i or that agent i is less risk averse (and more risk tolerant) than agent j when $i \geq j$.

We have then the following classical result

Proposition 1 *Under Assumption (U), there exists a family of functions* $(f_i)_{i\in[0,\infty]}$ *such that*

- $f_i : [d,\infty) \to [d_i,\infty)$ *with* $d = \int d_i \mathrm{d}Q(i)$ *is infinitely differentiable and increasing*
- $\int f_i(x)\mathrm{d}Q(i) = x$ *for all* $x \in [d,\infty)$
- $U(x) = \int \lambda_i u_i(f_i(x))\mathrm{d}Q(i)$.

We will say that $(f_i)_{i\in[0,\infty]}$ is an efficient sharing rule associated with the maximization program of Eq. 1.

We recall the following well known results that relate the collective risk aversion and risk tolerance to the individual ones through the efficient sharing rule.

Proposition 2 (Wilson 1968; Hara et al. 2007) *Let us assume that (U) is satisfied. Let* x *be a given aggregate wealth and let* $(f_i)_{i\in I}$ *be the efficient sharing rules associated with the maximization program of Eq. 1. The collective absolute risk tolerance* $t(x) = -\frac{U'(x)}{U''(x)}$ *and the collective relative risk tolerance* $s(x) = -\frac{U'(x)}{xU''(x)}$ *are given by*

$$t(x) = \int t_i(f_i(x))\mathrm{d}Q(i),$$
$$s(x) = \int \frac{f_i(x)}{x} s_i(f_i(x))\mathrm{d}Q(i).$$

The relative risk tolerance $s(x)$ of the group is then an average of the individual levels of relative risk tolerances $s_i(f_i(x))$ weighted by the optimal individual shares of consumption. Analogously, the degree of relative risk aversion of the group is an average of the individual degrees of relative risk aversion.The group is then less risk averse than the most risk averse agent and more risk averse than the least risk-averse one. In terms of the example given in the introduction this implies, in particular, that a group composed by B and C will always be less willing to undertake risks than a group composed by A and B.

It is easy to show that $s'(x)$ is positive and then that the collective relative risk aversion $R(x) = -x\frac{U''(x)}{U'(x)}$ is decreasing in x. This fact has been underlined by Hara et al. (2007, Proposition 6). They further show (Corollary 7) that $R(x)$ approaches the degree of relative risk aversion of the most (least) risk averse agent as x converges to zero (infinity).

At this stage we consider very general utility functions and we may assume, without loss of generality, that all the members of the group are granted the same weight (it suffices to replace the utility function u_i by $\lambda_i u_i$, note that the LSPM property is not impacted by this modification). In the next we consider then the equally weighted Pareto optimum. We also assume that there exists an efficient fair allocation. In other words, there exists x^* such that $(x_i)_{i\in I}$, with $x_i = x^*$ for all i, is efficient.

The following proposition provides an analysis of how the aggregate consumption x is shared among the agents depending on the position of x relatively to the fair efficient allocation x^*

Proposition 3 *Under the Assumptions (U) and (LSPM), we have the following results.*

1. *For $x \geq x^*$, the optimal allocation $(x_i)_{i\in I}$ associated to the aggregate wealth x is such that $x_i \geq x^*$, for all i, and x_i increases with i. Furthermore, if all the utility functions are DARA then $t_i(x_i)$ increases with i.*
2. *For $x \leq x^*$, the optimal allocation $(x_i)_{i\in I}$ associated to the aggregate wealth x is such that $x_i \leq x^*$, for all i, and x_i decreases with i.*

Although of some interest by itself, this Proposition will also play an important role in the analysis that follows.

3 CARA utility functions

Let us consider constant absolute risk-aversion/tolerance utility functions of the form

$$u_i(x) = -\theta_i \exp\left(-\frac{x}{\theta_i}\right). \tag{2}$$

We have $t_i(x) = \theta_i$ and $t(x) = \int \theta_i \mathrm{d}Q(i)$. If the agents are indexed by their absolute levels of risk tolerance we have $\theta_i = i$ and the log-supermodularity assumption is satisfied. We have then $t(x) = E^Q\left[\tilde{\theta}\right]$ and the collective level of risk tolerance does not depend on the wealth allocation among the agents. It is immediate that FSD shifts on the distribution of the individual levels of risk tolerance lead to an increase of the collective level of absolute (and relative) risk tolerance. More heterogeneity, in the sense of shifts in the distribution of preferences that preserve the mean, have no effect on the group's risk tolerance. These results will serve as a useful benchmark.

4 CRRA utility functions

Let us consider constant relative risk-aversion/tolerance utility functions of the form

$$u_i(x) = \frac{1}{1-\frac{1}{b_i}} x^{1-\frac{1}{b_i}}. \tag{3}$$

where b_i is the level of relative risk tolerance of individual i and $\frac{1}{b_i}$ is his level of relative risk aversion. In such a setting, we have

$$A_i = \frac{1}{b_i x}, \quad R_i = \frac{1}{b_i}, \quad t_i = b_i x, \quad \text{and} \quad s_i = b_i.$$

Since the utility functions are no more defined up to a multiplicative constant, we do not assume anymore that the $\lambda_i = 1$ for all i. However, we still assume that there exists

a wealth level x^* for which the fair allocation $(x_i)_{i\in I}$ with $x_i = x^*$ for all i, is efficient. Note that the existence of such a fair allocation can always be granted through a judicious choice of the weights $(\lambda_i)_{i\in I}$. The first-order conditions for Pareto optimality give then that $\lambda_i\,(x^*)^{1-\frac{1}{b_i}}$ is independent of i. The Pareto problem can be rewritten as follows $U(x) = \max_{\int \frac{x_i}{x^*}\mathrm{d}Q(i)=\frac{x}{x^*}} \int \frac{1}{1-\frac{1}{b_i}} \left(\frac{x_i}{x^*}\right)^{1-\frac{1}{b_i}} \mathrm{d}Q(i)$, up to a multiplicative constant. If we renormalize the individual consumptions to measure them in terms of multiples of x^*, we are led to analyze the situation where $u_i(x) = \frac{1}{1-\frac{1}{b_i}} x^{1-\frac{1}{b_i}}$ and all the weights λ_i are equal to 1. Note that with this renormalization, $x = 1$ corresponds to the efficient fair allocation.

In the next we consider then the equally weighted Pareto optimum. Since the agents differ by only one characteristic, namely their level b_i of relative risk tolerance, we might index them by this characteristic or we may, in other words, assume that $b_i = i$. For a given function h, we may then write indifferently $\int h(b_i)\mathrm{d}Q(i)$ or $\int h(b)\mathrm{d}Q(b)$ or $E^Q[h(\tilde{b})]$. The level of relative risk-aversion is then decreasing with i and the log-supermodularity condition is immediately satisfied.

The following Proposition uses these assumptions to characterize precisely the functions defining collective preferences and the collective level of risk aversion.

Proposition 4 *In a group made of agents with constant but heterogeneous levels of relative risk aversion, we have at the equally weighted Pareto optimum*

$$U(x) = \int \frac{b_i}{b_i - 1} e^{(1-b_i)\Phi^{-1}(x)} \mathrm{d}Q(i)$$

with

$$\Phi\,(t) \equiv \int e^{-b_i t} \mathrm{d}Q(i).$$

The collective degree of relative risk aversion $R(x)$ is given by

$$R(x) \equiv -x\frac{U''\,(x)}{U'\,(x)} = \frac{x}{\int b_i \mathrm{d}\widetilde{Q}(i)} = \frac{1}{s(x)}. \tag{4}$$

with $\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}Q} \equiv \frac{e^{-b\Phi^{-1}(x)}}{\int e^{-b_i\Phi^{-1}(x)}\mathrm{d}Q(i)}$.

As seen in the Proof in Appendix, the Lagrange multiplier of the Pareto optimum problem is given by $q = \exp(\Phi^{-1}(x))$ and q is then the shadow price associated to the constraint $\sum_{i\in I} x_i = x$. We clearly have $\Phi(0) = 1$, which means that $\Phi^{-1}(x^*) = 0$ and $q(x^*) = 1$ for the efficient fair allocation $x^* = 1$. Since the agents are risk averse, high levels of aggregate wealth have a low shadow price and low levels of aggregate wealth have high shadow price and we can easily derive that $q(x) < 1$ for $x > x^*$and $q(x) > 1$ for $x < x^*$. This means, in particular, that we have $\Phi^{-1}(x) < 0$ for $x > 1$ and $\Phi^{-1}(x) > 0$ for $x < 1$.

4.1 A model with two agents

Mazzocco (2004) shows that, in a model with two agents, an increase in the level of risk tolerance of one of the agents might have an ambiguous impact on the collective level of risk tolerance. It increases for some levels of aggregate wealth while it decreases for other levels of aggregate wealth. Since this is only stated on a numerical example in Mazzocco (2004), let us clearly express this result.

Proposition 5 *In a model with two agents with $b_1 < b_2$, there exists $\overline{x} \leq 1$ such that a small increase of b_2 leads to an increase of the collective level of risk tolerance $t(x)$ for all $x \geq \overline{x}$ and to a decrease of the collective level of risk tolerance $t(x)$ for all $x \leq \overline{x}$.*

Recall that after our normalization, $x^* = 1$ corresponds to the fair efficient allocation. Proposition 5 means then that an increase of the risk tolerance level of the most risk tolerant agent increases (decreases) the collective level of risk tolerance for levels of wealth above (below) a given threshold that is below the fair efficient allocation. Note that the threshold $\overline{x}$ depends on b_1 and b_2. This means that for x above the fair allocation, any increase of b_2 increases the collective risk tolerance. In fact, above the fair allocation, an increase of b_2 also increases the weight granted to b_2 leading to an increase of the collective level of risk tolerance. For a wealth level x below the fair allocation, the impact of an increase of b_2 is less clear. Indeed, we showed in Proposition 3 that for low levels of aggregate wealth the least risk tolerant agent (the most risk averse) has a larger share of the total wealth and an increase of b_2 leads to an increase of the weight granted to b_1 (the share of total wealth of agent 1). The increase of b_2 has then two effects in opposite directions: an increase of one of the terms of the average (namely the greatest one) and an increase of the weight of the smallest one. Since the second effect does not exist for $x = 1$, the first effect continues to dominate for x above a given threshold $\overline{x} \leq 1$ while the second effect dominates for $x \leq \overline{x}$.

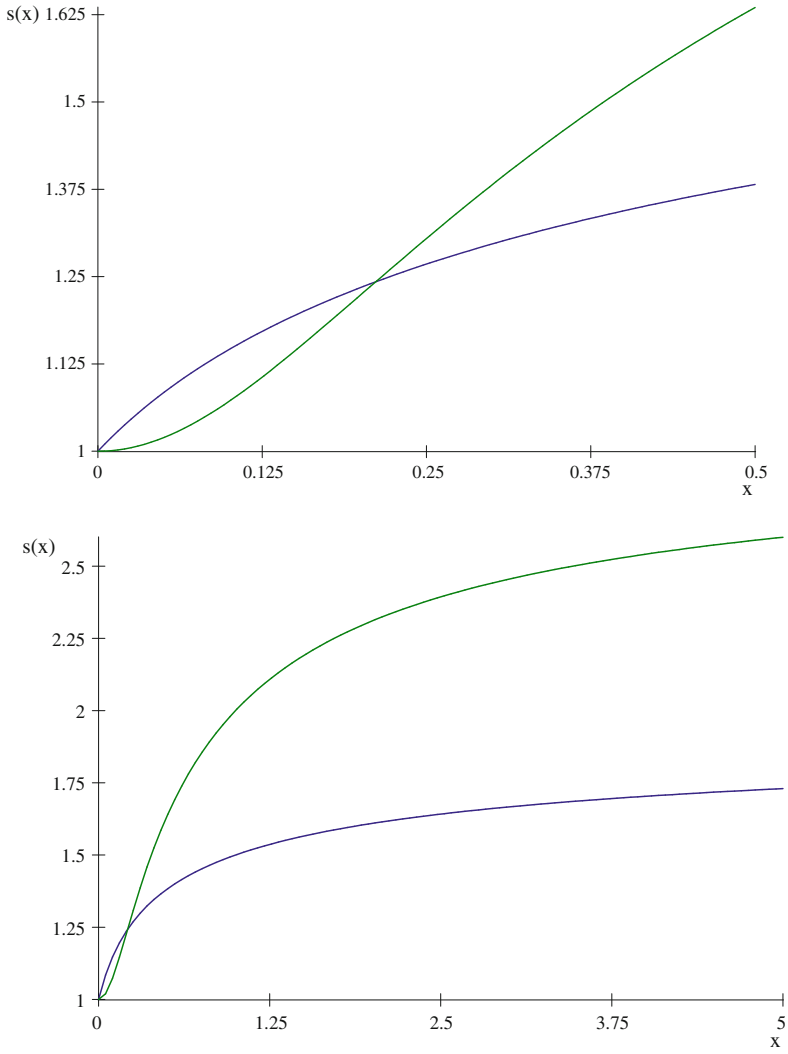
Figures 1–2 illustrate the impact of an increase of b_2.

The next Proposition analyzes more in detail how the collective level of risk tolerance evolves as a function of b_2.

Proposition 6 *In a model with two agents with $b_1 < b_2, b_1$ being given, for $x \geq \frac{1}{2}$, the function $b_2 \to t_x(b_2)$ is increasing on (b_1, ∞) with $\lim_{b_2 \to b_1} t_x(b_2) = xb_1$ and $t_x(b_2) \sim_{b_2 \to \infty} \left(x - \frac{1}{2}\right) b_2$. For $x < \frac{1}{2}$, there exists $b^*(x, b_1) > b_1$ such that the function $b_2 \to t_x(b_2)$ is increasing on $(b_1, b^*(x, b_1))$ and decreasing on $(b^*(x, b_1), \infty)$ with $\lim_{b_2 \to b_1} t_x(b_2) = xb_1$ and $\lim_{b_2 \to \infty} t_x(b_2) = xb_1$.*

In summary, for (very) low levels of wealth, increasing the risk tolerance of the more risk tolerant agent has an ambiguous impact on the collective attitude towards risk. Proposition 6 characterizes precisely the conditions for this paradoxical result to occur. Figures 3, 4, and 5 illustrate the different possible shapes for $b_2 \to t_x(b_2)$ (or equivalently of $b_2 \to s_x(b_2)$). The asymptotic behavior of $t_x(b_2)/b_2$ is illustrated in Figs. 6 and 7.

Propositions 5 and 6 establish the behavior of collective preferences as a function of one of the agent's risk tolerance. Another important question is what happens



Figs. 1–2 At two different scales, we represent the collective relative risk tolerance as a function of the total wealth, with $b_1 = 1$ and $b_2 = 2$ for the *curve* with a higher (lower) level of collective relative risk tolerance for low (high) levels of wealth and $b'_1 = 1$ and $b'_2 = 3$ for the other *curve*. *Both curves* converge slowly to the associated level of relative risk tolerance as can be shown on the second figure. An increase of the risk tolerance level of the most risk tolerant agent leads to an increase (decrease) of the collective level of risk tolerance above (below) a given threshold $x^* \leq 1$. With our parameters, we have $x^* = 0.21$. When $b_2 = 2$ and b'_2 is in the neighborhood of b_2, we have $x^* = 0.14$. When $b'_2 = 3$ and b_2 is in the neighborhood of b'_2, we have $x^* = 0.27$

to the collective level of risk tolerance when *both* agents become more risk tolerant. In the next Proposition we show that the ambiguous impact disappears when we consider a uniform increase of risk tolerance across the agents, but that for non-uniform increases in risk tolerance the collective degree of risk tolerance may still be lower.

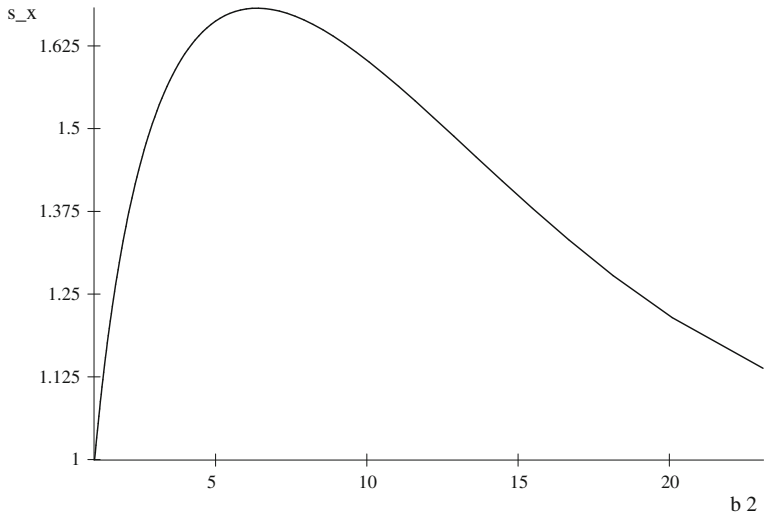


Fig. 3 In a setting with two agents with levels of relative risk tolerance b_1 and b_2, we represent the collective relative risk tolerance $s_x(b_2)$ as a function of b_2 for $b_1 = 1$ and for $x = 0.4$. For $b_2 = b_1 = 1$, s_x is equal to 1. The collective relative risk tolerance increases then decreases with b_2 and $\lim_{b_2 \to \infty} s_x(b_2) = b_1$

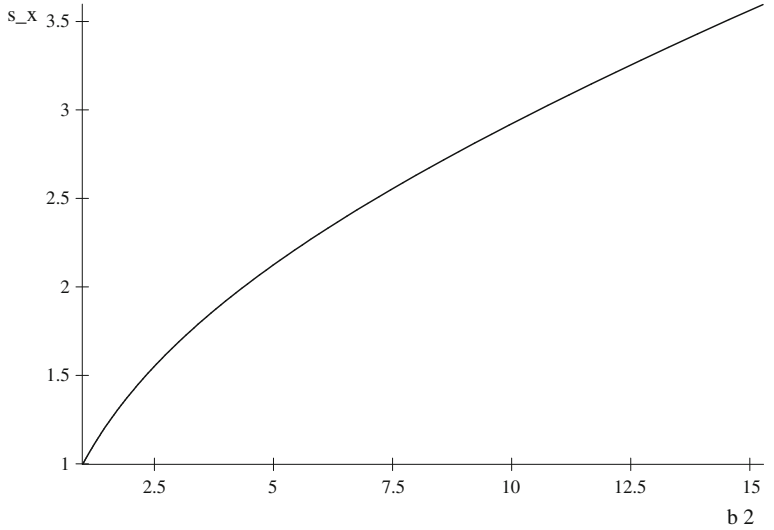


Fig. 4 In a setting with two agents with levels of relative risk tolerance b_1 and b_2, we represent the collective relative risk tolerance $s_x(b_2)$ as a function of b_2 for $b_1 = 1$ and for $x = 0.55$. For $b_2 = b_1 = 1$, s_x is equal to 1. The collective relative risk tolerance increases with b_2

Proposition 7 *Let $b_1 < b_2$ be given.*

1. *Let us consider a uniform increase of the individual levels of risk tolerance of the form $b_1 + h$ and $b_2 + h$ with $h > 0$. The associated level of collective risk tolerance $t_x(h)$ increases with h for all x.*

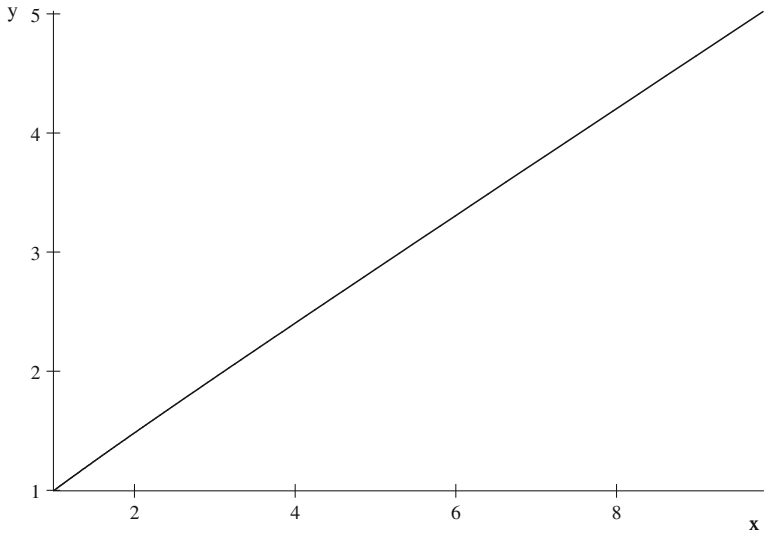


Fig. 5 In a setting with two agents with levels of relative risk tolerance b_1 and b_2, we represent the aggregate relative risk tolerance $s_x(b_2)$ as a function of b_2 for $b_1 = 1$ and for $x = 0.9$. For $b_2 = b_1 = 1$, s_x is equal to 1. The aggregate relative risk tolerance increases with b_2 and is almost linear

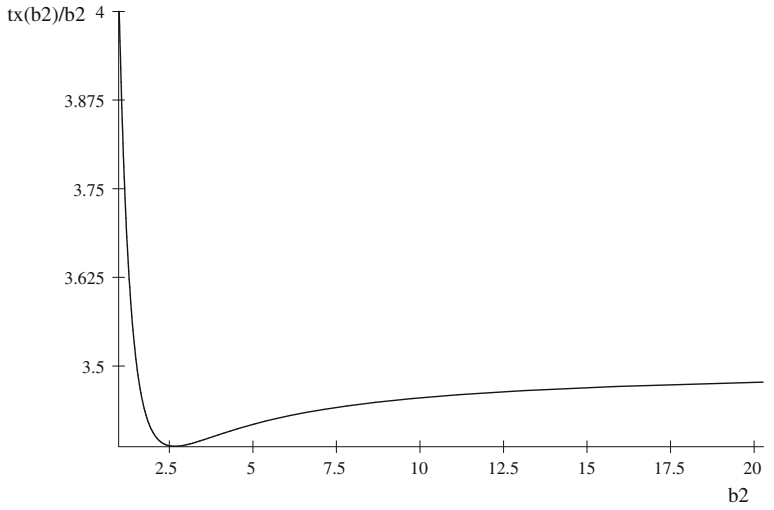


Fig. 6 In a setting with two agents with levels of relative risk tolerance b_1 and b_2, we represent the ratio $t_x(b_2)/b_2$ for $x = 4$ and $b_1 = 1$. For $b_2 = b_1 = 1$ it is immediate that $t_x(b_2)/b_2 = t_x(b_2) = x$. The ratio converges to $x - 0.5 = 3.5$ when b_2 converges to ∞

2. *Let $k > 0$ be given and let us consider an increase of the individual levels of risk tolerance of the form $b_1 + kh$ and $b_2 + h$ with $h > 0$. The associated level of collective risk tolerance $t_x(h)$ increases with h for $x \leq 1$ if $k \geq \frac{b_1}{b_2}$ and increases with h for $x \geq 1$ if $k \leq 1$. In particular, $t_x(h)$ increases with h for all x if $k \in [\frac{b_1}{b_2}, 1]$.*

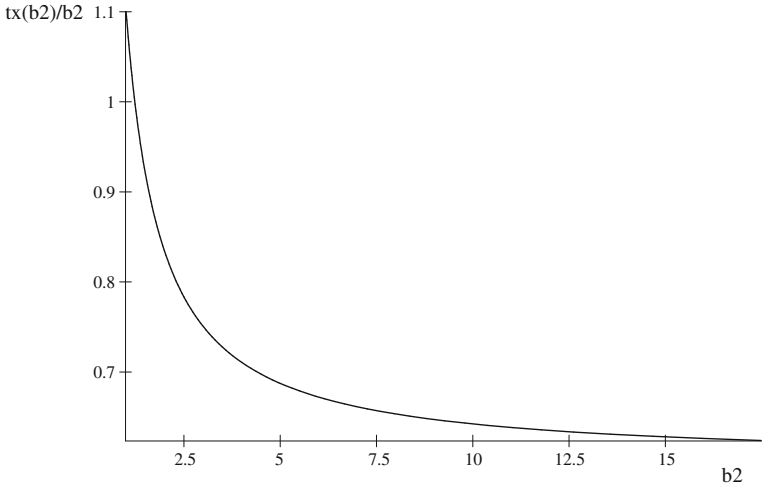


Fig. 7 In a setting with two agents with levels of relative risk tolerance b_1 and b_2, we represent the ratio $t_x(b_2)/b_2$ for $x = 1.1$ and $b_1 = 1$. For $b_2 = b_1 = 1$ it is immediate that $t_x(b_2)/b_2 = t_x(b_2) = x$. The ratio converges to $x - 0.5 = 0.6$ when b_2 converges to ∞

Let us illustrate the second point by two extreme situations. For k very small (near to 0), the shifts we are considering are almost of the form $(b_1, b_2) \to (b_1, b_2 + \varepsilon)$ that have already been considered in Proposition 5. These shifts increase the risk tolerance level of the second agent and also increase its weight for $x \geq 1$. These shifts lead then to an unambiguous increase of the aggregate level of risk tolerance for $x \geq 1$. For k near infinity (and h very small), the shifts we are considering are almost of the form $(b_1, b_2) \to (b_1 + \varepsilon, b_2)$ and such shifts increase the risk tolerance level of the first agent and also increase its weight when $x \leq 1$. These shifts lead then to an unambiguous increase of the aggregate level of risk tolerance for $x \leq 1$. The proposition shows that there is a range for k for which the shifts have an unambiguous impact without restrictions on x. However, for small levels of k we cannot conclude that the collective level of risk tolerance is higher.[4]

We are also interested in the impact of more heterogeneity among our two agents. The next result shows that more heterogeneity leads to a higher collective risk tolerance level for high wealth levels (above the fair efficient allocation) and to a lower collective risk tolerance level for low wealth levels (below the fair efficient allocation).

Proposition 8 *Let b_1 and b_2 be given with $b_1 < b_2$ and let us consider a shift of the form $b_1 - h$ and $b_2 + h$ with $h > 0$. The associated level of collective risk tolerance $t_x(h)$ increases (resp. decreases) with h for $x \geq 1$ (for $x \leq 1$).*

This result is very intuitive. We have already seen that the collective level of risk tolerance is near the risk tolerance level of the most risk tolerant agent for high levels of wealth and is near the risk tolerance level of the least risk tolerant level agent for low

[4] This should not be 'too' surprising given our previous results since for $k = 0$ we are back in the situation in which only b_2 increases.

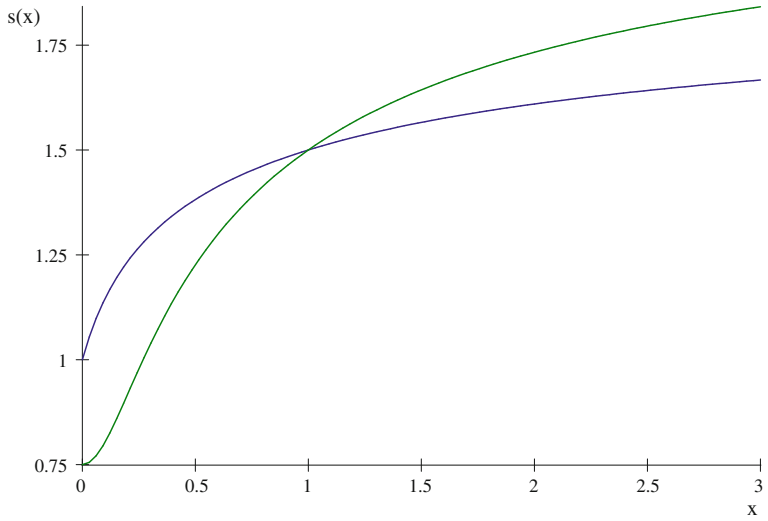


Fig. 8 In a setting with two agents with levels of relative risk tolerance b_1 and b_2, we represent the collective relative risk tolerance as a function of the total wealth with $b_1 = 1$ and $b_2 = 2$ for the *curve* with a higher (lower) level of collective relative risk tolerance for low (high) levels of wealth and $b'_1 = \frac{3}{4}$ and $b'_2 = \frac{9}{4}$ for the other *curve*. The shift from (b_1, b_2) to (b'_1, b'_2) is a Mean Preserving Spread as in Proposition 8 and the *two curves* cross at $x^* = 1$

levels of wealth. More heterogeneity leads to an increase of the risk tolerance level of the most risk tolerant agent and to a decrease of the risk tolerance level of the least risk tolerant agent. This leads then to an increase of the collective level of risk tolerance for high levels of wealth and to a decrease of the collective level of risk tolerance for low levels of wealth. Proposition 8 permits to give a precise meaning to high and low levels of wealth since the fair efficient allocation appears to be the relevant threshold. Figure 8 illustrates this result.

4.2 General populations

The different results of the previous section permit to see that first-order stochastic dominance shifts do not guarantee an increase in the collective degree of risk tolerance. In this section we initially consider whether a stronger notion of first-order stochastic dominance leads to an unambiguous impact on the group's degree of risk tolerance. Then we evaluate the effect of more heterogeneity within a group. In order to treat the problem in a quite general setting, we consider from now on general populations described by a distribution on $I = [0, \infty)$. To relate the results in this more general setting to those obtained in the 2-agent framework, we will attach a specific attention to distributions with a 2-point support.

Since FSD is not a good candidate to obtain comparative static results, let us recall the following definition corresponding to a stronger notion of first-order dominance.

Definition 1 Monotone Likelihood Ratio Dominance (MLR). Let P and Q denote two probability measures on $I = [0, \infty)$. We say that P dominates Q in the sense of MLR

$(P \succcurlyeq_{\mathrm{MLR}} Q)$ if there exist numbers $0 \leq \alpha \leq \beta \leq \infty$ and a nondecreasing function $h : [\alpha, \beta] \to [0, \infty]$ such that $P([0, \alpha)) = Q((\beta, \infty)) = 0$ and $\mathrm{d}P(i) = h(i)\mathrm{d}Q(i)$.

In other words, an MLR-dominated shift for a given probability measure puts less weight for higher values of i. This concept is widely used in the statistical literature and was first introduced in the context of portfolio problems by Landsberger and Meilijson (1990). MLR dominance is stronger than FSD and, in particular, an MLR-dominated shift for a given distribution reduces the mean.

When the supports of P and Q are reduced to two points, (b_1^1, b_2^1) with $b_1^1 < b_2^1$ for P and (b_1^2, b_2^2) with $b_1^2 < b_2^2$ for Q, we necessarily have $b_1^2 < b_2^2 \leq b_1^1 < b_2^1$ or $b_1^2 = b_1^1$ and $b_2^2 = b_2^1$. In the first case, any average of the risk tolerance levels in the support of Q is smaller than any average of the risk tolerance levels in the support of P and the collective risk tolerance level is higher under P. The (most) interesting case is when both probability measures have the same support. We then have two populations with the same set of possible levels of individual level of risk tolerance b_1 and b_2 but with different proportions of agents in each category: a proportion p_1 (resp. $p_2 = 1 - p_1$) of agents that have an individual level b_1 (resp. b_2) of risk tolerance under P and a proportion q_1 (resp. $q_2 = 1 - q_1$) of agents that have an individual level b_1 (resp. b_2) of risk tolerance under Q. The MLR dominance is characterized in this setting by $\frac{p_2}{p_1} \geq \frac{q_2}{q_1}$ (or equivalently $q_1 \geq p_1$ or $q_2 \leq p_2$).

Proposition 9 *Let us consider two populations. In the first one, we have proportions p_1 (resp. $p_2 = 1 - p_1$) of agents with an individual level b_1 of risk tolerance (resp. b_2). In the second one, we have proportions q_1 (resp. $q_2 = 1 - q_1$) of agents with an individual level b_1 of risk tolerance (resp. b_2). If we assume that $\frac{p_2}{p_1} \geq \frac{q_2}{q_1}$, then the collective level of risk tolerance is higher in the first population.*

In summary, when the support of the population is reduced to two points, an MLR dominant shift in the degree of risk tolerance of the members of the group increases the group's risk tolerance, so such shift clearly characterizes the notion of a "more risk tolerant group". The following proposition generalizes the impact of MLR shifts for distributions with more general supports.

Proposition 10 *Let us consider two populations characterized by two distributions P and Q of individual levels of risk tolerance. If $P \succcurlyeq_{\mathrm{MLR}} Q$ then for all $x \leq 1$, we have $t_x^P \geq t_x^Q$. However, $P \succcurlyeq_{\mathrm{MLR}} Q$ does not guarantee an increase of the collective level of risk tolerance when $x > 1$.*

MLR provides then a satisfying answer to the impact of shifts for low levels of wealth (when $x \leq 1$), which corresponds to the case where the unilateral increase of one of the individual levels of risk tolerance failed to guarantee an increase of the aggregate level of risk tolerance. In the following Proposition we show that when the density function (introduced in Definition 1) $h = \frac{\mathrm{d}P}{\mathrm{d}Q}$ has an exponential growth rate, then we do have an unambiguous impact on collective risk tolerance.

Proposition 11 *Let us consider two populations characterized by two distributions P and Q on $[0, \infty)$of individual levels of risk tolerance such that $P \succcurlyeq_{\mathrm{MLR}} Q$ with $\frac{\mathrm{d}P}{\mathrm{d}Q}(b) = \lambda \exp(kb)$ for some positive k and λ. For all x, we have $t_x^P \geq t_x^Q$.*

We have seen in the 2-agent setting (Proposition 8) that more heterogeneity has a clear impact on the collective level of risk tolerance depending on the relative position of the aggregate wealth with respect to the fair efficient allocation. We are now interested in establishing the effect of "more heterogeneity" in the general setting. For this purpose, we introduce the following definition.

Definition 2 Portfolio Dominance (PD). Let Q_1 and Q_2 denote two probability measures on $I = [0, \infty)$. We say that Q_1 dominates Q_2 in the sense of PD ($Q_1 \succcurlyeq_{\text{PD}} Q_2$) if we have $\int v(i-a)\mathrm{d}Q_1(i) = 0 \Rightarrow \int v(i-a)\mathrm{d}Q_2(i) \leq 0$ for any real number a and any nonnegative and nonincreasing function v.

This concept has been introduced in the context of portfolio problems by Landsberger and Meilijson (1990) and further studied by Gollier (1997). In the portfolio context it is related to the degree of riskiness of the asset returns. In our context, it is related to the level of individual heterogeneity in relative risk tolerance. In particular, a mean preserving PD-dominated shift for a given distribution increases the variance (Jouini and Napp 2008, Proposition 3).

The following proposition uses this concept to characterize the impact of "more heterogeneity" in the distribution of individual preferences.

Proposition 12 *Let us consider two populations, respectively, characterized by distributions P and Q of individual levels of risk tolerance. If P and Q are symmetric with respect to some b^* with $\frac{\mathrm{d}Q}{\mathrm{d}P}$ nonincreasing before b^* and nondecreasing after b^* then $P \succcurlyeq_{\text{PD}} Q$ and $P \succcurlyeq_{\text{SSD}} Q$ and the collective level of risk tolerance t_x^Q under Q is higher than (resp. lower than) the aggregate level of risk tolerance t_x^P under P for $x \geq 1$ (for $x \leq 1$).*

The intuition is, in essence, the same as that in Proposition 8. At high wealth levels those individuals that have a high tolerance for risk are more representative of the collective level of risk tolerance, but the opposite is true at low wealth levels. More dispersion in the levels of risk tolerance of the group then leads to a higher (lower) level of collective risk tolerance for high (low) wealth levels.

5 The case of egalitarian groups

It is interesting to analyze the aggregate behavior in a model where all the agents consume the total consumption x. This is the case when x is a public good. This is also the case when x is a private good but simultaneously consumed by all the agents in the group. We may consider that both agents in a couple get utility from saving money or from holding consumption goods and consider these goods as owned by the couple and not shared among them through an efficient sharing rule. This is also the setup used in a number of recent experiments that compare the degree of risk aversion of groups with that of individuals (e.g., Shupp and Williams 2008; Baker et al. 2008; Masclet et al. 2009) and where the rewards of the group are exogenously divided equally among its members. We may imagine, for example, that such experiments reflect the widely observed regularity of partnerships with equal sharing rules.

In the next, the endowment in the consumption good is defined by a random variable x on the probability space (Ω, F, P) and the social utility function is given by

$$U(x) = \int \lambda_i u_i(x) \mathrm{d}Q(i). \tag{5}$$

where u_i is the utility function of agent i and λ_i is the weight granted to agent i.

We have then

$$R(x) \equiv -x\frac{U''(x)}{U'(x)} = -x\frac{\int \lambda_i u_i''(x)\mathrm{d}Q(i)}{\int \lambda_i u_i'(x)\mathrm{d}Q(i)} = \int R_i(x)\mathrm{d}P^u(i) \tag{6}$$

where P^u is the probability measure defined by $\frac{\mathrm{d}P^u}{\mathrm{d}Q}(i) = \frac{\lambda_i u_i'(x)}{\int \lambda_i u_i'(x)\mathrm{d}Q(i)}$.

Let us first analyze the case with two agents and CRRA functions. We have then $u_i(x) = \frac{1}{1-\frac{1}{b_i}} x^{1-\frac{1}{b_i}}, i = 1, 2$, and we take $\lambda_i = 1, i = 1, 2$, as in the previous section. Equation 6 can be rewritten as follows

$$R(x) = \frac{\frac{1}{b_1}x^{-\frac{1}{b_1}} + \frac{1}{b_2}x^{-\frac{1}{b_2}}}{x^{-\frac{1}{b_1}} + x^{-\frac{1}{b_2}}} = \frac{R_1 x^{-R_1} + R_2 x^{-R_2}}{x^{-R_1} + x^{-R_2}}$$

and the aggregate relative risk aversion is a weighted arithmetic average of the individual levels of relative risk aversion. As in the private good case, the collective level of relative risk aversion decreases with x and it approaches the degree of relative risk aversion of the most (least) risk averse agent in the economy as x converges to zero (infinity). Notice also that the weights are given by the individual marginal utilities and the highest weight is granted to the lowest (highest) individual level relative risk aversion for $x > 1$ (for $x < 1$). An increase of the individual level of risk tolerance of the most risk tolerant agent might then have an ambiguous impact. We have the following result:

Proposition 13 *In a model with a public good, two agents and CRRA functions with $b_1 < b_2$, there exists $\overline{x} \leq 1$ such that a small increase of b_2 leads to an increase of the collective level of risk tolerance $t(x)$ for all $x \geq \overline{x}$ and to a decrease of the collective level of risk tolerance $t(x)$ for all $x \leq \overline{x}$.*

In particular, this means that FSD shifts of the individual levels of risk tolerance are not sufficient to increase the collective level of risk tolerance. The next result illustrates the impact of a mean preserving spread on the individual levels of risk aversion in a 2-agent setting.

Proposition 14 *In a model with a public good, two agents and CRRA functions with $b_1 < b_2$ (or equivalently $R_1 > R_2$), a shift of the form $R_1 - h$ and $R_2 + h$ with $h > 0$ increases (decreases) the aggregate level of relative risk aversion $R(x)$ for $x \leq 1$ (for $x \geq 1$). It increases (decreases) the collective level of risk tolerance $t(x)$ for $x \geq 1$ (for $x \leq 1$).*

In the next we characterize, in a general distribution setting, the impact of MLR shifts on the collective level of risk tolerance/aversion. Note that the following result is obtained for very general utility functions.

Proposition 15 *Let us consider two populations, respectively, characterized by distributions P_1 and P_2 on (I, ι). Under Assumptions (U) and (LSPM) and if $P_2 \succcurlyeq_{\text{MLR}} P_1$ then the collective level of risk aversion (risk tolerance) under P_1 is higher (lower) than under P_2.*

The result in Proposition 15 is quite powerful. Under the assumption that the members of the group consume the same endowment, and under weak restrictions on the individual utility functions, if the individual levels of risk aversion are more concentrated on high values (in the sense of MLR dominance) then the collective level of risk aversion will be higher for all endowment levels. In this case MLR dominance provides a clear characterization of comparative collective risk aversion.[5]

The next proposition analyzes the impact of more heterogeneity on the individual levels of risk aversion. For a given distribution P on (I, ι) and for a given x, we denote by P^x the image measure of P by $i \to -\frac{u_i''(x)}{u_i'(x)}$. The measure P^x describes the distribution of the individual levels of risk aversion at a given wealth level x.

Proposition 16 *Let us consider two populations, respectively, characterized by distributions P_1 and P_2 on (I, ι) and let us assume that (U) is satisfied. If, for a given x, $u_i'(x)$ is nondecreasing in i, $-\frac{u_i''(x)}{u_i'(x)}$ is decreasing with i[6] and $P_2^x \succcurlyeq_{\text{PD}} P_1^x$, then the collective level of risk aversion (risk tolerance), at x, under P_2 is higher (lower) than under P_1.*

In particular, if there exists x^* such that all the individual marginal utilities $u_i'(x^*)$ are equal, then $u_i'(x)$ is nondecreasing in i for $x \leq x^*$. It suffices then to have that $-\frac{u_i''(x)}{u_i'(x)}$ is decreasing with i and $P_2^x \succcurlyeq_{\text{PD}} P_1^x$ to conclude that the aggregate level of risk aversion, at x, under P_2 is higher than under P_1. In other words, under weak assumptions on the utility function, less heterogeneity in risk aversion, in the sense of PD dominance, implies a higher level of collective risk aversion when wealth is sufficiently low.

6 Conclusion

Mazzocco (2004) established the counter-intuitive result that an increase in the level of risk tolerance of one of the individuals in a couple may reduce their collective degree of risk tolerance. We studied precisely the conditions for this phenomenon to occur. More generally, we established conditions under which groups with individual

[5] Note that Proposition 15 can easily be extended for higher order collective preferences towards risk. For example, if we assume that $u''(x, i)$ is LSPM in (x, i), then an MLR-dominant shift decreases the collective degree of absolute (and relative) prudence.

[6] Note that this last condition is just a little bit stronger than the LSPM condition.

levels of risk tolerance more concentrated on high values and groups that are more heterogeneous will display higher risk tolerance, both with efficient risk-sharing and with an exogenous egalitarian sharing rule. Our results permit to better characterize differences in risk taking behavior between groups and individuals and among groups with different distributions of risk preferences.

It should be possible to design experiments to evaluate if our results are consistent with elicited risk attitudes of groups and individuals. Shupp and Williams (2008) compare the willingness to pay for lotteries of small groups and individuals in a setup similar to that of Sect. 5. They conclude that, for most lotteries, group choices are significantly different from the mean of the individual choices (groups tend to be more risk averse than individuals for low-expected-value lotteries but less risk averse than individuals for high-expected-value lotteries). These results are consistent with a large number of studies in social psychology that show "risky" and "cautious" shifts in group risk-taking behavior relative to the mean of the individual choices (see e.g., Clark 1971). We have seen that there is no reason to believe that the group's willingness to pay (derived from the collective preferences), and more generally the willingness to take risks, should be the same as the mean of the individual members' willingness to pay. In particular, even with CRRA individual preferences, the fact that the group's relative risk aversion decreases with x implies that "cautious" shifts should be more prevalent in low-expected-value (low-stakes) lotteries while "risky" shifts should be more prevalent in high-expected-value (high-stakes) lotteries, precisely what Shupp and Williams (2008) found.[7] It would be interesting to further explore the experimental relevance of our results by proceeding to inter-groups comparisons. For instance, to analyze the difference in risk attitudes between two couples that only differ by the risk aversion level of one of the members, e.g., the man.

Acknowledgments We thank the participants at the FUR XIV conference for their comments. Jouini and Napp acknowledge the financial support of the ANR (Risk Project) as well as of the Fondation du Risque (Groupama Chair).

Appendix

Proof of Proposition 1 Let us denote by φ the function defined by $\varphi(q) = \int \left(u_i'\right)^{-1}\left(\frac{q}{\lambda_i}\right)\mathrm{d}Q(i)$. By Inada's conditions and since u_i is increasing and strictly concave for all i, φ is well defined on $(0,\infty)$ and decreasing. Furthermore, from the monotone convergence Theorem we have $\lim_{x\to 0}\varphi(q) = \infty$ and $\lim_{x\to\infty}\varphi(q) = d$. Let us then define $f_i : [d,\infty) \to [d_i,\infty)$ by $f_i(x) = \left(u_i'\right)^{-1}\left(\frac{\varphi^{-1}(x)}{\lambda_i}\right)$, we clearly have $x = \int f_i(x)\mathrm{d}Q(i)$ and $\lambda_i u_i'(f_i(x)) = \varphi^{-1}(x)$ for all i and is independent of i. The family $(f_i(x))$ satisfies then the first-order conditions of the maximization program defined by Eq. 1 and since this program is concave we have $U(x) = \int \lambda_i u_i(f_i(x))\mathrm{d}Q(i)$. □

[7] Eliaz et al. (2006) show that risky and cautious shifts in groups can be seen as a failure of expected utility theory.

Proof of Proposition 3 Since the fair allocation $x_i = x^*, i \in I$, is efficient and since we granted the same weight to all the agents, the first-order conditions for Pareto optimality give us that $u_i'(x)$ is independent of i. Since $\frac{\partial u}{\partial x}(x, i)$ is LSPM, integrating $\frac{\partial}{\partial x} \log u_i'(x)$ between $x \geq x^*$ and x^*, gives us that $u_i'(x)$ increases with i. Let us consider the Pareto allocation $(x_i)_{i \in I}$ associated to the aggregate wealth $x \geq x^*$. We have that $u_i'(x_i)$ is independent of i. We also necessarily have $x_{i_0} \geq x^*$ for some i_0 and then $u_{i_0}'(x_{i_0}) \leq u_{i_0}'(x^*)$ by concavity of the utility functions. We have then $u_i'(x_i) \leq u_i'(x^*)$ for all i and consequently $x_i \geq x^*$ for all i. Since $u_i'(x)$ increases with i and $u_i'(x)$ is independent of i and since $u_i'(x)$ is decreasing in x, we have that x_i increases with i. By the LSPM property $t(x, i)$ is increasing in i and by the DARA property, it is increasing in x. We have then that $t(x_i, i)$ is increasing in i.

For $x \leq x^*$, integrating $\frac{\partial}{\partial x} \log u_i'(x)$ between x and x^*, gives us that $u_i'(x)$ decreases with i. The same kind of arguments as above give that x_i decreases with i. □

Proof of Proposition 4 The first-order condition gives us the optimal allocation of agent $i, x_i = q^{-b_i}$, where q is the Lagrange multiplier. Using the resource constraint we obtain $\int q^{-b_i} \mathrm{d}Q(i) = x$ or $\int \exp(-b_i \ln q) \mathrm{d}Q(i) = x$. We obtain that $q = \exp(\Phi^{-1}(x))$ hence $x_i = \mathrm{e}^{-b_i \Phi^{-1}(x)}$. Plugging this back into the utility function we obtain the representative agent's utility.

The degree of relative risk aversion of the representative agent can be derived directly from the expression of the representative agent utility function or using Proposition 2. We have $s = \int s_i \frac{x_i}{x} \mathrm{d}Q(i)$. Since $x_i = \mathrm{e}^{-b_i \Phi^{-1}(x)}$ the result follows. □

Proof of Proposition 5 Let us consider b_1 as given and denote by $t_x(b_2)$ the aggregate risk tolerance level when the risk tolerance level of agent 2 is given by $b_2 \geq b_1$. We have

$$t_x(b_2) = \frac{1}{2} b_1 \exp(-b_1 a_x(b_2)) + \frac{1}{2} b_2 \exp(-b_2 a_x(b_2))$$

where $a_x(b_2)$ is the solution of

$$\psi(a_x(b_2), b_1, b_2) = x$$

with

$$\psi(a, b_1, b_2) = \frac{1}{2} \exp(-b_1 a) + \frac{1}{2} \exp(-b_2 a).$$

We get after computations

$$\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) = \frac{1}{2} \frac{\exp\left(-(b_1 + b_2) a_x(b_2)\right)}{b_1 \exp(-b_1 a_x(b_2)) + b_2 \exp(-b_2 a_x(b_2))} \varphi(a_x(b_2), b_1, b_2)$$

with

$$\varphi(a, b_1, b_2) = b_1 + b_2 \exp((b_1 - b_2)\, a) - a\, (b_2 - b_1)\, b_1.$$

For $x \geq 1$, we have $a_x(b_2) \leq 0$ and $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) > 0$. The aggregate level of risk tolerance is an increasing function of b_2.

Let us now focus on the case $x \leq 1$. It is easy to check that $a_x(b_2)$ is always positive for $x \in (0, 1)$ and decreases with x from ∞ to 0. It is also easy to see that $\varphi(a, b_1, b_2)$

is decreasing in a, positive for $a = 0$ and converges to $-\infty$ when a converges to ∞. There exists then a level $x^* < 1$ such that $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) = 0$. For $x < x^*$ (resp. $x > x^*$), we have $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) < 0$ (resp. $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) > 0$). □

Proof of Proposition 6 We have

$$t_x(b_2) = \frac{1}{2} b_1 \exp(-b_1 a_x(b_2)) + \frac{1}{2} b_2 \exp(-b_2 a_x(b_2))$$

where $a_x(b_2)$ is the solution of

$$\frac{1}{2} \exp(-b_1 a_x(b_2)) + \frac{1}{2} \exp(-b_2 a_x(b_2)) = x.$$

We know that

$$\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) = \varphi(a_x(b_2), b_1, b_2) = b_1 + b_2 \exp((b_1 - b_2)\, a_x(b_2)) - a_x(b_2)\,(b_2 - b_1)\, b_1.$$

We have already seen that $\varphi(a, b_1, b_2)$ is decreasing in a, positive for $a = 0$ and converges to $-\infty$ when a converges to ∞. Let us denote by $a(b_1, b_2)$ the solution of $\varphi(a, b_1, b_2) = 0$. Let us first consider the case $x \leq \frac{1}{2}$. The function $a_x(b_2)$ is clearly decreasing with b_2 and we have $\lim_{b_2 \to b_1} a_x(b_2) = -\frac{\ln x}{b_1}$ and $\lim_{b_2 \to \infty} a_x(b_2) = -\frac{\ln 2x}{b_1}$. Furthermore, since $\frac{\partial \varphi}{\partial a}$ is decreasing, $\frac{\partial a}{\partial b_2}$ has the same sign as $\frac{\partial \varphi}{\partial b_2}$. Direct computations give $\frac{\partial \varphi}{\partial b_2} = u - ab_1 - ab_2 u$ with $u = \exp((b_1 - b_2)\, a)$ and a such that $b_1 + b_2 u - b_1 a(b_2 - b_1) = 0$. Substituting a in $\frac{\partial \varphi}{\partial b_2}$ gives $\frac{\partial \varphi}{\partial b_2} = (b_1 - b_2)^{-1} b_1^{-1} \left(b_2 b_1 u + b_1^2 + b_1^2 u + b_2^2 u^2\right) < 0$. The function $a(b_1, b_2)$ decreases then with b_2 and we have $\lim_{b_2 \to b_1} a(b_1, b_2) = \infty$ and $\lim_{b_2 \to \infty} a_x(b_2) = 0$. There exists then b_2^* such that $a_x(b_2) = a(b_1, b_2)$ and $\frac{d}{db_2} t_x(b_2) = 0$.

It is immediate that for $a_x(b_2) > a(b_1, b_2)$ we have $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) < 0$ and for $a_x(b_2) < a(b_1, b_2)$ we have $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2) > 0$. It suffices to show that $a_x(b_2)$ and $a(b_1, b_2)$ cross only once to establish the result. Let us consider b_2^* such that $a_x(b_2) = a(b_1, b_2)$ and let us compute $\frac{\mathrm{d}}{\mathrm{d}b_2} (a_x(b_2) - a(b_1, b_2))$ at b_2^*. By definition, we have $a(b_1, b_2^*) = a_x(b_2^*)$ and we denote it by a^*. Direct computations give

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}b_2} \left(a_x(b_2^*) - a(b_1, b_2^*)\right) &= \frac{\frac{\partial \varphi}{\partial b_2}}{\frac{\partial \varphi}{\partial a}}(a^*, b_1, b_2^*) - \frac{\frac{\partial \psi}{\partial b_2}}{\frac{\partial \psi}{\partial a}}(a^*, b_1, b_2^*) \\ &= \frac{A}{\left(b_1 \mathrm{e}^{-a^* b_1} + b_2^* \mathrm{e}^{-a^* b_2^*}\right)\left(b_1 + b_2^* \mathrm{e}^{a^*(b_1 - b_2^*)}\right)(b_2^* - b_1)}. \end{aligned}$$

with

$$\begin{aligned} A = \Big(&a^* b_1^2 \mathrm{e}^{-a^* b_1} + a^* b_1^2 \mathrm{e}^{-a^* b_2^*} + (a^* b_1 b_2^* \mathrm{e}^{-a^* b_1} + a^* b_1 b_2^* \mathrm{e}^{-a^* b_2^*} - b_1 \mathrm{e}^{-a^* b_1} \\ &- b_2^* \mathrm{e}^{-a^* b_2^*}) \mathrm{e}^{a^*(b_1 - b_2^*)}\Big). \end{aligned}$$

By definition, we have $\varphi(a^*, b_1, b_2^*) = 0$. Replacing $b_2^* \mathrm{e}^{a^*(b_1 - b_2^*)}$ by $a^* \left(b_2^* - b_1\right) b_1 - b_1$, we get

$$\frac{\mathrm{d}}{\mathrm{d}b_2}\left(a_x(b_2^*) - a(b_1, b_2^*)\right) = \frac{b_1^2 a^* \left(\mathrm{e}^{-a^* b_1} + \mathrm{e}^{-a^* b_2^*} + a^*(b_2^* - b_1)\mathrm{e}^{-a^* b_2^*}\right)}{\left(b_1 \mathrm{e}^{-a^* b_1} + b_2^* \mathrm{e}^{-a^* b_2^*}\right)\left(b_1 + b_2^* \mathrm{e}^{a^*(b_1 - b_2^*)}\right)\left(b_2^* - b_1\right)} > 0.$$

We have then that $\frac{\mathrm{d}}{\mathrm{d}b_2} a_x(b_2) > \frac{\mathrm{d}}{\mathrm{d}b_2} a(b_1, b_2)$ each time these two functions cross. This means that they can cross only once.

For $x \geq 1$, we have already shown that $\frac{\mathrm{d}}{\mathrm{d}b_2} t_x(b_2)$ is positive for all b_2 and $t_x(b_2)$ is then increasing.

For $1 \geq x \geq \frac{1}{2}$, let us show that there is no crossing between $a_x(b_2)$ and $a(b_1, b_2)$. For that purpose let us consider $a = a_x(b_2) = a(b_1, b_2)$. We have then

$$\frac{1}{2}\exp(-b_1 a) + \frac{1}{2}\exp(-b_2 a) = x$$
$$\text{and } b_1 + b_2 \exp((b_1 - b_2)\, a) - a\,(b_2 - b_1)\, b_1 = 0$$

which can be rewritten as follows

$$\exp((b_1 - b_2)\, a) = 2x \exp(b_1 a) - 1$$
$$\text{and } \exp((b_1 - b_2)\, a) = \frac{a\,(b_2 - b_1)\, b_1 - b_1}{b_2}$$

which gives

$$2x \exp(b_1 a) - 1 = \frac{a\,(b_2 - b_1)\, b_1 - b_1}{b_2}$$
$$\text{or } 2x b_2 \exp(b_1 a) = (a b_1 + 1)\,(b_2 - b_1)$$

Remark that $\exp(b_1 a) \geq 1 + b_1 a$ and we have then $2x b_2 \leq (b_2 - b_1)$ or $(2x - 1)\, b_2 \leq -b_1$ which is impossible since we assumed $2x - 1 \geq 0$.

For $x \geq \frac{1}{2}$, it is easy to check that $\lim_{b_2 \to b_1} a_x(b_2) = -\frac{\ln x}{b_1}$ and $a_x(b_2) \sim_{b_2 \to \infty} -\frac{\ln(2x-1)}{b_2}$. The limits of t_x derive from there. □

Proof of Proposition 7 It suffices to prove directly the second point. We have

$$t_x(h) = \frac{1}{2}(b_1 + kh)\exp(-(b_1 + kh)a_x(h)) + \frac{1}{2}(b_2 + h)\exp(-(b_2 + h)a_x(h))$$

where $a_x(h)$ is the solution of $\psi(a_x(h), b_1, b_1 + kh) = x$.

We get after computations

$$\frac{\mathrm{d}}{\mathrm{d}h} t_x(0) \equiv \phi(a_x(0), b_1, b_2)$$

where

$$\phi(a, b_1, b_2) = kb_1 \exp((b_2 - b_1)\, a) + b_2 \exp((b_1 - b_2)\, a) + b_1 + kb_2 + a\,(b_2 - b_1)\,(kb_2 - b_1).$$

For $x \leq 1$, we have $a_x(0) \geq 0$ and it suffices to impose $k \geq \frac{b_1}{b_2}$ to have $\frac{\mathrm{d}}{\mathrm{d}h} t_x(0) \geq 0$ for all x.

For $x \geq 1$, we have $a_x(0) \leq 0$ and it suffices to remark that $\frac{\partial^2 \phi}{\partial a^2}$ is positive, that $\frac{\partial \phi}{\partial a}(0, b_1, b_2) = (b_2 - b_1)(b_2 + b_1)(k - 1)$ and $\phi\,(0, b_1, b_2) = (k + 1)\,(b_1 + b_2) > 0$. It suffices to impose $k \leq 1$ to have $\frac{\partial \phi}{\partial a}(0, b_1, b_2) \leq 0$ and then $\phi(a, b_1, b_2) \geq 0$ for $a \leq 0$ and hence $\frac{\mathrm{d}}{\mathrm{d}h} t_x(0) \geq 0$. □

Proof of Proposition 8 We have

$$t_x(h) = \frac{1}{2}\,(b_1 - h) \exp\,(-(b_1 - h)a_x(h)) + \frac{1}{2}\,(b_2 + h) \exp\,(-(b_2 + h)a_x(h))$$

where $a_x(h)$ is the solution of $\psi(a_x(h), b_1 - h, b_2 + h) = x$. We want to show that for $x < 1$, we have $\frac{\mathrm{d}t_x}{\mathrm{d}h}(0) < 0$ and for $x > 1$ we have $\frac{\mathrm{d}t_x}{\mathrm{d}h}(0) < 0$.

We have $\frac{\mathrm{d}a_x}{\mathrm{d}h}(0) = \frac{a(\exp(-ab_1) - \exp(-ab_2))}{b_1 \exp(-ab_1) + b_2 \exp(-ab_2)}$ where $a = a_x(0)$, and

$$\frac{\mathrm{d}t_x}{\mathrm{d}h}(0) = -\frac{1}{2}\exp\,(-ab_1) + \frac{1}{2}\exp\,(-ab_2) + \frac{1}{2}ab_1 \exp\,(-ab_1) - \frac{1}{2}ab_2 \exp\,(-ab_2) - \left(\frac{1}{2}b_1^2 \exp\,(-ab_1) + \frac{1}{2}b_2^2 \exp\,(-ab_2)\right)\left(\frac{a\,(\exp(-ab_1) - \exp(-ab_2))}{b_1 \exp(-ab_1) + b_2 \exp(-ab_2)}\right)$$

which is of the same sign as $g(a)$ with

$$\begin{aligned} g(a) &= (b_1 \exp\,(-ab_1) + b_2 \exp\,(-ab_2))\,(\exp\,(-ab_2) - \exp\,(-ab_1) \\ &\quad + ab_1 \exp\,(-ab_1) - ab_2 \exp\,(-ab_2)) \\ &\quad - (a\,(\exp(-ab_1) - \exp(-ab_2))\,(b_1^2 \exp\,(-ab_1) + b_2^2 \exp\,(-ab_2)) \\ &= -\,b_1 \exp\,(-2ab_1) + b_2 \exp\,(-2ab_2) - (b_2 - b_1 - ab_1^2 + ab_2^2) \\ &\quad \times\ \exp\,(-a\,(b_1 + b_2)) \end{aligned}$$

which has the same sign as $-\ell(a)$ with

$$\ell(a) = b_1 \exp\,(-a(b_1 - b_2)) - b_2 \exp\,(-a(b_2 - b_1)) + (b_2 - b_1 - ab_1^2 + ab_2^2).$$

We have $\lim_{a \to \infty} \ell(a) = \infty$, $\ell(0) = 0$ and $\lim_{a \to -\infty} \ell(a) = -\infty$. We also have

$$\ell'(a) = (b_1 - b_2)(-b_2 - b_1 - b_1 \exp\,(a(b_2 - b_1)) - b_2 \exp\,(a(b_1 - b_2)) > 0.$$

The function ℓ is then negative on $\mathbb{R}_-$ and positive on $\mathbb{R}_+$. Since $a > 0$ for $x < 1$ and $a < 0$ for $x > 1$ this gives the result. □

Proof of Proposition 9 We denote by t_x^P (resp. t_x^Q) the aggregate level of risk tolerance in the first (resp. second) population when the aggregate wealth is x. We have

$$t_x^P = p_1 b_1 \exp(-b_1 a_x^P) + p_2 b_2 \exp(-b_2 a_x^P)$$

where a_x^P is the solution of

$$p_1 \exp(-b_1 a_x^P) + p_2 \exp(-b_2 a_x^P) = x.$$

We have similar formulas for t_x^Q and we have that $t_x^P \geq t_x^Q$ if and only

$$\frac{q_1 \exp(-b_1 a_x^Q)}{p_1 \exp(-b_1 a_x^P)} \geq \frac{q_2 \exp(-b_2 a_x^Q)}{p_2 \exp(-b_2 a_x^P)} \text{or equivalently} p_1 \exp(-b_1 a_x^P) \leq q_1 \exp(-b_1 a_x^Q) \tag{7}$$

If $x \leq 1$, we have $a_x^P \leq a_x^Q$ and the result is immediate. Let us now consider the case $x \geq 1$. Note that $\exp(-b_1 a_x^Q)$ and $\exp(-b_2 a_x^Q)$ correspond to the Pareto optimal allocations x_1^Q and x_2^Q in population Q while $\exp(-b_1 a_x^P)$ and $\exp(-b_2 a_x^P)$ correspond to the Pareto optimal allocation x_1^P and x_2^P in population P. We want to prove that $\frac{p_1}{q_1} x_1^P \leq x_1^Q$. Remark that the allocation $(\frac{p_1}{q_1} x_1^P, \frac{p_2}{q_2} x_2^P)$ is feasible in population Q. Let us compare $u_1'\left(\frac{p_1}{q_1} x_1^P\right)$ and $u_2'\left(\frac{p_2}{q_2} x_2^P\right)$ or $\left(\frac{p_1}{q_1} x_1^P\right)^{-\frac{1}{b_1}}$ and $\left(\frac{p_2}{q_2} x_2^P\right)^{-\frac{1}{b_2}}$. By construction we have $\left(x_1^P\right)^{-\frac{1}{b_1}} = \left(x_2^P\right)^{-\frac{1}{b_2}}$ and since $\frac{p_1}{q_1} \leq 1$ and $\frac{p_2}{q_2} \geq 1$ we have $u_1'\left(\frac{p_1}{q_1} x_1^P\right) \geq u_2'\left(\frac{p_2}{q_2} x_2^P\right)$. Since Pareto optimal allocations are characterized by the condition $u_1'\left(x_1^Q\right) = u_2'\left(x_2^Q\right)$ and since the allocation $(\frac{p_1}{q_1} x_1^P, \frac{p_2}{q_2} x_2^P)$ is feasible, we necessarily have, by concavity of u_1 and u_2, $\frac{p_1}{q_1} x_1^P \leq x_1^Q$. □

Proof of Proposition 10 We first prove that for $x \leq 1$ an MLR shift is suficient to increase the aggregate level of risk tolerance. Suppose that $P \succcurlyeq_{\text{MLR}} Q$. Since the MLR order is stronger than the FSD order, we have for all nondecreasing function h, $E^P\left[h\left(\tilde{b}\right)\right] \geq E^Q\left[h\left(\tilde{b}\right)\right]$, hence $\Phi_Q(t) \geq \Phi_P(t)$ for $t \geq 0$. Since Φ_Q and Φ_P are decreasing, then for all $x \leq 1, \Phi_P^{-1}(x) \leq \Phi_Q^{-1}(x)$. Since $\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}\widetilde{P}} = \frac{\mathrm{d}Q}{\mathrm{d}P} \mathrm{e}^{-b(\Phi_Q^{-1}(x) - \Phi_P^{-1}(x))} \frac{E^P\left[\mathrm{e}^{-b\Phi_P^{-1}(x)}\right]}{E^Q\left[\mathrm{e}^{-b\Phi_Q^{-1}(x)}\right]}$, we obtain that $\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}\widetilde{P}}$ is the product (modulo a constant) of the decreasing function $\mathrm{e}^{-b(\Phi_Q^{-1}(x) - \Phi_P^{-1}(x))}$ with the decreasing function $\frac{\mathrm{d}Q}{\mathrm{d}P}$ (both of them being positive) and is then decreasing and $\widetilde{P} \succeq_{\text{MLR}} \widetilde{Q}$ which gives $E^{\tilde{Q}}\left[\tilde{b}\right] \leq E^{\tilde{P}}\left[\tilde{b}\right]$ and $t_x^P \geq t_x^Q$ or $R_x^P \leq R_x^Q$.

However, MLR does not guarantee an increase of the aggregate level of risk tolerance when $x > 1$ as shown in the next counter-example.

Let us consider a model in which the distribution of the risk tolerances is given by $\exp(-b) 1_{b \geq 0}$ and let us consider a shift obtained through multiplication of the exponential density by $1 + \varepsilon 1_{b > b^*}$ (for some $b^* > 0$ and some $\varepsilon > 0$) and by

renormalization. The function $b \to 1 + \varepsilon 1_{b>b^*}$ is nondecreasing and the shift is MLR. The aggregate level of risk tolerance in the initial population is given by $t(x) = \int_0^\infty b \exp(-b) \exp(-ab) \mathrm{d}b$ where a solves $\int_0^\infty \exp(-b) \exp(-ab) \mathrm{d}b = x$. We obtain $a = \frac{1}{x} - 1$ and $t(x) = \frac{1}{(a+1)^2} = x^2$.

After the shift, the aggregate level of risk tolerance is given by

$$t_\varepsilon(x) = \frac{\int_0^\infty b \exp(-b) \exp(-a_\varepsilon b) \mathrm{d}b + \varepsilon \int_{b^*}^\infty b \exp(-b) \exp(-a_\varepsilon b) \mathrm{d}b}{\int_0^\infty \exp(-b) \mathrm{d}b + \varepsilon \int_{b^*}^\infty \exp(-b) \mathrm{d}b} \tag{8}$$
$$= \frac{\frac{1}{(a_\varepsilon+1)^2} + \varepsilon \left(\frac{1}{a_\varepsilon+1} b^* \exp\left(-(a_\varepsilon+1)b^*\right) + \frac{1}{(a_\varepsilon+1)^2} \exp\left(-(a_\varepsilon+1)b^*\right) \right)}{1 + \varepsilon \exp\left(-b^*\right)}$$

where a_ε solves $\frac{\int_0^\infty \exp(-b) \exp(-a_\varepsilon b) \mathrm{d}b + \varepsilon \int_{b^*}^\infty \exp(-b) \exp(-a_\varepsilon b) \mathrm{d}b}{\int_0^\infty \exp(-b) \mathrm{d}b + \varepsilon \int_{b^*}^\infty \exp(-b) \mathrm{d}b} = x$ or $\frac{\frac{1}{a_\varepsilon+1} + \varepsilon \frac{1}{a_\varepsilon+1} \exp\left(-(a_\varepsilon+1)b^*\right)}{1+\varepsilon \exp(-b^*)} = x$. We have then $\frac{1}{a+1} = \frac{\frac{1}{a_\varepsilon+1} + \varepsilon \frac{1}{a_\varepsilon+1} \exp\left(-(a_\varepsilon+1)b^*\right)}{1+\varepsilon \exp(-b^*)}$. Let us consider the difference $t_\varepsilon(x) - t(x)$. It is positively proportional to

$$\Delta = \left(\frac{1}{\beta^2} + \varepsilon \left(\frac{1}{\beta} y^* \exp\left(-\beta y^*\right) + \frac{1}{\beta^2} \exp\left(-\beta y^*\right) \right) \right) \left(1 + \varepsilon \exp\left(-y^*\right)\right) \tag{9}$$
$$- \left(\frac{1}{\beta} + \varepsilon \frac{1}{\beta} \exp\left(-\beta y^*\right) \right)^2$$

where $\beta = a_\varepsilon + 1$ and where $\frac{1}{a+1}$ has been replaced by its value as a function of β. This quantity is of the form $\mu\varepsilon + \nu\varepsilon^2$ and since we want it to be positive for all $\varepsilon > 0$ we have to check if μ is positive. But $\mu = \frac{(\exp(\beta b^*) - \exp(b^*) + \beta b^* \exp(b^*))}{\exp(\beta b^*) \exp(b^*) \beta^2}$ and is positively proportional to $\exp(\beta b^*) - \exp(b^*) + \beta b^* \exp(b^*)$. It is easy to remark that for $\beta = \frac{1}{b^{*2}}$ and b^* sufficiently large, this quantity is negative. Let us chose then a pair (β, b^*) for which this quantity is negative and let us take ε sufficiently small such that the quantity Δ itself is negative and let us finally take $x = \frac{\frac{1}{(a_\varepsilon+1)^2} + \varepsilon \left(\frac{1}{a_\varepsilon+1} b^* \exp\left(-(a_\varepsilon+1)b^*\right) + \frac{1}{(a_\varepsilon+1)^2} \exp\left(-(a_\varepsilon+1)b^*\right) \right)}{1+\varepsilon \exp(-b^*)}$. The resulting shift leads then to a decrease of the collective level of risk tolerance at x. □

Proof of Proposition 11 We just have to consider the case where $x \geq 1$. We have

$$\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}\widetilde{P}} = \frac{\mathrm{d}Q}{\mathrm{d}P} \mathrm{e}^{-b(\Phi_Q^{-1}(x) - \Phi_P^{-1}(x))} \frac{E^P[\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_Q^{-1}(x)}]} \tag{10}$$
$$= \exp(-kb - b(\Phi_Q^{-1}(x) - \Phi_P^{-1}(x))).$$

To conclude, it is sufficient to show that $k + (\Phi_Q^{-1}(x) - \Phi_P^{-1}(x)) > 0$. We have $E^P[\mathrm{e}^{-b(\Phi_Q^{-1}(x)+k)}] = \frac{E^Q[\mathrm{e}^{-b\Phi_Q^{-1}(x)}]}{E^Q[\exp(kb)]} < x$ and since $\Phi_P^{-1}(x)$ is characterized by $E^P[\mathrm{e}^{-b\Phi_P^{-1}(x)}] = x$ we have $k + (\Phi_Q^{-1}(x) - \Phi_P^{-1}(x)) > 0$. □

Proof of Proposition 12 Assume that P and Q are symmetric with respect to some b^* with $\frac{\mathrm{d}Q}{\mathrm{d}P}$ nonincreasing before b^* and nondecreasing after b^* then $P \succcurlyeq_{\mathrm{PD}} Q$ and $P \succcurlyeq_{\mathrm{SSD}} Q$ (Jouini and Napp 2008). Let us denote by Φ_P and Φ_Q the functions, respectively, defined by $\Phi_P(t) = \int \mathrm{e}^{-bt}\mathrm{d}P(b)$ and $\Phi_Q(t) = \int \mathrm{e}^{-bt}\mathrm{d}Q(b)$. Since e^{-bt} is decreasing and convex for $t \geq 0$, we have by SSD, $\Phi_Q(t) \geq \Phi_P(t)$ for all $t \geq 0$. For $x \leq 1$, $\Phi_P^{-1}(x)$ and $\Phi_Q^{-1}(x)$ are positive. Furthermore, both Φ_Q and Φ_P are decreasing and we have then $\Phi_Q^{-1}(x) \geq \Phi_P^{-1}(x)$ for all $x \leq 1$. Since $b \to \mathrm{e}^{-b\Phi_P^{-1}(x)}$ is decreasing and positive, we have by PD, $\frac{E^P[b\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^P[\mathrm{e}^{-b\Phi_P^{-1}(x)}]} \geq \frac{E^Q[b\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_P^{-1}(x)}]} = E^{P_1}[b]$ with $\frac{\mathrm{d}P_1}{\mathrm{d}Q} = \frac{\mathrm{e}^{-b\Phi_P^{-1}(x)}}{E^Q[\mathrm{e}^{-b\Phi_P^{-1}(x)}]}$. Let us consider Q_1 defined by $\frac{\mathrm{d}Q_1}{\mathrm{d}Q} = \frac{\mathrm{e}^{-b\Phi_Q^{-1}(x)}}{E^Q[\mathrm{e}^{-b\Phi_Q^{-1}(x)}]}$. We have $\frac{\mathrm{d}Q_1}{\mathrm{d}P_1} = \frac{E^Q[\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_Q^{-1}(x)}]}\mathrm{e}^{-b\left(\Phi_Q^{-1}(x)-\Phi_P^{-1}(x)\right)}$ and is decreasing. Hence $E^{P_1}[b] \geq E^{Q_1}[b] = \frac{E^Q[b\mathrm{e}^{-b\Phi_Q^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_Q^{-1}(x)}]}$. We have then the result for $x \leq 1$. For $x \geq 1$, since both distributions are symmetric with respect to b^*, we have $\frac{E^P[b\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^P[\mathrm{e}^{-b\Phi_P^{-1}(x)}]} = \frac{E^P[(2b^*-b)\mathrm{e}^{-(2b^*-b)\Phi_P^{-1}(x)}]}{E^P[\mathrm{e}^{-(2b^*-b)\Phi_P^{-1}(x)}]} = 2b^* - \frac{E^P[b\mathrm{e}^{b\Phi_P^{-1}(x)}]}{E^P[\mathrm{e}^{b\Phi_P^{-1}(x)}]}$ and $\frac{E^Q[b\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_P^{-1}(x)}]} = 2b^* - \frac{E^Q[b\mathrm{e}^{b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{b\Phi_P^{-1}(x)}]}$. Since $\Phi_P^{-1}(x)$ is negative for $x \geq 1$, the function $b \to \mathrm{e}^{b\Phi_P^{-1}(x)}$ is decreasing and positive, and we have by PD that $\frac{E^P[b\mathrm{e}^{b\Phi_P^{-1}(x)}]}{E^P[\mathrm{e}^{b\Phi_P^{-1}(x)}]} \geq \frac{E^Q[b\mathrm{e}^{b\Phi_Q^{-1}(x)}]}{E^Q[\mathrm{e}^{b\Phi_Q^{-1}(x)}]}$ or $E^{P_1}[b] = \frac{E^Q[b\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_P^{-1}(x)}]} \geq \frac{E^P[b\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^P[\mathrm{e}^{-b\Phi_P^{-1}(x)}]}$. Since $\frac{\mathrm{d}Q_1}{\mathrm{d}P_1} = \frac{E^Q[\mathrm{e}^{-b\Phi_P^{-1}(x)}]}{E^Q[\mathrm{e}^{-b\Phi_Q^{-1}(x)}]}\mathrm{e}^{-b\left(\Phi_Q^{-1}(x)-\Phi_P^{-1}(x)\right)}$, it is now increasing, we have then $E^{P_1}[b] \leq E^{Q_1}[b]$ which gives the result. □

Proof of Proposition 13 Let us denote by $t(x,h)$ the aggregate level of risk tolerance at x when the individual levels of risk tolerance are given by b_1 and $b_2 + h$, we have

$$t(x,h) = \frac{\exp(-\frac{\ln x}{b_1}) + \exp(-\frac{\ln x}{b_2+h})}{\frac{1}{b_1}\exp(-\frac{\ln x}{b_1}) + \frac{1}{b_2+h}\exp(-\frac{\ln x}{b_2+h})}$$

and

$$\frac{\partial t}{\partial h}(x,0) = \frac{b_1\left((b_2-b_1)(\ln x)\,\mathrm{e}^{-\frac{\ln x}{b_1}} + b_1b_2\left(\mathrm{e}^{-\frac{\ln x}{b_1}} + \mathrm{e}^{-\frac{\ln x}{b_2}}\right)\right)\mathrm{e}^{-\frac{\ln x}{b_2}}}{b_2\left(b_1\mathrm{e}^{-\frac{\ln x}{b_2}} + b_2\mathrm{e}^{-\frac{\ln x}{b_1}}\right)^2}$$

and we clearly have $\frac{\partial t}{\partial h}(x,0) > 0$ for $x > 1$. Furthermore, if we denote by $L(x)$ the quantity $L(x) = (b_2-b_1)(\ln x) + b_1b_2\left(1 + \mathrm{e}^{\ln x\left(\frac{1}{b_1}-\frac{1}{b_2}\right)}\right)$, we have $\frac{\mathrm{d}L}{\mathrm{d}\ln x} =$

$(b_2 - b_1)\left(1 + \mathrm{e}^{\ln x\left(\frac{1}{b_1} - \frac{1}{b_2}\right)}\right) > 0$ and $\lim_{x\to 0} L(x) = -\infty$ and $L(1) = 2b_1 b_2 > 0$. There exists then $x^* < 1$ such that $\frac{\partial t}{\partial h}(x, 0) < 0$ for $x < x^*$ and $\frac{\partial t}{\partial h}(x, 0) > 0$ for $x > x^*$. □

Proof of Proposition 14 Let us denote by $R(x, h)$ the aggregate level of risk aversion at x when the individual levels of risk aversion are given by $R_1 - h$ and $R_2 + h$, we have $R(x, h) = \frac{(R_1-h)\exp(-\ln x(R_1-h))+(R_2+h)\exp(-\ln x(R_2+h))}{\exp(-\ln x(R_1-h))+\exp(-\ln x(R_2+h))}$ and $\frac{\partial R}{\partial h}(x, 0) = -\frac{\mathrm{e}^{-2\ln x R_1} - \mathrm{e}^{-2\ln x R_2} + 2\ln x(R_2-R_1)\mathrm{e}^{-\ln x R_1}\mathrm{e}^{-\ln x R_2}}{\left(\mathrm{e}^{-\ln x R_1} + \mathrm{e}^{-\ln x R_2}\right)^2}$ which is clearly negative for $x > 1$ and positive for $x < 1$. □

Proof of Proposition 15 Let us denote by $U(x, 1)$ and $U(x, 2)$ the social utility functions, respectively, associated to P_1 and P_2. We denote, respectively, by $F(i, 1)$ and $F(i, 2)$ the cumulative distributions of P_1 and P_2. We have

$$U'(x, j) = \int \frac{\partial u}{\partial x}(x, i) F'(i, j)\mathrm{d}i, \quad j = 1, 2.$$

By assumption, $\frac{\partial u}{\partial x}(x, i)$ is log-supermodular. Furthermore, since $P_2 \succcurlyeq_{\mathrm{MLR}} P_1$, $F'(i, j)$ is also LSPM. By Karlin's Theorem $U'(x, j)$ is log-supermodular. Therefore, $\frac{\partial \ln U'(x,j)}{\partial x}$ increases with j or in other words

$$-\frac{U''(x, 1)}{U'(x, 1)} \geq -\frac{U''(x, 2)}{U'(x, 2)}$$

which gives $R_1(x) \geq R_2(x)$. □

Proof of Proposition 16 Let us denote by $U(x, 1)$ and $U(x, 2)$ the social utility functions, respectively, associated to P_1 and P_2. Since $\frac{\partial u}{\partial x}(x, i)$ is increasing in i and $-\frac{u_i''(x)}{u_i'(x)}$ is decreasing in i, there exists a nonnegative and decreasing function Ψ^x such that $E\left[u_i'(x)\middle| -\frac{u_i''(x)}{u_i'(x)}\right] = \Psi^x\left(-\frac{u_i''(x)}{u_i'(x)}\right)$ for $i \in I$. We have then

$$-\frac{U''(x, j)}{U'(x, j)} = \frac{\int -\frac{u_i''(x)}{u_i'(x)} u_i'(x)\mathrm{d}P_j(i)}{\int u_i'(x, j)\mathrm{d}P_j(i)}, \; j = 1, 2$$
$$= \frac{E^{P_j^x}\left[X\Psi^x(X)\right]}{E^{P_j^x}\left[\Psi^x(X)\right]}.$$

By Portfolio Dominance we immediately have $\frac{E^{P_2^x}\left[X\Psi^x(X)\right]}{E^{P_2^x}\left[\Psi^x(X)\right]} \geq \frac{E^{P_1^x}\left[X\Psi^x(X)\right]}{E^{P_1^x}\left[\Psi^x(X)\right]}$. □

References

Arrow KJ (1971) Essays in the theory of risk-bearing. North-Holland, Amsterdam
Baker R, Laury S, Williams A (2008) Comparing small-group and individual behavior in lottery-choice experiments. South Econ J 75:367–382
Barsky RB, Juster FT, Kimball MS, Shapiro MD (1997) Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study. Quart J Econ 112:537–579
Bone J, Hey J, Suckling J (1999) Are groups more (or less) consistent than individuals? J Risk Uncertain 18:63–81
Charness G, Karni E, Levin D (2007) Individual and group decision making under risk: an experimental study of Bayesian updating and violations of first-order stochastic dominance. J Risk Uncertain 35:129–148
Clark R (1971) Group induced shift toward risk: a critical appraisal. Psychol Bull 76:251–270
Eliaz K, Ray D, Razin R (2006) Choice shifts in groups: a decision-theoretic basis. Am Econ Rev 96:1321–1332
Gollier C (1997) A note on portfolio dominance. Rev Econ Stud 64:147–150
Gollier C (2001) Wealth inequality and asset pricing. Rev Econ Stud 68:181–203
Gollier C (2007) Whom should we believe? Aggregation and heterogeneous beliefs. J Risk Uncertain 35:107–127
Hara C, Huang J, Kuzmics C (2007) Representative consumer's risk aversion and efficient risk-sharing rules. J Econ Theory 137:652–672
Jouini E, Napp C (2008) On Abel's concept of doubt and pessimism. J Econ Dyn Control 32:3682–3694
Landsberger M, Meilijson I (1990) Demand for risky financial assets: a portfolio analysis. J Econ Theory 50:204–213
Masclet D, Colombier N, Denant-Boemont L, Loheac Y (2009) Group and individual risk preferences: a lottery-choice experiment with self-employed and salaried workers. J Econ Behav Organ 70:470–484
Mazzocco M (2004) Savings, risk sharing and preferences for risk. Am Econ Rev 94:1169–1182
Pratt JW (1964) Risk aversion in the small and in the large. Econometrica 32:122–136
Pratt JW, Zeckhauser RJ (1989) The impact of risk sharing on efficient decision. J Risk Uncertain 2:219–234
Samuelson PA (1956) Social indifference curves. Quart J Econ 70:1–22
Shupp R, Williams A (2008) Risk preference differentials of small groups and individuals. Econ J 118:258–283
Wilson R (1968) The theory of syndicates. Econometrica 36:119–132


THE UNIVERSITY OF CHICAGO PRESS JOURNALS

Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility
Author(s): John C. Harsanyi
Source: *Journal of Political Economy*, Vol. 63, No. 4 (Aug., 1955), pp. 309-321
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/1827128
Accessed: 04-09-2016 10:35 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Political Economy*

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

CARDINAL WELFARE, INDIVIDUALISTIC ETHICS, AND INTERPERSONAL COMPARISONS OF UTILITY[1]

JOHN C. HARSANYI
University of Queensland

I

THE naïve concept of social welfare as a sum of intuitively measurable and comparable individual cardinal utilities has been found unable to withstand the methodological criticism of the Pareto school. Professor Bergson[2] has therefore recommended its replacement by the more general concept of a social welfare function, defined as an arbitrary mathematical function of economic (and other social) variables, of a form freely chosen according to one's personal ethical (or political) value judgments. Of course, in this terminology everybody will have a social welfare function of his own, different from that of everybody else, except to the extent to which different individuals' value judgments happen to coincide with one another. Actually, owing to the prevalence of individualistic value judgments in our society, it has been generally agreed that a social welfare function should be an increasing function of the utilities of individuals: if a certain situation, X, is preferred by an individual to another situation, Y, and if none of the other individuals prefers Y to X, then X should be regarded as socially preferable to Y. But no other restriction is to be imposed on the mathematical form of a social welfare function.

Recently, however, Professor Fleming[3] has shown that if one accepts one further fairly weak and plausible ethical postulate, one finds one's social welfare function to be at once restricted to a rather narrow class of mathematical functions so as to be expressible (after appropriate monotone transformation of the social welfare and individual utility indexes if necessary) as the weighted sum of the individuals' utilities. This does not mean, of course, a return to the doctrine that the existence of an additive cardinal utility function is intuitively self-evident. The existence of such a function becomes, rather, the consequence of the ethical postulates adopted and is wholly dependent on these postulates. Still, Fleming's results do in a sense involve an unexpected revival of some views of the pre-Pareto period.

In this paper I propose, first of all, to examine the precise ethical meaning of Fleming's crucial postulate and to show that it expresses an *individualistic* value judgment going definitely beyond the generally adopted individualistic postu-

[1] I am indebted to my colleagues at the University of Queensland, Messrs. R. W. Lane and G. Price, for helpful comments. Of course, the responsibility for shortcomings of this paper and for the opinions expressed in it is entirely mine.

[2] A. Bergson (Burk), "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, LII (February, 1938), 310–34, and "Socialist Economics," in *A Survey of Contemporary Economics*, ed. H. S. Ellis (Philadelphia, 1949), esp. pp. 412–20.

[3] M. Fleming, "A Cardinal Concept of Welfare," *Quarterly Journal of Economics*, LXVI (August, 1952), 366–84. For a different approach to the same problem see L. Goodman and H. Markovitz, "Social Welfare Functions Based on Individual Rankings," *American Journal of Sociology*, Vol. LVIII (November, 1952).

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

late mentioned earlier, though it represents, as I shall argue, a value judgment perfectly acceptable according to common ethical standards (Sec. II). I shall also attempt to show that, if both social and individual preferences are assumed to satisfy the von Neumann–Morgenstern-Marschak axioms about choices between uncertain prospects, even a much weaker ethical postulate than Fleming's suffices to establish an additive cardinal social welfare function (Sec. III). In effect, it will be submitted that a mere logical analysis of what we mean by value judgments concerning social welfare and by social welfare functions leads, without any additional ethical postulates, to a social welfare function of this mathematical form (Sec. IV). Finally, I shall turn to the problem of interpersonal comparisons of utility, which gains new interest by the revival of an additive cardinal welfare concept, and shall examine what logical basis, if any, there is for such comparisons (Sec. V).

II

Fleming expresses his ethical postulates in terms of two alternative conceptual frameworks: one in terms of an "*ideal utilitarianism*" of G. E. Moore's type, the other in terms of a *preference* terminology more familiar to economists. Though he evidently sets greater store by the first approach, I shall adopt the second, which seems to be freer of unnecessary metaphysical commitments. I have also taken the liberty of rephrasing his postulates to some extent.

Postulate A (*asymmetry of social preference*).—If "from a social standpoint"[4] situation X is preferred to situation Y, then Y is not preferred to X.

Postulate B (*transitivity of social preference*).—If from a social standpoint X is preferred to Y, and Y to Z, then X is preferred to Z.

Postulate C (*transitivity of social indifference*).—If from a social standpoint neither of X and Y is preferred to the other, and again neither of Y and Z is preferred to the other, then likewise neither of X and Z is preferred to the other.

These three postulates are meant to insure that "social preference" establishes a *complete ordering* among the possible social situations, from which the existence of a social welfare function (at least of an ordinal type) at once follows. (Actually, two postulates would have sufficed if, in the postulates, "weak" preference, which does not exclude the possibility of indifference, had been used instead of "strong" preference.)

Postulate D (*positive relation of social preferences to individual preferences*).—If a given individual i prefers situation X to situation Y, and none of the other individuals prefers Y to X, then X is preferred to Y from a social standpoint.

As already mentioned Postulate D expresses a generally accepted individualistic value judgment.

Finally, Fleming's Postulate E states essentially that on issues on which two individuals' interests (preferences) conflict, all other individuals' interests being unaffected, social preferences should depend exclusively on comparing the relative social importance of the interests at stake of each of the two individuals concerned. In other words, it requires that

[4] Of course, when I speak of preferences "from a social standpoint," often abbreviated to "social" preferences and the like, I always mean preferences based on a given individual's value judgments concerning "social welfare." The foregoing postulates are meant to impose restrictions on *any* individual's value judgements of this kind, and thus represent, as it were, value judgments of the second order, that is, value judgments concerning value judgments. Later I shall discuss the concept of "preferences from a social standpoint" at some length and introduce the distinctive term "ethical preferences" to describe them (in Sec. IV). But at this stage I do not want to prejudge the issue by using this terminology.

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

the distribution of utilities between each pair of individuals should be judged separately on its own merits, independently of how utilities (or income) are distributed among the other members of the community.

Postulate E (independent evaluation of the utility distribution[5] between each pair of individuals).—(1) There are at least three individuals. (2) Suppose that individual i is indifferent between situations X and X' and also between situations Y and Y', but prefers situations X and X' to situations Y and Y'. Suppose, further, that individual j is also indifferent between X and X' and between Y and Y', but (unlike individual i) prefers Y and Y' to X and X'. Suppose also that all other individuals are indifferent between X and Y, and likewise between X' and Y'.[6] Then social preferences should always go in the same way between X and Y as they do between X' and Y' (that is, if from a social standpoint X is preferred to Y, then X' should also be preferred to Y'; if from a social standpoint X and Y are regarded as indifferent, the same should be true of X' and Y'; and if from a social standpoint Y is preferred to X, then Y' should also be preferred to X').

Postulate E is a natural extension of the individualistic value judgment expressed by Postulate D. Postulate D already implies that if the choice between two situations X and Y happens to affect the interests of the individuals i and j only, without affecting the interests of anybody else, social choice must depend exclusively on i's and j's interests—provided that i's and j's interests *agree* in this matter. Postulate E now adds that in the assumed case social choice must depend exclusively on i's and j's interests (and on weighing these two interests one against the other in terms of a consistent ethical standard), even if i's and j's interests are in *conflict*. Thus both postulates make social choice dependent solely on the *individual* interests directly affected.[7] They leave no room for the separate interests of a superindividual state or of impersonal cultural values[8] (except for the ideals of equity incorporated in the ethical postulates themselves).

At first sight, Postulate E may look inconsistent with the widespread habit of judging the "fairness" or "unfairness" of the distribution of income between two individuals, not only on the basis of these two people's personal conditions and needs, but also on the basis of comparing

[5] The more general term "utility distribution" is used instead of the term "income distribution," since the utility enjoyed by each individual will, in general, depend not only on his own income but also, owing to external economies and diseconomies of consumption, on other people's incomes.

[6] It is not assumed, however, that the other individuals are (like i and j) indifferent between X and X' and between Y and Y'. In effect, were this restrictive assumption inserted into Postulate E, this latter would completely lose the status of an independent postulate and would become a mere corollary of Postulate D.

[7] In view of consumers' notorious "irrationality," some people may feel that these postulates go too far in accepting the consumers' sovereignty doctrine. These people may reinterpret the terms in the postulates referring to individual preferences as denoting, not certain individuals' actual preferences, but rather their "true" preferences, that is, the preferences they *would* manifest under "ideal conditions," in possession of perfect information, and acting with perfect logic and care. With some ingenuity it should not be too difficult to give even some sort of "operational" meaning to these ideal conditions, or to some approximation of them, acceptable for practical purposes. (Or, alternatively, these terms may be reinterpreted as referring even to the preferences that these individuals *ought* to exhibit in terms of a given ethical standard. The latter interpretation would, of course, deprive the postulates of most of their individualistic meaning.)

[8] These postulates do not exclude, however, the possibility that such consideration may influence the relative weights given to different individuals' utilities within the additive social welfare function. Even by means of additional postulates, this could be excluded only to the extent to which the comparison of individual utilities can be put on an objective basis independent of individual value judgments (see Sec. V).

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

their incomes with the incomes of the other members of their respective social groups. Thus people's judgments on the income distribution between a given worker and his employer will also depend on the current earnings of other similar workers and employers. But the conflict with Postulate E is more apparent than real. In a society with important external economies and diseconomies of consumption, where the utility of a given income depends not only on its absolute size but also on its relation to other people's incomes, it is not inconsistent with Postulate E that, in judging the income distribution between two individuals, other people's incomes should also be taken into account. An income distribution between a given worker and a given employer, which in the original situation seemed perfectly "fair" in terms of a given ethical standard, may require adjustment in the worker's favor, once wages have generally gone up, since the worsening of this worker's position relative to that of his fellows must have reduced him to a lower level of utility.

Postulate E requires that the distribution of *utility* between two individuals (once the utility levels of the two individuals are given) should always be judged independently of how utility and income are distributed among other members of the society. In the absence of external economies and diseconomies of consumption, this would necessarily also mean judging the distribution of *income* between two individuals independently of the incomes of others. In the presence of such economies and diseconomies, however, when the utility level of any person depends not only on his own income but also on other persons' incomes, it is not inconsistent with Postulate E that our value judgment on the distribution of income between two individuals should be influenced by the income distribution in the rest of the society—in so far as the income distribution in the rest of the society affects the utility levels of these two individuals themselves and consequently the distribution of utility between them. Postulate E demands only that, once these effects have been allowed for, the distribution of income in the rest of the society must not have any further influence on our value judgment.

III

In accordance with prevalent usage in welfare economics, Fleming's postulates refer to social or individual preferences between *sure prospects* only. However, it seems desirable to have both sorts of preferences defined for choices between *uncertain prospects* as well. More often than not, we have to choose in practice between social policies that promise given definite results only with larger or smaller probabilities. On the other hand, if we subscribe to some sort of individualistic ethics, we should like to make social attitude toward uncertainty somehow dependent on individual attitudes toward it (at least if the latter do not manifest too patent and too great an inconsistency and irrationality).

Since we admit the possibility of external economies and diseconomies of consumption, both social and individual prospects will, in general, specify the amounts of different commodities consumed and the stocks of different goods held by all individuals at different future dates (up to the time horizon adopted), together with their respective probabilities.

As the von Neumann–Morgenstern axioms[9] or the Marschak postulates[10]

[9] See J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (2d ed.; Princeton, 1947), pp. 641 ff.

[10] J. Marschak, "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*,

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

equivalent to them (which latter I shall adopt) are essential requirements for rational behavior, it is natural enough to demand that both social and individual preferences[11] should satisfy them. This gives us:

Postulate a.—Social preferences satisfy Marschak's Postulates I, II, III′, and IV.

Postulate b.—Individual preferences satisfy the same four postulates.

In addition, we need a postulate to secure the dependence of social preferences on individual preferences:

Postulate c.—If two prospects P and Q are indifferent from the standpoint of every individual, they are also indifferent from a social standpoint.

Postulate c once more represents, of course, an individualistic value judgment—though a very weak one, comparable to Fleming's Postulate D rather than to his Postulate E.

I propose to show that Postulate c suffices to establish that the cardinal social welfare function defined by Postulate a can be obtained as a weighted sum of the cardinal individual utility functions defined by Postulate b (on the understanding that the zero point of the social welfare function is appropriately chosen).

Theorem I.—There exists a social welfare function such that its actuarial value is maximized by choices conformable to the social preferences given. This social welfare function is unique up to linear transformation.

Theorem II.—For each individual there exists a utility function such that its actuarial value is maximized by choices conformable to the individual's preferences. This utility function is unique up to linear transformation.

Both theorems follow from Marschak's argument.

Let W denote a social welfare function satisfying Theorem I and U_i denote a utility function of the i'th individual, satisfying Theorem II. Moreover, let W be chosen so that $W = 0$ if for all the n individuals $U_1 = U_2 = \ldots = U_n = 0$.

Theorem III.—W is a single-valued function of $U_1, U_2, \ldots, U_n$. This follows, in view of Theorems I and II, from Postulate c.

Theorem IV.—W is a homogeneous function of the first order of $U_1, U_2, \ldots, U_n$.

Proof.—We want to show that, if the individual utilities $U_1 = u_1$; $U_2 = u_2$; $\ldots$; $U_n = u_n$ correspond to the social welfare $W = w$, then the individual utilities $U_1 = k \cdot u_1$; $U_2 = k \cdot u_2$; $\ldots$; $U_n = k \cdot u_n$ correspond to the social welfare $W = k \cdot w$.

This will be shown first for the case where $0 \leqq k \leqq 1$. Suppose that prospect O represents $U_1 = U_2 = \ldots = U_n = 0$

XVIII (1950), 111–41, esp. 116–21. Marschak's postulates can be summarized as follows. *Postulate I (complete ordering)*: The relation of preference establishes a complete ordering among all prospects. *Postulate II* (continuity): If prospect P is preferred to prospect R, while prospect Q has an intermediate position between them (being preferred to R but less preferred than P), then there exists a mixture of P and R, with appropriate probabilities, such as to be exactly indifferent to Q. *Postulate III′ (sufficient number of nonindifferent prospects)*: There are at least four mutually nonindifferent prospects. *Postulate IV (equivalence of mixture of equivalent prospects)*: If prospects Q and Q' are indifferent, then, for any prospect P, a given mixture of P and Q is indifferent to a similar mixture of P and Q', (that is, to a mixture of P and Q' which has the same probabilities for the corresponding constituent prospects).

Postulate I is needed to establish the existence of even an *ordinal* utility (or welfare) function, while the other three postulates are required to establish the existence of a *cardinal* utility (or welfare) function. But, as Postulates II and III are almost trivial, Postulate IV may be regarded as being decisive for cardinality as against mere ordinality.

[11] There are reasons to believe that, in actuality, individual preferences between uncertain prospects do not always satisfy these postulates of rational behavior (for example, owing to a certain "love of danger"; see Marschak, *op. cit.*, pp. 137–41). In this case we may fall back again upon the preferences each individual *would* manifest under "ideal conditions" (see n. 5).

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

for the different individuals and consequently represents $W = 0$ for society, while prospect P represents $U_1 = u_1$; $U_2 = u_2$; . . . ; $U_n = u_n$ for the former and $W = w$ for the latter. Moreover, let Q be the mixed prospect of obtaining either prospect O (with the probability $1 - p$) or prospect P (with the probability p). Then, obviously, Q will represent $U_1 = p \cdot u_1$; $U_2 = p \cdot u_2$; . . . ; $U_n = p \cdot u_n$ for the individuals and $W = p \cdot w$ for society. Now, if we write $k = p$, a comparison between the values of the variables belonging to prospect P and those belonging to prospect Q will, in view of Theorem III, establish the desired result for the case where $0 \leqq k \leqq 1$ (p, being a probability, cannot be <0 or >1).

Next let us consider the case where $k < 0$. Let us choose prospect R so that prospect O becomes equivalent to the mixed prospect of obtaining either prospect R (with the probability p) or prospect P (with the probability $1 - p$). A little calculation will show that in this case prospect R will represent $U_1 = (1 - 1/p) \cdot u_1$; $U_2 = (1 - 1/p) \cdot u_2$; . . . ; $U_n = (1 - 1/p) \cdot u_n$ for the different individuals and $W = (1 - 1/p) \cdot w$ for society. If we now write $k = 1 - 1/p$, a comparison between the variables belonging to R and those belonging to P will establish the desired result for the case $k < 0$ (by an appropriate choice of the probability p, we can make k equal to any negative number).

Finally, the case where $k > 1$ can be taken care of by finding a prospect S such that prospect P becomes equivalent to the mixed prospect of obtaining either S (with a probability p) or O (with a probability $1 - p$). Then this prospect S will be connected with the values $U_1 = 1/p \cdot u_1$; $U_2 = 1/p \cdot u_2$; . . . ; $U_n = 1/p \cdot u_n$ and $W = 1/p \cdot w$. If we now write $k = 1/p$ we obtain the desired result for the case where $k > 1$ (by an appropriate choice of p we can make k equal to any number >1).

Theorem V.—W is a weighted sum of the individual utilities, of the form

$$W = \Sigma a_i \cdot U_i ,$$

where a_i stands for the value that W takes when $U_i = 1$ and $U_j = 0$ for all $j \neq i$.

Proof.—Let S_i be a prospect representing the utility U_i to the ith individual and the utility zero to all other individuals. Then, according to Theorem IV, for S_i we have $W = a_i \cdot U_i$.

Let T be the mixed prospect of obtaining either S_1 or S_2 or . . . S_n, each with probability $1/n$. Then T will represent the individual utilities U_1/n, U_2/n, . . . , U_n/n and the social welfare

$$W = \frac{1}{n} \cdot \Sigma a_i \cdot U_i .$$

In view of Theorem IV, this directly implies that if the individual utility functions take the values U_1, U_2, . . . , U_n, respectively, the social welfare function has the value

$$W = \Sigma a_i \cdot U_i ,$$

as desired.[12]

IV

In the pre-Pareto conceptual framework, the distinction between social welfare and individual utilities was free of ambiguity. Individual utilities were assumed to be directly given by introspection, and social welfare was simply their sum. In the modern approach, however, the distinction is far less clear. On the one hand, our social welfare concept has

[12] If we want a formal guaranty that no individual's utility can be given a negative weight in the social welfare function, we must add one more postulate (for instance, Postulate D of Sec. II).

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

come logically nearer to an individual utility concept. Social welfare is no longer regarded as an objective quantity, the same for all, by necessity. Rather, each individual is supposed to have a social welfare function of his own, expressing his own individual values—in the same way as each individual has a utility function of his own, expressing his own individual taste. On the other hand, our individual utility concept has come logically nearer to a social welfare concept. Owing to a greater awareness of the importance of external economies and diseconomies of consumption in our society, each individual's utility function is now regarded as dependent not only on this particular individual's economic (and noneconomic) conditions but also on the economic (and other) conditions of all other individuals in the community—in the same way as a social welfare function is dependent on the personal conditions of all individuals.

At the same time, we cannot allow the distinction between an individual's social welfare function and his utility function to be blurred if we want (as most of us do, I think) to uphold the principle that a social welfare function ought to be based not on the utility function (subjective preferences) of *one* particular individual only (namely, the individual whose value judgments are expressed in this welfare function), but rather on the utility functions (subjective preferences) of *all* individuals, representing a kind of "fair compromise" among them.[13] Even if both an individual's social welfare function and his utility function in a sense express his own individual preferences, they must express preferences of different sorts: the former must express what this individual prefers (or, rather, would prefer) on the basis of impersonal social considerations alone, and the latter must express what he actually prefers, whether on the basis of his personal interests or on any other basis. The former may be called his "ethical" preferences, the latter his "subjective" preferences. Only his "subjective" preferences (which define his utility function) will express his preferences in the full sense of the word as they actually are, showing an egoistic attitude in the case of an egoist and an altruistic attitude in the case of an altruist. His "ethical" preferences (which define his social welfare function) will, on the other hand, express what can in only a qualified sense be called his "preferences": they will, by definition, express what he prefers only in those possibly rare moments when he forces a special impartial and impersonal attitude upon himself.[14]

In effect, the ethical postulates pro-

[13] This principle is essentially identical with Professor Arrow's "nondictatorship" postulate in his *Social Choice and Individual Values* (New York, 1951), p. 30 (see also n. 12).

[14] Mr. Little's objection to Arrow's nondictatorship postulate (see Little's review article in the *Journal of Political Economy*, LX [October, 1952], esp. 426–31) loses its force, once the distinction between "ethical" and "subjective" preferences is noted. It does, then, make sense that an individual should morally *disapprove* (in terms of his "ethical" preferences) of an unequal income distribution which benefits him financially, and should still *prefer* it (in terms of his "subjective" preferences) to a more egalitarian one or should even *fight* for it—behavior morally regrettable but certainly not logically inconceivable.

Arrow's distinction between an individual's "tastes" (which order social situations only according to their effects on his own consumption)and his "values" (which take account also of external economies and diseconomies of consumption and of ethical considerations, in ordering social situations) does not meet the difficulty, since it does not explain how an individual can without inconsistency accept a social welfare function conflicting with his own "values." This can be understood only if his social welfare functions represents preferences of another sort than his "values" do. (Of course, in my terminology Arrow's "values" fall in the class of "subjective" preferences and not in the class of "ethical" preferences, as is easily seen from the way in which he defines them.)

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

posed in Sections II and III—namely, Postulates D, E, and *c*—can be regarded as simply an implicit definition of what sort of "impartial" or "impersonal" attitude is required to underlie "ethical" preferences: these postulates essentially serve to exclude nonethical subjective preferences from social welfare functions. But this aim may also be secured more directly by explicitly defining the impartial and impersonal attitude demanded.

I have argued elsewhere[15] that an individual's preferences satisfy this requirement of impersonality if they indicate what social situation he would choose if he did not know what his personal position would be in the new situation choosen (and in any of its alternatives) but rather had an equal *chance* of obtaining any of the social positions[16] existing in this situation, from the highest down to the lowest. Of course, it is immaterial whether this individual does not in fact know how his choice would affect his personal interests or merely disregards this knowledge for a moment when he is making his choice. As I have tried to show,[17] in either case an impersonal choice (preference) of this kind can in a technical sense be regarded as a choice between "uncertain" prospects.

This implies, however, without any additional ethical postulates that an individual's impersonal preferences, if they are rational, must satisfy Marschak's axioms and consequently must define a cardinal social welfare function equal to the arithmetical mean[18] of the utilities of all individuals in the society (since the arithmetical mean of all individual utilities gives the actuarial value of his uncertain prospect, defined by an equal probability of being put in the place of any individual in the situation chosen).

More exactly, if the former individual has any objective criterion for comparing his fellows' utilities with one another and with his own (see Sec. V), his social welfare function will represent the unweighted mean of these utilities, while in the absence of such an objective criterion it will, in general, represent their weighted mean, with arbitrary weights depending only on his personal value judgments. In the former case social welfare will in a sense be an objective quantity, whereas in the latter case it will contain an important subjective element; but even in this latter case it will be something very different from the utility function of the individual concerned.[19]

V

There is no doubt about the fact that people do make, or at least attempt to make, interpersonal comparisons of utility, both in the sense of comparing different persons' total satisfaction and in the

[15] See my "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking," *Journal of Political Economy*, LXI (October, 1953), 434–35.

[16] Or, rather, if he had an equal chance of being "put in the place of" any individual member of the society, with regard not only to his objective social (and economic) conditions, but also to his subjective attitudes and tastes. In other words, he ought to judge the utility of another individual's position not in terms of his own attitudes and tastes but rather in terms of the attitudes and tastes of the individual actually holding this position.

[17] *Op. cit.*

[18] Obviously, the (unweighted or weighted) *mean* of the individual utilities defines the same social welfare function as their *sum* (weighted by the same relative weights), except for an irrelevant proportionality constant.

[19] The concept of ethical preferences used in this section implies, of course, an ethical theory different from the now prevalent subjective attitude theory, since it makes a person's ethical judgments the expression, not of his subjective attitudes in general, but rather of certain special unbiased impersonal attitudes only. I shall set out the philosophic case for this ethical theory in a forthcoming publication. (For a similar view, see J. N. Findlay, "The Justification of Attitudes," *Mind*, N.S., LXIII [April, 1954], 145–61.)

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

sense of comparing increments or decrements in different persons' satisfaction.[20] The problem is only what logical basis, if any, there is for such comparisons.

In general, we have two indicators of the utility that *other* people attach to different situations: their preferences as revealed by their actual choices, and their (verbal or nonverbal) expressions of satisfaction or dissatisfaction in each situation. But while the use of these indicators for comparing the utilities that a *given* person ascribes to different situations is relatively free of difficulty, their use for comparing the utility that *different* persons ascribe to each situation entails a special problem. In actual fact, this problem has two rather different aspects, one purely metaphysical and one psychological, which have not, however, always been sufficiently kept apart.

The *metaphysical* problem would be present even if we tried to compare the utilities enjoyed by different persons with identical preferences and with identical expressive reactions to any situation. Even in this case, it would not be inconceivable that such persons should have different susceptibilities to satisfaction and should attach different utilities to identical situations, for, in principle, identical preferences may well correspond to different absolute levels of utility (as long as the ordinal properties of all persons' utility functions are the same[21]), and identical expressive reactions may well indicate different mental states with different people. At the same time, under these conditions this logical possibility of different susceptibilities to satisfaction would hardly be more than a metaphysical curiosity. If two objects or human beings show similar behavior in *all* their relevant aspects open to observation, the assumption of some unobservable hidden difference between them must be regarded as a completely gratuitous hypothesis and one contrary to sound scientific method.[22] (This principle may be called the "principle of unwarranted differentiation." In the last analysis, it is on the basis of this principle that we ascribe mental states to other human beings at all: the denial of this principle would at once lead us to solipsism.[23] Thus in the case of persons with similar preferences and expressive reactions we are fully entitled to assume that they derive the same utilities from similar situations.

In the real world, of course, different people's preferences and their expressive reactions to similar situations may be rather different, and this does represent a very real difficulty in comparing the utilities enjoyed by different people—a difficulty in addition to the metaphysical difficulty just discussed and independent of it. I shall refer to it as the *psychological* difficulty, since it is essentially a question of how psychological differences between people in the widest sense (for example,

[20] See I. M. D. Little, *A Critique of Welfare Economics* (Oxford, 1950), chap. iv. I have nothing to add to Little's conclusion on the *possibility* of interpersonal comparisons of utility. I only want to supplement his argument by an analysis of the *logical basis* of such comparisons. I shall deal with the problem of comparisions between total utilities only, neglecting the problem of comparisons between differences in utility, since the social welfare functions discussed in the previous sections contain only total utilities of individuals.

[21] Even identical preferences among uncertain prospects (satisfying the Marschak axioms) are compatible with different absolute levels of utility, since they do not uniquely determine the zero points and the scales of the corresponding cardinal utility functions.

[22] By making a somewhat free use of Professor Carnap's distinction, we may say that the assumption of different susceptibilities of satisfaction in this case, even though it would not be against the canons of *deductive* logic, would most definitely be against the canons of *inductive* logic.

[23] See Little, *A Critique of Welfare Economics*, pp. 56–57.

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

differences in consumption habits, cultural background, social status, and sex and other biological conditions, as well as purely psychological differences, inborn or acquired) affect the satisfaction that people derive from each situation. The problem in general takes the following form. If one individual prefers situation X to situation Y, while another prefers Y to X, is this so because the former individual attaches a *higher* utility to situation X, or because he attaches a *lower* utility to situation Y, than does the latter—or is this perhaps the result of both these factors at the same time? And, again, if in a given situation one individual gives more forcible signs of satisfaction or dissatisfaction than another, is this so because the former feels more intense satisfaction or dissatisfaction, or only because he is inclined to give stronger expression to his feelings?

This psychological difficulty is accessible to direct empirical solution to the extent to which these psychological differences between people are capable of change, and it is therefore possible for some individuals to make direct comparisons between the satisfactions open to one human type and those open to another.[24] Of course, many psychological variables are not capable of change or are capable of change only in some directions but not in others. For instance, a number of inborn mental or biological characteristics cannot be changed at all, and, though the cultural patterns and attitudes of an individual born and educated in one social group can be considerably changed by transplanting him to another, usually they cannot be completely assimilated to the cultural patterns and attitudes of the second group. Thus it may easily happen that, if we want to compare the satisfactions of two different classes of human beings, we cannot find any individual whose personal experiences would cover the satisfactions of both these classes.

Interpersonal comparisons of utility made in everyday life seem, however, to be based on a different principle (which is, of course, seldom formulated explicitly). If two individuals have opposite preferences between two situations, we usually try to find out the psychological differences responsible for this disagreement and, on the basis of our general knowledge of human psychology, try to judge to what extent these psychological differences are likely to increase or decrease their satisfaction derived from each situation. For example, if one individual is ready at a given wage rate to supply more labor than another, we tend in general to explain this mainly by his having a lower disutility for labor if his physique is much more robust than that of the other individual and if there is no ascertainable difference between the two individuals' economic needs; we tend to explain it mainly by his having a higher utility for income (consumption goods) if the two individuals' physiques are similar and if the former evidently has much greater economic needs (for example, a larger family to support).

Undoubtedly, both these methods of tackling what we have called the "psychological difficulty" are subject to rather large margins of error.[25] In general, the greater the psychological, biological, cultural, and social differences between two

[24] On the reliability of comparisons between the utility of different situations before a change in one's "taste" (taken in the broadest sense) and after it, see the first two sections of my "Welfare Economics of Variable Tastes," *Review of Economic Studies*, XXI, (1953–54), 204–8.

[25] Though perhaps it would not be too difficult to reduce these margins quite considerably (for example, by using appropriate statistical techniques), should there be a need for more precise results.

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

people, the greater the margin of error attached to comparisons between their utility.

Particular uncertainty is connected with the second method, since it depends on our general knowledge of psychological laws, which is still in a largely unsatisfactory state.[26] What is more, all our knowledge about the psychological laws of satisfaction is ultimately derived from observing how changes in different (psychological and other) variables affect the satisfactions an individual obtains from various situations. We therefore have no direct empirical evidence on how people's satisfactions are affected by the variables that, for any particular individual, are *not* capable of change. Thus we can, in general, judge the influence of these "unchangeable" variables only on the basis of the correlations found between these and the "changeable" variables, whose influence we can observe directly. For instance, let us take sex as an example of "unchangeable" variables (disregarding the few instances of sex change) and abstractive ability as an example of "changeable" variables. We tend to assume that the average man finds greater satisfaction than the average woman does in solving mathematical puzzles *because*, allegedly, men in general have greater abstractive ability than women. But this reasoning depends on the implicit assumption that differences in the "unchangeable" variables, if unaccompanied by differences in the "changeable" variables, are in themselves immaterial. For example, we must assume that men and women equal in abstractive ability (and the other relevant characteristics) would tend to find the same satisfaction in working on mathematical problems.

Of course, the assumption that the "unchangeable" variables in themselves have no influence is *ex hypothesi* not open to direct empirical check. It can be justified only by the a priori principle that, when one variable is alleged to have a certain influence on another, the burden of proof lies on those who claim the existence of such an influence.[27] Thus the second method of interpersonal utility comparison rests in an important sense on empirical evidence more indirect[28] than that underlying the first method. On the other hand, the second method has the advantage of also being applicable in those cases where no one individual can possibly have wide enough personal experience to make direct utility comparisons in terms of the first method.

In any case, it should now be sufficiently clear that interpersonal compari-

[26] Going back to our example, for instance, the disutility of labor and the utility of income are unlikely to be actually independent variables (as I have tacitly assumed), though it may not always be clear in which way their mutual influence actually goes. In any case, income is enjoyed in a different way, depending on the ease with which it has been earned, and labor is put up with in a different spirit, depending on the strength of one's need for additional income.

[27] This principle may be called the "principle of unwarranted correlation" and is again a principle of inductive logic, closely related to the principle of unwarranted differentiation referred to earlier.

[28] There is also another reason for which conclusions dependent on the principle of unwarranted correlation have somewhat less cogency than conclusions dependent only on the principle of unwarranted differentiation. The former principle refers to the case where two individuals differ in a certain variable X (in our example, in sex) but where there is no special evidence that they differ also in a certain other variable Y (in susceptibility to satisfaction). The latter principle, on the other hand, refers to the case where there is no ascertainable difference at all between the two individuals in any observable variable whatever, not even in X (in sex). Now, though the assumption that these two individuals differ in Y (in susceptibility to satisfaction) would be a gratuitous hypothesis in either case, obviously it would be a less unnatural hypothesis in the first case (where there is some observed difference between the two individuals) than in the second case (where there is none).

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

sons of utility are not value judgments based on some ethical or political postulates but rather are factual propositions based on certain principles of inductive logic.

At the same time, Professor Robbins[29] is clearly right when he maintains that propositions which purport to be interpersonal comparisons of utility often contain a purely *conventional* element based on ethical or political value judgments. For instance, the assumption that different individuals have the same susceptibility to satisfaction often expresses only the egalitarian value judgment that all individuals should be treated equally rather than a belief in a factual psychological equality between them. Or, again, different people's total satisfaction is often compared on the tacit understanding that the gratification of wants regarded as "immoral" in terms of a certain ethical standard shall not count. But in order to avoid confusion, such propositions based on ethical or political restrictive postulates must be clearly distinguished from interpersonal comparisons of utility without a conventional element of this kind.

It must also be admitted that the use of conventional postulates based on personal value judgments may sometimes be due not to our free choice but rather to our lack of the factual information needed to give our interpersonal utility comparisons a more objective basis. In effect, if we do not know anything about the relative urgency of different persons' economic needs and still have to make a decision, we can hardly avoid acting on the basis of personal guesses more or less dependent on our own value judgments.

On the other hand, if the information needed is available, individualistic ethics consistently requires the use, in the social welfare function, of individual utilities not subjected to restrictive postulates. The imposition of restrictive ethical or political conventions on the individual utility functions would necessarily qualify our individualism, since it would decrease the dependence of our social welfare function on the actual preferences and actual susceptibilities to satisfaction, of the individual members of the society, putting in its place a dependence on our own ethical or political value judgments (see nn. 5 and 6).

To sum up, the more complete our factual information and the more completely individualistic our ethics, the more the different individuals' social welfare functions will converge toward the same objective quantity, namely, the unweighted sum (or rather the unweighted arithmetic mean) of all individual utilities. This follows both from (either of two alternative sets of) ethical postulates based on commonly accepted individualistic ethical value judgments and from the mere logical analysis of the concept of a social welfare function. The latter interpretation also removes certain difficulties connected with the concept of a social welfare function, which have been brought out by Little's criticism of certain of Arrow's conclusions.

Of course, the practical need for reaching decisions on public policy will require us to formulate social welfare functions—explicitly or implicitly—even if we lack the factual information needed for placing interpersonal comparisons of utility on an objective basis. But even in this case, granting the proposed ethical postulates (or the proposed interpretation of

[29] See L. Robbins, "Robertson on Utility and Scope," *Economica*, N.S., XX (1953), 99–111, esp. 109; see also his *An Essay on the Nature and Significance of Economic Science* (2d ed.; London, 1948), chap. vi; and his "Interpersonal Comparisons of Utility," *Economic Journal*, XLIII (December, 1938), 635–41.

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

the concept of a social welfare function), our social welfare function must take the form of a weighted sum (weighted mean) of all individual utility functions, with more or less arbitrary weights chosen according to our own value judgments.

There is here an interesting analogy with the theory of statistical decisions (and, in general, the theory of choosing among alternative hypotheses). In the same way as in the latter, it has been shown[30] that a rational man (whose choices satisfy certain simple postulates of rationality) must act *as if* he ascribed numerical subjective probabilities to all alternative hypotheses, even if his factual information is insufficient to do this on an objective basis—so in welfare economics we have also found that a rational man (whose choices satisfy certain simple postulates of rationality and impartiality) must likewise act *as if* he made quantitative interpersonal comparisons of utility, even if his factual information is insufficient to do this on an objective basis.

Thus if we accept individualistic ethics and set public policy the task of satisfying the preferences of the individual members of the society (deciding between conflicting preferences of different individuals according to certain standards of impartial equity), our social welfare function will always tend to take the form of a sum (or mean) of individual utilities; but whether the weights given to these individual utilities have an objective basis or not will depend wholly on the extent of our factual (psychological) information.

[30] See Marschak's discussion of what he calls "Ramsey's norm," in his paper on "Probability in the Social Sciences," in *Mathematical Thinking in the Social Sciences*, ed. P. F. Lazarsfeld (Glencoe, Ill., 1954), Sec. I, esp. pp. 179–87; also reprinted as No. 82 of "Cowles Commission Papers" (N.S.).

For a survey of earlier literature see K. J. Arrow, Alternative Approaches to the Theory of Choice in Risk-taking Situations," *Econometrica*, XIX (October, 1951), 404–37, esp. 431–32, and the references there quoted.

This content downloaded from 147.46.17.142 on Sun, 04 Sep 2016 10:35:58 UTC
All use subject to http://about.jstor.org/terms

The Non-Existence of Representative Agents

Matthew O. Jackson* and Leeat Yariv†‡

December 2018

Abstract

We characterize environments in which there exists a representative agent: an agent who inherits the structure of preferences of the population that she represents. The existence of such a representative agent imposes strong restrictions on individual utility functions—requiring them to be *linear* in the allocation and additively separable in any parameter that characterizes agents' preferences (e.g., a risk aversion parameter, a discount factor, etc.). Commonly used classes of utility functions (exponentially discounted utility functions, CRRA or CARA utility functions, logarithmic functions, etc.) do not admit a representative agent.

JEL Classification Numbers: D72, D71, D03, D11, E24

Keywords: Representative Agents, Preference Aggregation, Revealed Preference, Collective Decisions

1 Introduction

1.1 Overview

Groups of people, in aggregate, can behave very differently from individuals. For example, the classic Sonnenschein-Mantel-Debreu Theorem (Sonnenschein, 1973; Mantel, 1974; Debreu, 1974) illustrated that even if individuals each satisfy standard conditions on their demand functions, some of the most vital of those conditions—e.g., the weak axiom of revealed preference—are lost when demands are aggregated.

This can be problematic, a model that admits arbitrary aggregate behavior is hard to work with. Thus, a usual approach is to assume aggregate behavior that is well-behaved,

*Department of Economics, Stanford University, the Santa Fe Institute, and CIFAR. http://www.stanford.edu/∼jacksonm e-mail: jacksonm@stanford.edu

†Department of Economics, Princeton University. http://www.princeton.edu/yariv e-mail: lyariv@princeton.edu

‡We thank William Brainard, three anonymous referees, and the Editor for very helpful suggestions. Financial support from the NSF (grant SES-1629613) is gratefully acknowledged.

Electronic copy available at: https://ssrn.com/abstract=2684776

implicitly presuming that agents in the underlying economy satisfy certain necessary restrictions for the model to be consistent. In particular, the literature has often assumed the existence of a well-behaved *representative agent*, one whose choices or preferences reflect the aggregate choices of society. The notion itself can be traced back to Edgeworth (1881) and Marshall (1890).[1] Since the publication of the Lucas Critique (1976), micro-founding economic models has become pervasive. Given the challenges of analyzing heterogeneous societies, the use of a representative agent as a modeling tool has become standard practice.

The existence of one sort of representative agent was theoretically founded in the mid-twentieth century by Gorman (1953, 1961). Gorman showed that in order to have a representative Marshallian demand function for an economy, such that the representative demand at the aggregate income level is equal to the sum of individual demands, agents' indirect utility functions have to take a particular restrictive form, termed the "Gorman Form," and have identical dependence on income.

Specifically, let $D(p, y)$ denote a Marshallian demand as a function of a vector of prices p and an income level y. Gorman's (1953) results imply that in order for there to exist a representative D such that

$$D(p, \sum_i y_i) = \sum_i D_i(p, y_i)$$

for all vectors of individual income levels y_i, it must be that the agents have linear and identical Engel curves, up to a parallel shift. As Gorman (1961) showed later, this imposes strong restrictions on the preferences in society—essentially requiring that they either be quasi-linear in income, or identical (up to a normalization) and homothetic.

Although Gorman's results are discouraging, most of the settings that researchers have analyzed with representative agents are not modeled through Marshallian demand functions nor do they require that a representation hold for all distributions of income. Most models involve decisions that are far more restricted. For instance, representative agents have been used to analyze how agents make consumption and savings decisions in the face of returns to savings that are impacted by various policies (e.g., Lucas, 1978), how agents choose their labor supply in the face of a tax schedule (e.g., Chamley, 1986), and how agents select public goods (e.g., Rogoff, 1990). Even though these decision problems involve maximizing a utility function with respect to some resource constraints, none of them fit into the Gorman setting.

Instead of presuming a common demand function, researchers assume that there is a single agent in the economy and specify that agent's preferences rather than their demand function. This allows a derivation of the agent's behavior in reaction to various influences and policies, as well as the analysis of inefficiencies and welfare. Ultimately, models often specify an agent with some characteristics a, who has a utility function of the form $V(x, a)$, where x can correspond to one dimension of consumption, a stream of consumption, etc. The agent's choice of x can then be subject to various feasibility constraints. If, for example, a captures the agent's income and x stands for a bundle of goods, then this formalization

[1] Edgeworth (1881) referred to a "representative particular," while Marshall (1890) referred to a "representative firm".

Electronic copy available at: https://ssrn.com/abstract=2684776

can be translated back into the Gorman form. But if, instead, a captures the level of risk aversion, or a discount factor, or a political ideology, etc., then this specification no longer fits the Gorman framework. In particular, the conditions under which this specification can represent a heterogeneous society do not follow as a corollary from Gorman's results.

Although researchers are generally careful not to claim that such a representative-agent formulation is a valid substitute for the analysis of a heterogeneous population, that hope is implicit. Such results are clearly of much more limited interest if there does not exist *any* population in which individual agents' preferences also take the form $V(x, a_i)$, with heterogeneous characteristics a_i, for which the analysis of one agent with preferences described by $V(x, a)$ for some characteristics a, could ever be consistent with.

Thus, we ask whether there exists *at least one* possible set of weights λ_i—e.g., representing the relative fractions of different groups in the population—such that if the population is comprised of λ_i agents in group i, each with a preference parameter a_i, then there exists some representative agent with preference parameter a for whom the utility of the average outcome is a proxy for the average utility. For private goods, this restriction takes the form:

$$V(\sum \lambda_i x_i, a) = \sum_i \lambda_i V(x_i, a_i), \tag{1}$$

while for common consumption, or a public good, this restriction takes the form:

$$V(x, a) = \sum_i \lambda_i V(x, a_i).$$

One requires this sort of formulation, for instance, when one looks for a policy that maximizes society's utilitarian welfare. For example, in Chamley (1986), a representative agent chooses labor supply, consumption, and savings decisions, responding to a taxation schedule given by government. The taxation schedule affects the relative returns to labor and savings, and thus the resulting consumption streams. Denote an agent i's labor, savings, and consumption choices, be it dynamic or static, by x_i. Denote that agent's risk-aversion parameter by a_i. The agent's utility can then be represented by some $V(x_i, a_i)$. The planner is presumed to choose a taxation schedule that maximizes the overall welfare subject to some revenue constraints. If the society is heterogeneous, then a planner who has a welfare function that is represented as a weighted sum of utilities of the agents would be evaluating the welfare of profiles of the agents' choices, $(x_i)_i$, by evaluating $\sum_i \lambda_i V(x_i, a_i)$ for some weights $(\lambda_i)_i$. In Chamley's formulation, which is quite typical, the planner maximizes the utility of a single "representative agent," and takes the form $V(x, a)$. In order for that agent to actually "represent" a heterogeneous society beyond a single agent, it must be that the planner's choice of x is somehow related to the actual choices $(x_i)_i$, either in per-capita value or some transformation thereof. Thus, the representative agent is evaluating some $V(\sum_i \lambda_i x_i, a)$, where the parameter a allows for any sort of transformation of $\sum_i \lambda_i x_i$ necessary to make things work. Having this evaluation of $V(\sum_i \lambda_i x_i, a)$ "represent" $\sum_i \lambda_i V(x_i, a_i)$, so that it yields the same welfare evaluations, is then the requirement that (1) hold.

Electronic copy available at: https://ssrn.com/abstract=2684776

The question that we pose is simply whether there exists *at least one* possible heterogeneous society that could be represented in this way. We show that the classes of utility functions that admit a representative agent are quite restricted.

Before providing our main results, and discussing more of the literature, we offer a couple of simple examples to illustrate the issues arising with the representative-agent assumption. One could imagine that if the source of heterogeneity in preferences is differences in discount factors or risk-aversion coefficients, then finding a representative agent would be easy, and there would exist some aggregate discount factor or risk-aversion coefficient that would capture the overall evaluation of utilitarian welfare. The following examples illustrate why this is not the case. We illustrate both of these with common consumption. The representation problem is even more restrictive when consumption is heterogeneous.

Example 1 (CRRA Utility Functions): Consider a population of n agents with CRRA (isoelastic) utility functions. Each agent i is identified by a CRRA parameter $a_i \in (0,1)$ and gets a utility from reward x given by

$$V(x;a_i) = \frac{x^{1-a_i}-1}{1-a_i}.$$

A representative agent would have utility proportional to some convex combination of the population. Namely, for a profile of coefficients of relative risk aversion $\mathbf{a} \equiv (a_1,...,a_n)$, up to an affine transformation, her utility of the common reward x would be given by:

$$U(x,\mathbf{a}) = \sum_i \lambda_i V(x;a_i) = \sum_i \lambda_i \frac{x^{1-a_i}-1}{1-a_i},$$

for some positive weights λ_i.

Straightforward calculations show that, whenever the a_i's are not all identical, the resulting coefficient of relative risk aversion,

$$-\frac{xU''(x,\mathbf{a})}{U'(x,\mathbf{a})} = \frac{\sum_i \lambda_i a_i x^{-a_i}}{\sum_i \lambda_i x^{-a_i}},$$

changes with x.

This means that the representative agent cannot be characterized by a utility function $V(x,a)$ that satisfies the same property (constant relative risk aversion) satisfied by the utility functions of all members of the population she represents.

If we move to a setting with private allocations, the problem becomes even starker. In this case, the weighted sum of agents' utilities is

$$\sum_i \lambda_i V(x_i;a_i) = \sum_i \lambda_i \frac{x_i^{1-a_i}-1}{1-a_i}.$$

This cannot be represented by any function of $\sum_i \lambda_i x_i$ when the x_i's differ, unless all the a_i's are 0—so all agents have to be risk neutral and evaluating a linear function.

Electronic copy available at: https://ssrn.com/abstract=2684776

Example 2 (Exponential Discounting): Consider a population of n agents who assess consumption streams with a horizon of T. Assume consumption at time t is $x(t) \in [0,1]$. Suppose each individual is characterized by a discount factor $a_i \in [0,1]$. The resulting utility of individual i is generated by exponentially discounting the consumption stream. That is,

$$V(\mathbf{x}; a_i) = \sum_{t=1}^{T} a_i^{t-1} x(t).$$

For a profile of discount factors, $\mathbf{a} \equiv (a_1, ..., a_n)$, the representative agent would be characterized, up to an affine transformation, by the utility function

$$U(\mathbf{x}; \mathbf{a}) = \sum_{i=1}^{n} \lambda_i V(\mathbf{x}; a_i) = \sum_{t=1}^{T} \sum_{i=1}^{n} \lambda_i a_i^{t-1} x(t),$$

for some positive weights λ_i, as before.

The effective discount factor corresponding to the representative agent at any time $t \geq 1$ is then

$$\frac{\sum_{i=1}^{n} \lambda_i a_i^t}{\sum_{i=1}^{n} \lambda_i a_i^{t-1}}.$$

Given any heterogeneity in the a_i's, this effective discount factor increases in t.[2] Again, the representative agent has a utility function that has fundamentally different properties than those corresponding to the underlying population. In this setting, the representative agent is time inconsistent and exhibits a present bias, even though all members of the population are time consistent.

Our main results below prove that these examples are not special. We fully characterize the classes of preferences for which representative agents exist. That is, we identify the conditions under which the population's preferences can be represented by an agent who has preferences in the same class. As we show, such a representative agent exists only if there are extreme restrictions on individual utility functions. When consumption is common, we show that only parametrized classes of utility functions that are *separable in agents' utility parameters* admit representative agents. The assumption that consumption is common applies to environments corresponding to members of a household sharing consumption and savings, or a community—a neighborhood, a state, or a country—benefiting from a common public good, etc. When consumption is private, corresponding to settings of consumer behavior, and encompassing the original examples offered by Lucas (1978), the existence of a representative agent turns out to be even more demanding. In this case, we show that *only utility functions that are linear in consumption and additively separable in agents' utility parameters* admit representative agents.

[2] For a more general analysis of the issue of finding a common discount factor, as well as related references, see Jackson and Yariv (2015).

Electronic copy available at: https://ssrn.com/abstract=2684776

The literature using representative agents almost never assumes linear utility functions, nor utility functions that are additively separable in agents' individual preference parameters. It then follows from our results that the commonly used classes of utility functions (logarithmic, weighting mean and variance, prospect theoretic, etc., in addition to exponentially-discounted utilities and CRRA or CARA utilities as described above) cannot be aggregated to generate a representative agent who is characterized by preferences from the same class. Thus, our results imply the only possible underlying societies that could rationalize most representative-agent models are ones in which some agents must have preferences outside of any standard class.

1.2 Related Literature

We are certainly not the first to point out issues with the use of representative agents. Beyond Gorman's contributions discussed above, the notion has endured scrutiny practically since its inception, and actively since the beginning of the twentieth century. For instance, one of its most vehement early criticisms appeared in Robbins (1928) (e.g., see Kirman, 1992 and Hartley, 1996 for surveys).

Nonetheless, as mentioned above, the publication of the Lucas Critique (1976) brought new life to micro-founding economic models using the representative-agent construct. Examples of ensuing models relying on representative agents abound. For instance, classical business cycle theories posit that observed aggregate fluctuations of an economy are partly driven by decisions of a representative household (e.g., Kydland and Prescott, 1982; King, Plosser, and Rebelo, 1988). In these models, the cyclical variation of aggregate consumption and employment is a consequence of the continuous optimization by a household that trades goods and leisure intertemporally in response to exogenous factors and movements in prices. Representative-agent models have also been used in the design of tax systems (e.g., see Chamley, 1986; Judd, 1985; and literature that followed), to estimate tax rates on factor incomes and consumption (e.g., Mendoza, Razin, and Tesar, 1994), moral hazard and adverse selection constraints in insurance markets (see Prescott and Townsend, 1984 and literature that followed), and so forth.

Although much of this literature mentions a background assumption of population homogeneity, the fragility of some of its conclusions to heterogeneity have been inspected only fairly recently and in particular contexts. For instance, An, Chang, and Kim (2009) consider an economy with some heterogeneity, incomplete capital markets, and indivisible labor supply. They illustrate that, in their setting, a "representative household" would correspond to an agent with non-concave utility. Constantinides (1982) studies the challenges of heterogeneity in an asset pricing model. Gollier (2001) shows that a mean-preserving spread of endowments among agents that are otherwise identical can affect the level of the equity premium and the risk-free rate in an Arrow-Debreu exchange economy. More recently, Kaplan, Moll, and Violante (2018) highlight the potential importance of accounting for household heterogeneity in monetary policy. Mazzoco (2004) studies households' saving decisions when

Electronic copy available at: https://ssrn.com/abstract=2684776

members have heterogenous risk preferences and make efficient joint choices. Among other results, he shows that an increase in risk aversion and prudence of one household member can reduce the household's risk aversion and prudence, absent harsh restrictions on preferences. In a completely different realm, Mongin (1998) considers a group of individuals satisfying the axioms of subjective expected utility. He shows that when either individuals' beliefs or utility functions are sufficiently heterogeneous, it is impossible to aggregate their preferences while respecting Pareto efficiency and the axioms of subjective expected utility. Golman (2011) examines the existence of representative agents in quantal-response equilibria. There are several other papers that illustrate the sensitivity of the representative-agent framework to heterogeneity of various sorts in particular environments. Our contribution is in highlighting a basic and general principle that drives all such observations.

As mentioned above, there is also a literature characterizing conditions under which aggregate demand, or aggregate behavior, features similar properties to underlying demands (for more recent references, see Chiappori and Ekelend, 1999). In contrast, our focus is on whether any modeler who assumes some properties of a representative agent's preferences must be making errors when presuming that these preferences are consistent with the preferences of an underlying heterogeneous population satisfying similar properties. As discussed above, this question is not answered by the demand-based literature, and yet it covers many, if not most, of the settings in which representative agents are used.

The insights of this paper are in the spirit of Jackson and Yariv (2015) and several of the papers cited there, which showed that there is no utilitarian aggregation of exponentially-discounted preferences that satisfies time consistency.[3] Here we show that such impossibilities are a much more pervasive phenomenon—applying to many different preference formulations and for quite general sources of heterogeneity—and can be argued quite directly.

Our results also provide insight into the observed differences between individual and group decision making. It is well-documented in the experimental and empirical literature that groups exhibit different behavioral patterns than individuals in various environments, ranging from choices between uncertain and thereby risky alternatives—chronicled in a large literature starting from Wallach, Kogan, and Bem (1962)—to choices of timing of events (see Ibanez, Czermak, and Sutter, 2009; Schaner, 2015; and references therein), allocation decisions (Cason and Mui, 1997; Ambrus, Greiner, and Pathak, 2015), etc. Our conclusions are in line with these observations when groups behave in line with some convex combination of their members' preferences. In fact, experimental evidence suggests that group members place substantial weight on utilitarian motives (e.g., Charness and Rabin, 2002; Jackson and Yariv, 2014). The characterization that we provide then suggests that if individual behavior is inconsistent with linear utilities, there is no reason to expect groups composed of such individuals to echo choices of any well-behaved individual.

[3]See also Apesteguia and Ballester (2016) for a similar approach considering several stochastic models of choice.

Electronic copy available at: https://ssrn.com/abstract=2684776

2 Representative Agents with Private Allocations

We first consider the case in which individuals each have their own allocation and the representative agent evaluates the aggregate/average allocation. For example, the allocation could stand for consumption, investment, and/or savings levels. Agents may exhibit heterogeneity in their discount factors, risk aversion parameters, or other preference parameters, as well as their endowments of human capital, wealth, and so on.

Formally, $n \geq 2$ agents evaluate allocations, generically denoted by x that come from some set D_x, which is a closed and convex subset of $\mathbb{R}^\ell_+$ for some ℓ. Also, we assume that there exists some $x \in D_x$ for which x is positive in all dimensions and $0 \leq y \leq x$ implies that $y \in D_x$.[4]

The heterogeneity of agents' preferences is captured by an index $a \in D_a$, where D_a is some index set. Depending on the application, the parameter a would represent an agent's risk aversion parameter, discount factor, endowment of human capital or wealth, etc.

Utility functions are functions $V : D_x \times D_a \to \mathbb{R}$ that are continuous in the allocation (the first variable).[5]

We say that there exists *a representative agent* with private allocations if there exists *some* $(\lambda_1, ..., \lambda_n) \in [0, 1)^n$, where $\sum_{i=1}^{n} \lambda_i = 1$,[6] such that for *some* $(a_1, \ldots, a_n) \in D_a^n$, there exists $\bar{a} \in D_a$ for which for all $(x_1, \ldots, x_n) \in D_x^n$:[7]

$$\sum_{i=1}^{n} \lambda_i V(x_i; a_i) = V(\sum_{i=1}^{n} \lambda_i x_i; \bar{a}).$$

When utility functions are concave, a Pareto optimal allocation is a solution to the maximization $\sum_{i=1}^{n} \lambda_i V(x_i; a_i)$ for some weights. The utilitarian social-welfare function corresponds

[4] We could also assume that D_x is an open set. Given that our utility functions are continuous, they extend to points of closure, and the results generalize. One can admit allocations with negative dimensions (so D_x is a closed convex subset of $\mathbb{R}^\ell$), simply by presuming there is an open ball in D_x containing 0 and the proof extends. In addition, by simply translating the utility functions (which preserves our definition of representative agents), it becomes without loss of generality to presume that $0 \in D_x$. So, technically, the necessary assumption for our results amounts to presuming that D_x contains an open ball.

[5] Using techniques from Corollary 3 of Rado and Baker (1987), one can extend the key lemmas in our proof to hold for Lebesgue measurable functions, but the proof will be more transparent with continuous functions on D_x and preferences are generally assumed to be continuous in representative-agent models.

[6] The proofs of Theorems 1 and 2 below can be extended to the case in which the $\sum_{i=1}^{n} \lambda_i$ is not required to be one, provided that D_x is unbounded above. It is then then still required that $\lambda_i > 0$ for at least two agents (which is implied above since $\lambda_i < 1$ for all i and the sum is 1). Otherwise, the setting boils down to one with a single agent and representation is trivial.

[7] Note that we have not placed any restrictions on how V depends on a, and so this allows $V(\cdot : \bar{a})$ to be any arbitrary function $W : D_x \to \mathbb{R}$ that is continuous in x.

Electronic copy available at: https://ssrn.com/abstract=2684776

to the special case in which each coefficient λ_i, $i = 1, ..., n$, is the fraction of the population characterized by preference parameter a_i and allocation x_i per person.

One could also contemplate situations in which welfare is assessed with a set of weights that does not necessarily coincide with the respective fractions of "types" in the population. In that case, the weights on the right-hand-side of the definition of a representative agent may differ from those on the left-hand-side. In Section 6, we show that such an assumption leads to even harsher restrictions on utility functions and requires them to be independent of the allocation altogether. We maintain the definition above since it corresponds to most applications of representative agents in the literature, and since it provides for more "conservative" insights in that it places weaker restrictions on preferences.

The existence of a representative agent is a type of convexity requirement on the space of utility functions. Indeed, consider the special case in which the x_i's are all equal. The existence of a representative agent requires that a convex combination of individuals' utility functions is in the same class to which those individual utility functions belong.

In our formulation, $\bar{a}$ is the representative agent's preference parameter. The representative agent's utility function is often assumed to take the form $be^{c-\bar{a}x}$, $(c+bx)^{\bar{a}}/\bar{a}$, $\bar{a}\ log(x)$, etc.[8]

Note that this setting fits a classic example in Lucas (1978). Lucas considers choices of consumption levels. He assumes individuals make a consumption versus savings decision each period, and the sequence of consumption decisions determines the remaining savings decisions. In that example, a_i would be the agent's initial holdings of the asset, and x_i would be the agent's consumption choice in a given period.

Lucas assumes a single consumer represents the entire population, which, as he notes, would be valid if all agents in the economy were identical. Clearly, the presumption that all agents in the economy are identical is made purely for tractability, as agents in most economies have very different wealth levels and thus face different consumption versus savings trade-offs (as was certainly noted before, see the above literature review). The hope is that the single "representative" consumer analysis provides insights into more general settings than those pertaining to a homogeneous society. If so, the modeler could consider a single representative agent, with utility $V(x, \bar{a})$ instead of considering a heterogeneous society. As in Lucas (1978), the ultimate goal is to analyze welfare-maximizing policies. In particular, we want to be able to compare two different policies—inducing, say, $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$—according to the social welfare function $\sum_i \lambda_i V(x_i; a_i)$. In order for the researcher to simplify her analysis and consider a representative agent with some preference parameter $\bar{a}$, whose preferences over aggregate bundles correspond to the welfare evaluation of the heterogeneous bundles, evaluations of the form $\sum_i \lambda_i V(x_i; a_i)$ should be equivalent to those of $V(\sum_i \lambda_i x_i; \bar{a})$ for some $\bar{a}$.

[8] In these formulations, b and c are taken as constants. For instance, the form $be^{c-\bar{a}x}$ with $b = 1$ and $c = 0$ would correspond to a representative agent with a CARA utility function and the form $(c+bx)^{\bar{a}}/\bar{a}$ with $c = 0$ and $b = 1$ would correspond to a representative agent with a CRRA utility function.

Electronic copy available at: https://ssrn.com/abstract=2684776

The following is the characterization of utility functions that admit the existence of a representative agent when allocations are private.

THEOREM 1 *There exists a representative agent $\bar{a}$ in the case of private allocations, relative to some $\lambda \in [0,1)^n$ and some $(a_1, \ldots, a_n) \in D_a^n$, if and only if $V(x;a) = c \cdot x + h(a)$ for all $x \in D_x$ and $a \in \{a_i \mid \lambda_i > 0\} \cup \{\bar{a}\}$, where $c \in \mathbb{R}^\ell$ and $h : D_a \to \mathbb{R}$ satisfies $h(\bar{a}) = \sum_i \lambda_i h(a_i)$.*

The structure characterized by Theorem 1 requires linearity in the allocation x and additive separability in the type parameter a. It is clearly not satisfied by utility functions that are commonly used in economic modeling. For example, strictly concave utility functions do not satisfy the restriction, nor do CRRA or CARA utility functions, nor do exponentially-discounted utilities. In such cases, the theorem implies that assuming a representative agent whose utility is taken from the same class of heterogeneous individuals' preferences would generate inaccurate estimates of aggregate behavior and welfare.

If we additionally require the representative-agent restriction hold for *all* preference profiles, then the structural implications of Theorem 1 then apply to all preference parameters. In particular, if we assume that $D_a = [0,1]$ and that $V(x;a)$ is continuous in a, the existence of a representative agent is tantamount to $V(x;a) = c \cdot x + h(a)$ for all $(x,a) \in D_x \times D_a$, where $c \in \mathbb{R}^\ell$ and $h : D_a \to \mathbb{R}$ is any continuous function.

The proofs of our results appear in Section 6. Intuitively, if a representative agent exists, a marginal change in the private allocation x_i of any agent i has a proportional effect on the allocation the representative agent considers (where the proportional factor corresponds to the individual's weight in society). The only way to get marginal utility calculations line up for all agents is to have linearity in x.

3 Representative Agents with Common Alternatives

The case of private allocations applies to most of the work built upon representative agents in macroeconomics and finance. We now expand the analysis to admit alternatives that are jointly evaluated. For example, in household decision making, expenditures and savings are often common across household members. Furthermore, common consumption is central to many models of political economy and public finance. In these models, agents make decisions over the level of some public good. As we now show, the existence of a representative agent in such environments is still very restrictive, but entails a different sort of separability.

We maintain the same basic structure of preference heterogeneity as above. For the rest of the paper, we add the conditions that D_a is $[0,1]$, that utility functions $V(x;a)$ are continuous in a, and that there exists at least one $x^* \in D_x$ for which $V(x^*;a)$ is strictly monotone (increasing or decreasing) in a.[9]

[9]The assumption that $V(x;a)$ is continuous in a simplifies our proof presentation, but is, in fact, not necessary. The continuity of $V(x^*;a)$ in a is implied by montonicity combined with the existence of a representative agent defined below, and is all that is required for our main result.

Electronic copy available at: https://ssrn.com/abstract=2684776

The restriction that $V(x^*; a)$ is monotone in a for some $x^* \in D_x$ is weak and satisfied for many classes of commonly used utility functions. For instance, exponential discounting, CRRA, and CARA satisfy the condition. Although we maintain this condition for presentation simplicity and since it allows for most cases covered in the literature, we note that the proofs imply that this condition can, in fact, be weakened to a requirement that $V(x^*; a)$ be piece-wise monotonic for some $x^* \in D_x$, which is satisfied for practically all preference specifications appearing in the literature.[10]

We say that there exists *a representative agent* with common alternatives if there exists *some* $(\lambda_1, ..., \lambda_n) \in [0, 1)^n$, where $\sum_{i=1}^{n} \lambda_i = 1$, such that for any $(a_1, ..., a_n) \in D_a^n$, there exists $\bar{a} \in [0, 1]$ for which:

$$\sum_{i=1}^{n} \lambda_i V(x; a_i) = V(x; \bar{a})$$

for all x.

In the case of common alternatives, an alternative definition that would require our restriction to hold for only one profile of preference parameters $(a_1, ..., a_n)$ could be trivially satisfied in a mechanical fashion. For instance, for any two continuous functions $f(x), g(x)$ such that $f(x) > g(x)$ for all x, defining $V(x; a) = f(x)$ for low values of a, $V(x; a) = g(x)$ for high values of a, and $V(x; a) = \frac{1}{2}f(x) + \frac{1}{2}g(x)$ for intermediate values of a would suffice for the existence of a representative agent with respect to particular weights and particular preference parameters. This is why we require the restriction to hold for *all* preference parameters.

Theorem 2 characterizes the class of utility functions that admit a representative agent.

THEOREM **2** *There exists a representative agent with common alternatives if and only if* $V(x; a) = h(a)f(x) + g(x)$ *for all* $(x, a) \in D_x \times D_a$, *for some continuous functions* $h(a), f(x)$, *and* $g(x)$ *such that* $h(\cdot)$ *is monotone, and* $f(x^*) \neq 0$.

Although the restrictions implied by the existence of a representative agent for common alternatives are weaker than those for private allocations, they are still sufficiently strong as to rule out nearly all commonly assumed utility functions. From the examples mentioned so far, exponential discounting, CARA and CRRA utility functions with risk-aversion parameters do not satisfy the restrictions of Theorem 2, nor do concave loss functions with bliss points serving as parameters—e.g., single-peaked preferences.

[10] Without some such assumption, one admits the possibility that all preferences are completely independent of types, in which case there is no meaningful heterogeneity in the population and everyone has the same preferences. In such cases, a representative agent exists trivially. This requirement is not needed in the case of private allocations since there agents can differ in their consumption. That potential variation imposes a stronger requirement on a representative agent, even with identical preferences.

Electronic copy available at: https://ssrn.com/abstract=2684776

One contrast between the common-alternative and private-allocation cases pertains to the class of concave utility functions. Certainly, a mixture of concave functions is concave. Thus, when considering the full class of concave functions, with common consumption, a representative agent does exist, and is characterized by the convex combination of agents' utility functions. Of course, that function must look quite different from the functions that are being aggregated (as required by the theorem above). This does not violate the theorem since there is no representation of the class of all concave functions that satisfies the monotonicity requirement.

4 Strongly Representative Agents

We now consider a more demanding notion of a representative agent. Under this variant, the representative agent's preference parameter $\bar{a}$ is the weighted average of individual agents' preference parameters. For instance, suppose an empiricist observes individual preference parameters with noise and erroneously assumes the population is homogenous. A natural estimate for the preference parameter corresponding to that population, as well as its legitimate representative agent under the assumption of homogeneity, would be the average of observed parameters. In the context of discount factor estimations, see the survey by Frederick, Loewenstein, and O'Donoghue (2002) for examples. With a large population of individuals, the estimated average parameter may not be biased. However, as we now show, welfare assessments based on the estimated utility function may be inaccurate. In fact, the classes of utility functions admitting such strongly representative agents are even more restrictive than those identified above.

As before, we start with the case of private allocations.[11]

We say that there exists a *strongly representative agent* with private allocations if there exists some $(\lambda_1, ..., \lambda_n) \in [0,1)^n$, where $\sum_{i=1}^{n} \lambda_i = 1$, such that for all $(a_1, \ldots, a_n) \in D_a^n$ and $(x_1, \ldots, x_n) \in D_x^n$:

$$\sum_{i=1}^{n} \lambda_i V(x_i; a_i) = V\left(\sum_{i=1}^{n} \lambda_i x_i; \sum_{i=1}^{n} \lambda_i a_i\right). \tag{2}$$

PROPOSITION 1 *There exists a strongly representative agent when allocations are private if and only if there exist constants b_1, b_2, and $c \in \mathbb{R}^\ell$ such that $V(x; a) = c \cdot x + b_1 a + b_2$ for all $(x, a) \in D_x \times D_a$.*

Proposition 1 states that a strongly representative agent exists only when utility functions are additively separable in the preference parameter and the allocation, and *linear* in both.

[11] As mentioned, we maintain our assumptions from Section 3 that D_a is $[0,1]$, that $V(x; a)$ is continuous in a, and that there exists at least one $x^* \in D_x$ for which $V(x^*; a)$ is strictly monotone (increasing or decreasing) in a.

Electronic copy available at: https://ssrn.com/abstract=2684776

The intuition is similar to that described for Theorem 1. If a strongly representative agent exists, a marginal change in either the allocation or the parameter of any individual has a proportional effect on the representative agent's allocation and utility parameter (where the proportional factor corresponds to the individual's weight in society). This implies the linear structure of utility functions.

With common alternatives, we analogously say that there exists *a strongly representative agent* with common alternatives if there exists some $(\lambda_1, ..., \lambda_n) \in [0, 1)^n$, where $\sum_{i=1}^{n} \lambda_i = 1$, such that for any $a_1, ..., a_n \in D_a^n$ and $x \in D_x$:

$$\sum_{i=1}^{n} \lambda_i V(x; a_i) = V\left(x; \sum_{i=1}^{n} \lambda_i a_i\right). \tag{3}$$

PROPOSITION **2** *There exists a strongly representative agent when alternatives are common if and only if there exist continuous functions $f(x), g(x)$ such that $V(x; a) = af(x) + g(x)$ for all $(x, a) \in D_x \times D_a$.*

The intuition behind Proposition 2 is again similar to the intuition provided for previous results. When an average representative agent exists, a marginal change in one individual's utility parameter has a proportional impact on the marginal change of the representative agent's utility parameter. This maps into a linearity requirement with respect to the utility parameter a.

5 Discussion

The assumption of a representative agent, who has preferences representing the aggregate of the population, is commonplace in modern economics. We have shown that for the representative agent to inherit the structure of preferences in the population that she represents, extreme restrictions need to be satisfied. In particular, utility functions need to be additively separable in allocations and parameters characterizing preferences; and, in the case including private allocations, *linear* in the allocation. Unfortunately, these restrictions are not satisfied by any commonly used classes of utility functions. For instance, a society in which each agent is characterized by a CRRA or CARA utility does not admit a representative agent with a similar utility function.

While others have pointed out the challenges of using a representative-agent model when individuals in society are aggregated in particular ways or interact strategically, the results in this paper are more fundamental. They illustrate an impossibility result for wide classes of utility functions, including practically all those studied in the literature. Our results make it imperative that researchers give more careful consideration to the use and formulation of representative agents, and model and account for heterogeneity directly.

Electronic copy available at: https://ssrn.com/abstract=2684776

6 Proofs

We begin with some lemmas that provide the key structure behind the proofs. The lemmas provide a variation on the analysis of Pexider's equation.[12] The proofs use techniques developed in Azcél (1966, 1969), Eichhorn (1978), and Diewert (2011).

We begin with a proof about a representation on one dimension and then use that to prove results for more dimensions.

LEMMA **1** *Let $f(x)$ be a continuous function on $[0,t]$. Suppose that, for some $\lambda \in (0,1)$,*

$$f(\lambda x + (1-\lambda)y) = f(\lambda x) + f((1-\lambda)y) \text{ for all } x, y \in [0,t]$$

then $f(x) = cx$ for all $x \in [0,t]$, where c is a scalar.

Proof of Lemma 1: Let $t' = \min\{\lambda, 1-\lambda\}t$.

We first show that for any positive integer k and any $z \in [0,t']$, $f(z) = kf(z/k)$.

Let $x = \frac{z}{\lambda k}$ and $y = \frac{(k-1)z}{(1-\lambda)k}$. Note that, since $z \leq t'$ then by construction, $x, y \in [0,t]$. Then,

$$f(z) = f\left(\frac{z}{k}\right) + f\left(\frac{(k-1)z}{k}\right).$$

For $k = 1$ this establishes the claim. For $k \geq 2$, writing $\frac{(k-1)z}{k} = \lambda\frac{z}{\lambda k} + (1-\lambda)\frac{(k-2)z}{(1-\lambda)k}$, it follows that

$$f(z) = f\left(\frac{z}{k}\right) + f\left(\frac{z}{k}\right) + f\left(\frac{(k-1)z}{k}\right) = 2f\left(\frac{z}{k}\right) + f\left(\frac{(k-2)z}{k}\right).$$

Continuing recursively, establishes that $f(z) = kf(z/k)$ for all $z \in [0,t']$ and for all positive integers k.

Next, we show that this implies that $f(x) = cx$ for all $x \in [0,t']$. Let $c = \frac{f(t')}{t'}$. For any $x = \frac{m}{n}t'$, where m and n are integers such that $m < n$, we have $\frac{x}{m} = \frac{t'}{n}$ and therefore from above $f(x) = \frac{m}{n}f(t') = cx$. From continuity, it follows that $f(x) = cx$ for all $x \in [0,t']$.[13]

Now, suppose $\min\{\lambda, 1-\lambda\} = \lambda$ so that $\lambda z \in [0,t']$ for all $z \in [0,t]$. Then:

$$\begin{aligned} f(z) &= f(\lambda z) + f((1-\lambda)z) = c\lambda z + f((1-\lambda)z) = \\ &= c\lambda z + f(\lambda(1-\lambda)z) + f((1-\lambda)^2 z) = c(\lambda z + \lambda(1-\lambda))z + f((1-\lambda)^2 z) = \\ &= cz\lambda\sum_{i=0}^{\infty}(1-\lambda)^i + \lim_{n\to\infty} f((1-\lambda)^n z) = cz + \lim_{n\to\infty} f((1-\lambda)^n z). \end{aligned}$$

Since $f(0) = 0$ (which follows from $f(0) = kf(0)$) and f is continuous, it follows that $f(z) = cz$ for all $z \in [0,t]$. A similar argument follows for $\min\{\lambda, 1-\lambda\} = 1-\lambda$. ∎

[12] Our results in an earlier version of this paper presumed analytic functions. We thank an anonymous reviewer for suggesting that some variation of Pexider's equation might be used to strengthen our results.

[13] Note that were D_x unbounded, e.g. $D_x = [0,\infty)$, the proof would be completed here.

Electronic copy available at: https://ssrn.com/abstract=2684776

LEMMA **2** *Let $f(x)$ be a continuous function on D_x such that there exists some $\lambda \in (0,1)$ for which*

$$f(\lambda x + (1-\lambda)y) = f(\lambda x) + f((1-\lambda)y) \text{ for all } x, y \in D_x.$$

Then there exists $c \in \Re^\ell$ such that $f(x) = c \cdot x$ for all $x \in D_x$.

Proof of Lemma 2: First, note that $f(0) = 0$, since $\lambda \times 0 + (1-\lambda) \times 0 = 0$ and so $f(0) = f(0) + f(0) = 2f(0)$.

Next, note that, by assumption, there exists $x \in D_x$ that is positive in all dimensions such that $0 \leq y \leq x$ implies $y \in D_x$. Let $D' = \{y : 0 \leq y \leq x\}$. Let D'_j be the subset of D' such that $y \in D'_j$ implies $y_k = 0$ for all $k \neq j$.

Applying Lemma 1 to each D'_j implies that for each dimension j, there exists c_j for which $f(y) = c_j y_j$ whenever $y \in D'_j$.

Next, let $d = \min\{\lambda, 1-\lambda\}$ and $D'' = \{y : 0 \leq \frac{y}{d} \leq x\}$. For $y \in D''$, abuse notation and write y_j to denote the vector that is the projection of y onto its j-th dimension, and y_{-j} to be $y - y_j$. Then, for any $y \in D''$, by the definition of D'', it follows that $(1-\lambda)y_j + (1-\lambda)\frac{y_{-j}}{1-\lambda} \in D'$. Therefore,

$$f(y) = f\left(\lambda y_j + (1-\lambda)y_j + (1-\lambda)\frac{y_{-j}}{1-\lambda}\right) = \lambda c_j y_j + f\left((1-\lambda)y_j + (1-\lambda)\frac{y_{-j}}{1-\lambda}\right).$$

Repeating the argument for the second term, we get $f(y) = c_j y_j + f(y_{-j})$. Then, iterating on the remaining dimensions,

$$f(y) = c \cdot y.$$

for any $y \in D''$.

Next, let $d' = \max\{\lambda, 1-\lambda\}$ and consider any $z \in D_x$ such that $d'z \in D''$. Then,

$$f(z) = f(\lambda z) + f((1-\lambda)z) = \lambda c \cdot z + (1-\lambda)c \cdot z = c \cdot z.$$

For any $z \in D_x$, there exists some t for which $d'^t z \in D''$, and so by iterating on the above argument, the result follows. ∎

LEMMA **3** *Let $f_1(x)$, $f_2(x)$, and $f_3(x)$ be continuous functions on D_x. If, for some $\lambda \in (0,1)$,*

$$f_1(\lambda x + (1-\lambda)y) = \lambda f_2(x) + (1-\lambda) f_3(y) \text{ for all } x, y \in D_x,$$

then there exist constants $a, b \in \mathbb{R}$ and $c \in \mathbb{R}^\ell$ such that

$$\begin{aligned} f_1(x) &= c \cdot x + \lambda a + (1-\lambda)b \\ f_2(x) &= c \cdot x + a, \\ f_3(x) &= c \cdot x + b. \end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=2684776

Proof of Lemma 3: Let $x = 0$. Then

$$f_1((1-\lambda)y) = \lambda f_2(0) + (1-\lambda)f_3(y) \text{ for all } y \in D_x.$$

Define $a \equiv f_2(0)$. Then,

$$f_3(y) = \frac{1}{1-\lambda}\left[f_1((1-\lambda)y) - \lambda a\right].$$

Similarly, if we define $b \equiv f_3(0)$, we get:

$$f_2(x) = \frac{1}{\lambda}\left[f_1(\lambda x) - (1-\lambda)b\right].$$

Plugging into the assumed equality, we have:

$$\begin{aligned} f_1(\lambda x + (1-\lambda)y) &= \lambda f_2(x) + (1-\lambda)f_3(y) = \\ &= f_1(\lambda x) + f_1((1-\lambda)y) - \lambda a - (1-\lambda)b. \end{aligned}$$

Define $f(x) \equiv f_1(x) - \lambda a - (1-\lambda)b$. Then, from the last equality we have

$$f(\lambda x + (1-\lambda)y) = f(\lambda x) + f((1-\lambda)y).$$

Lemma 2 then implies that $f(x) = c \cdot x$ and the result follows. ∎

6.1 Proofs Pertaining to Private Allocations

Proof of Theorem 1: Suppose that for some $\lambda_1, ..., \lambda_n \in [0, 1)$ for which $\sum_{i=1}^{n} \lambda_i = 1$, and some $(a_1, \ldots, a_n) \in D_a^n$, there exists $\bar{a} \in D_a$ such that for all $(x_1, \ldots, x_n) \in D_x^n$:

$$\sum_{i=1}^{n} \lambda_i V(x_i; a_i) = V(\sum_{i=1}^{n} \lambda_i x_i; \bar{a}).$$

There must exist some i for which $0 < \lambda_i < 1$. Let $x_i = x$, $x_j = y$ for all $j \neq i$. It follows that

$$V(\lambda_i x + (1-\lambda_i)y; \bar{a}) = \lambda_i V(x; a_i) + (1-\lambda_i)\sum_{j \neq i} V(y; a_j),$$

for any $x \in D_x$ and $y \in D_x$ and the characterization of V follows from Lemma 3. In particular, the first application of the lemma uses $f_1(x) = V(x; \bar{a})$, $f_2(x) = V(x; a_i)$, and $f_3(x) = \sum_{j \neq i} V(\cdot; a_j)$ and so gives

$$\begin{aligned} V(x; \bar{a}) &= c \cdot x + h(\bar{a}), \\ V(x; a_i) &= c \cdot x + h(a_i), \\ \textstyle\sum_{j \neq i} V(\cdot; a_j) &= c \cdot x + b_i, \end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=2684776

where $h(\bar{a}) = \lambda_i h(a_i) + (1-\lambda_1)b_i$. Iterating to apply the lemma to any j for which $\lambda_j > 0$ (and necessarily $\lambda_j < 1$ by definition), one similarly gets that $h(\bar{a}) = \lambda_j h(a_j) + (1-\lambda_j)b_j$. These iterative applications of the lemma imply that for any j for which $\lambda_j > 0$,

$$V(x; a_j) = c \cdot x + h(a_j)$$

and

$$\sum_{i \neq j} V(x; a_i) = c \cdot x + b_j.$$

Thus, putting all of these together, it follows that $b_j = \sum_{i \neq j : \lambda_i > 0} \lambda_i h(a_i)$. Therefore, it follows that

$$h(\bar{a}) = \sum_{i:\lambda_i > 0} \lambda_i h(a_i) = \sum_i \lambda_i h(a_i),$$

as claimed and any extension of $h(\cdot)$ to D_a would do. The converse follows directly. ∎

A weaker definition of the existence of a representative agent would impose that for some $\lambda_i, \lambda_i' \in [0,1)$ for $i = 1, .., n$ with $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n \lambda_i' = 1$, and some $(a_1, \ldots, a_n) \in D_a^n$, there exists $\bar{a} \in D_a$ such that for all $(x_1, \ldots, x_n) \in D_x^n$:

$$\sum_{i=1}^n \lambda_i V(x_i; a_i) = V(\sum_{i=1}^n \lambda_i' x_i; \bar{a}).$$

Without loss of generality, assume $\lambda_1' \in (0,1)$. As in the proof of Theorem 1, Let $x_1 = x$, $x_2 = x_3 = ... = x_n = y$. It follows that

$$\begin{aligned} V(\lambda_1' x + (1-\lambda_1')y; \bar{a}) &= \lambda_1 V(x; a_1) + (1-\lambda_1) V(y; a_2) \\ &= \lambda_1' V_1(x; a_1) + (1-\lambda_1') V_2(y; a_2), \end{aligned}$$

where

$$V_1(x; a_1) = \frac{\lambda_1}{\lambda_1'} V(x; a_1) \quad \text{and} \quad V_2(y; a_2) = \frac{1-\lambda_1}{1-\lambda_1'} V(y; a_2).$$

Lemma 3 then implies that whenever $\lambda_1 \neq \lambda_1'$, $V(x; a)$ is independent of x (noting that all three functions must have the same c, which is not possible if $\lambda_1 \neq \lambda_1'$ and $c \neq 0$) .

Proof of Proposition 1: The proof follows combining the implications of the functional forms from Theorem 1 together with Proposition 2, as both representations hold by either simply fixing any profile of a_is, or working with all agents having the same allocation. ∎

Electronic copy available at: https://ssrn.com/abstract=2684776

6.2 Proofs Pertaining to Common Alternatives

We start with the proof of Proposition 2, which is useful for proving Theorem 2.

Proof of Proposition 2: We show that (3) implies that there exist continuous functions $f(x), g(x)$ such that $V(x;a) = af(x) + g(x)$ for all x, a, as the converse is straightforward. Let $a_1 = r$, $a_2 = a_3 = ... = a_n = s$, and $x_1 = x_2 = ... = x_n = x$. Then, the existence of a strongly representative agent, for some $\lambda_i \in (0,1)$, implies that:

$$V(x; \lambda_i r + (1-\lambda_i)s) = \lambda_i V(x;r) + (1-\lambda_i)V(x;s)$$

and Lemma 3 (now applied on the a dimension), together with the continuity of V in a, imply the result. ∎

Proof of Theorem 2: Let

$$h(a) \equiv V(x^*; a).$$

Notice that $h(\cdot)$ is monotone in a and, therefore, from continuity, $\text{Im}_h\, D_a = \tilde{D}_{h,a}$ is a compact set and $h^{-1} : \tilde{D}_{h,a} \to D_a$ is continuous and monotone as well. Now let

$$G(x;a) = V(x; h^{-1}(a)).$$

By our assumption on V, for some $\lambda_1, ..., \lambda_n \geq 0, \sum_{i=1}^{n} \lambda_i = 1$, and for any $a_1, ..., a_n \in D_a^n$, there exists $\bar{a}$ such that for all x,

$$\sum_{i=1}^{n} \lambda_i G(x; a_i) = G(x; \bar{a}),$$

In particular:

$$\sum_{i=1}^{n} \lambda_i G(x^*; a_i) = \sum_{i=1}^{n} \lambda_i a_i = G(x^*; \bar{a}) = \bar{a}.$$

Therefore, G satisfies the assumptions of Proposition 2, so that there exist continuous functions $f(x), g(x)$ such that

$$G(x;a) = af(x) + g(x),$$

which, in turn, implies that

$$V(x;a) = h(a)f(x) + g(x).$$

This establishes the theorem, as the converse is immediate. ∎

Electronic copy available at: https://ssrn.com/abstract=2684776

7 References

Ambrus, Attila, Ben Greiner, and Parag Pathak (2015), "How Individual Preferences are Aggregated in Groups: An Experimental Study," *Journal of Public Economics*, 129, 1-13.

An, Sungbae, Yongsung Chang, and Sun-Bin Kim (2009), "Can a Representative-Agent Model Represent a Heterogeneous-Agent Economy?," *American Economic Journal: Macroeconomics*, 1(2), 29-54.

Apesteguia, Jose and Miguel A. Ballester (2016), "Stochastic Representative Agent," mimeo.

Azcél, János (1966), *Lectures on Functional Equations and their Applications*, New York: Academic Press.

Azcél, János (1969), *On Applications and Theory of Functional Equations*, New York: Academic Press.

Cason, Timothy N. and Vai-Lam Mui (1997), "A Laboratory Study of Group Polarisation in the Team Dictator Game," *The Economic Journal*, 107(444), 1465-1483.

Chamley, Christophe (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54(3), 607-622.

Charness, Gary and Matthew Rabin (2002), "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 117(3), 817-869.

Chiappori, Pierre-Andre and Ivar Ekeland (1999), "Aggregation and Market Demand: An Exterior Differential Calculus Viewpoint," *Econometrica*, 67(6), 1435-1457.

Constantinides, George M. (1982) "Intertemporal asset pricing with heterogeneous consumers and without demand aggregation." *Journal of business*, 55(2), 253-267.

Debreu, Gerard (1974), "Excess-demand Functions," *Journal of Mathematical Economics*, 1, 15-21.

Diewert, W. Erwin (2011) *Index Number Theory and Measurment Economics*, Chapter 2, Mimeo, University of British Columbia.

Edgeworth, Francis Y. (1881), *Mathematical Psychics*, London: Kegan Paul.

Eichhorn, Wolfgang (1978), *Functional Equations in Economics*, Reading, MA: Addison-Wesley Publishing Company.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue (2002), "Time Discounting and Time Preference: A Criticial Review," *Journal of Economic Literature*, 40, 351-401.

Gollier, Christian (2001), "Wealth Inequality and Asset Pricing," *The Review of Economic Studies*, 68(1), 181-203.

Golman, Russell (2011) "Quantal response equilibria with heterogeneous agents," *Journal of Economic Theory*, 146:5, 2013-2028.

Gorman, William M. (1953), "Community Preference Fields," *Econometrica*, 21(1), 63-80.

Gorman, William M. (1961), "On a class of preference fields." " *Metroeconomica* 13.2 53-56.

Hartley, James E. (1996), "The Origins of the Representative Agent," *Journal of Economic Perspectives*, 10(2), 169-177.

Ibanez, Marcela, Simon Czermak, and Mattias Sutter (2009), "Searching for a better deal – On the Influence of Group Decision Making, Time Pressure and Gender on Search Behavior,"

Electronic copy available at: https://ssrn.com/abstract=2684776

Journal of Economic Psychology, 30(1), 1-10.

Jackson, Matthew O. and Leeat Yariv (2014), "Present Bias and Collective Choice in the Lab," *The American Economic Review*, 104(12), 4184-4204.

Jackson, Matthew O. and Leeat Yariv (2015), "Collective Dynamic Choice: The Necessity of Time Inconsistency," *American Economic Journal: Microeconomics*, 7(4), 150-178.

Judd, Kenneth L. (1985), "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28(1), 59-83.

Kaplan, Greg, Ben Moll, and Giovanni L. Violante (2018), "Monetary Policy according to HANK," *The American Economic Review*, 18(3), 697-743.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (1988), "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21(2-3), 195-232.

Kirman, Alan P. (1992), "Whom or What does the Representative Agent Represent," *Journal of Economic Perspectives*, 6(2), 117-136.

Kydland, Finn E. and Edward C. Prescott (1982), "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345-1370.

Lucas, Robert E. (1976), "Econometric Policy Evaluation: A Critique," in K. Brunner and A. H. Meltzer (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of Carnegie-Rochester Conference Series on Public Policy, Amsterdam: North-Holland.

Lucas, Robert E. (1978), "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429-1445.

Mantel, Rolf R. (1974), "On the Characterization of Aggregate Excess-demand," *Journal of Economic Theory* , 7, 348-353.

Marshall, Alfred (1890), *Principles of Economics*, London: Macmillan.

Mazzoco, Maurizio (2004), "Saving, Risk Sharing, and Preferences for Risk," *The American Economic Review*, 94(4), 1169-1182.

Mendoza, Enrique G., Assaf Razin, and Linda L. Tesar (1994), "Effective Tax Rates in Macroeconomics: Cross-country Estimates of Tax Rates on Factor Incomes and Consumption," *Journal of Monetary Economics*, 34(3), 297-323.

Mongin, Philippe (1998), "The Paradox of the Bayesian Experts and State-dependent Utility Theory," *Journal of Mathematical Economics*, 29, 331-361.

Prescott, Edward C. and Robert M. Townsend (1984), "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard," *Econometrica*, 52(1), 21-45.

Radó, F., and Baker, John A. (1987), "Pexider's equation and aggregation of allocations," *Aequationes Mathematicae*, 32(1), 227-239.

Robbins, Lionel (1928), "The Representative Firm," *Economic Journal*, 38, 387-404.

Rogoff, Kenneth, (1990), "Equilibrium Political Budget Cycles," *American Economic Review*, 81:1, 21-36.

Schaner, Simone (2015), "Do Opposites Detract? Intrahousehold Preference Heterogeneity and Inefficient Strategic Savings," *American Economic Journal: Applied Economics*, 7(2), 135-174.

Electronic copy available at: https://ssrn.com/abstract=2684776

Sonnenschein, Hugo (1973), "Do Walras' Identity and Continuity Characterize the Class of Community Excess-demand Functions?," *Journal of Economic Theory*, 6, 345-354.
Wallach, Michael A., Nathan Kogan, and Daryl J. Bem (1962), "Group Influence on Individual Risk Taking," *ETS Research Bulletin Series*, 1962(1), 1-39.

Electronic copy available at: https://ssrn.com/abstract=2684776

Preference Aggregation After Harsanyi

Matthias Hild
Christ's College, Cambridge

Richard Jeffrey
Department of Philosophy, Princeton University

Mathias Risse
Department of Philosophy, Princeton University

August 3, 1998

Justice, Political Liberalism, and Utilitarianism: Proceedings of the Caen Conference in Honor of John Harsanyi and John Rawls

Edited by Maurice Salles and John A. Weymark

1 Introduction

Consider a group of people whose preferences satisfy the axioms of one of the current versions of utility theory, such as von Neumann-Morgenstern (1944), Savage (1954), or Bolker-Jeffrey (1965). There are political and economic contexts in which it is of interest to find ways of aggregating these individual preferences into a group preference ranking. The question then arises of whether methods of aggregation exist in which the group's preferences also satisfy the axioms of the chosen utility theory, and in which at the same time the aggregation process satisfies certain plausible conditions (e.g., the Pareto conditions below).

The answer to this question is sensitive to details of the chosen utility theory and method of aggregation. Much depends on whether uncertainty, expressed in terms of probabilities, is present in the framework and, if so, on how the probabilities are aggregated. The goal of this paper is (a) to provide a conceptual map of the field of preference aggregation—with special emphasis, prompted by the occasion, on Harsanyi's aggregation result and its relations to other results—and (b) to present a new problem ("Flipping") which we see as leading to a new impossibility result.

The story begins with some bad news, roughly 50 years old, about "purely ordinal" frameworks, in which probabilities play no role.[1]

> **Arrow's General Possibility Theorem (1950, 1951, 1963)** No universally applicable non-dictatorial method of aggregating individual preferences into group preferences can satisfy both the Pareto Preference condition (Unanimous individual preferences are group preferences) and the condition of Independence of Irrelevant Alternatives (Group preference between two prospects depends only on individual preferences between those same prospects).

But for nearly as long we have had some good news about the *vN–M* (von Neumann-Morgenstern) framework, in which probabilities play an essential role:[2]

> **Harsanyi's Representation Theorem (1955)** If individual and group preferences all satisfy the *vN–M* axioms, if ("Pareto Indifference") the group is indifferent whenever all individuals are, and if ("Strong Pareto") group preference agrees with that of an individual whenever no individual has the opposite preference, then group utility is a linear function W of individual utilities.

[1] Sen (1970) chapter 3 provides an excellent exposition.

[2] Here and in sec. 2 we draw on Weymark's (1991) reconstruction of the Harsanyi theorem.

Both news items are accurate. Their differences stem from differences in the requirements they place on utility functions that count as representing a given preference ordering. Arrow's framework was "purely ordinal" in the sense that for a utility function to count as a representation of a preference ordering he only required the numerical ordering of utilities to agree with the given preference ordering of prospects. But in the von Neumann-Morgenstern framework, where the agent is assumed to have preferences between lotteries that yield particular outcomes with particular numerical probabilities, there is a second requirement: The place of a lottery in the preference ranking must correspond to the *eu* (the expected utility, the probability–weighted sum) of the utilities of its possible outcomes. In the *vN–M* framework utilities of outcomes and *eu's* of lotteries are uniquely determined by the preference ranking once a zero and a unit have been chosen.

Actual personal probabilities play no part in Harsanyi's aggregation process: even though individuals may have personal probabilities and use them to solve their own decision problems, the process does not aggregate these into group probabilities; it is only personal utilities for outcomes that are aggregated. These will determine social *eu's* for chancy prospects in which outcomes are assigned definite numerical probabilities. Harsanyi's result will be our main concern in section 2.

In various other frameworks, e.g., Savage's (1954), and Bolker and Jeffrey's (1965), personal probabilities as well as utilities are deducible from preferences. If both group and individual preferences are to be placed in these frameworks we need to decide how to use personal probabilities as well as personal utilities in the aggregation process—a decision that does not arise in the von Neumann-Morgenstern framework. There are two ways to go: "*ex ante*" and "*ex post*". (Harsanyi's own method of aggregation falls into neither of these categories, since personal probabilities have no place in his *vN–M* framework.) Both methods of aggregation face serious problems.

In *ex ante* aggregation (sec. 3) group *eu* is a function—say, W—of individual *eu's*. Here the question arises: under what conditions is the aggregate $W(eu_1, \ldots, eu_I)$ of individual *eu's* itself an *eu*? The answer is bad news for those who hope to use aggregation as a way of arriving at compromises among conflicting judgments of fact or value:[3]

> **Generic ex ante Impossibility Theorem.** In general, *ex ante* aggregation is possible only for groups that are highly homogeneous in their probability judgments or in their value judgments.

In *ex post* aggregation (sec. 4) individual *eu's* are first disintegrated into utilities and probabilities. These are then aggregated separately into group

[3]Among the bearers of bad tidings have been Broome (1987), (1990), Seidenfeld *et al.* (1989), and Mongin (1995).

utilities and group probabilities, which are finally reintegrated into group *eu's*. This blocks the difficulty that led to the generic *ex ante* possibility theorem. But later in sec. 4 we announce some new bad news for the *ex post* approach:

> **Flipping.** In *ex post* aggregation utility and probability profiles for individuals exist relative to which group preference between some pair of options reverses repeatedly or even endlessly as the analysis is refined, although individual preferences remain constant throughout these analyses.

Finally, we note that Harsanyi's good news is not vitiated by the flipping phenomenon, and we suggest a connection between that fact and a certain sort of individualism.

2 Harsanyi's Utilitarianism

In "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility" (1955) Harsanyi challenged Arrow's (1951, p. 9) thesis "that interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility." Both saw themselves as responding to Bergson's (1938, 1948) challenge "to establish an ordering of social states which is based on indifference maps of individuals".[4] But their responses were radically different, with Arrow reaffirming the ordinalism of the 1930's, and Harsanyi rejecting it in favor of von Neumann and Morgenstern's revived cardinalism, which he applied to social as well as individual preferences.

Needing cardinal utilities for game theory, von Neumann and Morgenstern had turned the tables on ordinalists who had argued that the significance of a numerical utility function for prospects $X, Y, \ldots$ is exhausted by the corresponding order relation ($\succeq$) of preference–or–indifference on those prospects:

$$u(X) \geq u(Y) \text{ iff } X \succeq Y \tag{1}$$

By replacing the old prospects $X, Y, \ldots$ by the set $\mathcal{G}$ of all gambles among them[5] and replacing the utilities $u(X), u(Y), \ldots$ by expected utilities $eu(P), eu(Q), \ldots$ relative to $P, Q, \ldots \in \mathcal{G}$ they obtained a preference relation with definite cardinal significance:

$$eu(P) \geq eu(Q) \text{ iff } P \succeq Q \tag{2}$$

[4] The words are Arrow's (1951, p. 9).

[5] In these gambles the probabilities of outcomes must be specified explicitly in numerical form, e.g., "Victory with probability .1, defeat with probability .9". The contrast is with specifications in terms of events for which different individuals might have different probabilities, e.g., "Victory if Ruritania joins us, defeat if it does not".

Here $eu(P) = u(X)P(X) + u(Y)P(Y) + \ldots$, and similarly for $eu(Q)$. In the presence of (2) the full set of monotone increasing transformations of u under which (1) is preserved shrinks to its positive affine subset.

Note that it is not eu's (or their ratios, or differences) that are invariant, but ratios of differences, ratios of "preference intensities":[6]

$$\frac{eu(P) - eu(Q)}{eu(R) - eu(S)} = \frac{\text{intensity of preference for P over Q}}{\text{intensity of preference for R over S}} \tag{3}$$

Harsanyi used Marschak's formulation of the *vN–M* theory. In Marschak's framework the outcomes $X, Y, \ldots$ of gambles are off stage; it is only members of the set $\mathcal{G}$ that appear on stage. But each outcome off stage is represented on stage by the member of $\mathcal{G}$ that assigns probability 1 to it and 0 to all the others.

Marschak's Postulates[7]
For $P, Q, R, S \in \mathcal{G}$ and $x, \tilde{x} \in [0,1]$ where $\tilde{x} = 1 - x$:
M_1 $\quad \succeq$ is a complete, transitive relation on $\mathcal{G}$.
M_2 $\quad$ If $P \succ Q \succ R$ then $xP + \tilde{x}R \approx Q$ for some x.
M_3 $\quad P \succ Q \succ R \succ S$ for some P, Q, R, S.
M_4 $\quad$ If $Q \approx R$ then $xP + \tilde{x}Q \approx xP + \tilde{x}R$ for all P, x.

Representation Theorem
Given $M_1 - M_4$ there exist functions eu satisfying (2). These are unique up to a positive affine transformation.

InTRApersonal comparison of preference intensities
To compare i's preference intensity for P_1 over P_2 with that for P_3 over P_4, select suitable test–gambles P_{14}, P_{23} from $\mathcal{G}$, i.e.,

$$P_{14} = \frac{1}{2}P_1 + \frac{1}{2}P_4, \quad P_{23} = \frac{1}{2}P_2 + \frac{1}{2}P_3, \tag{4}$$

and note their relative positions in i's preference ranking. It will turn out that $eu_i(P_1) - eu_i(P_2) \geq eu_i(P_3) - eu_i(P_4)$ iff $P_{14} \succeq_i P_{23}$, for by (2), the three conditions (5) are equivalent:

$$\frac{eu_i(P_1) - eu_i(P_2)}{eu_i(P_3) - eu_i(P_4)} \geq 1, \ \frac{eu_i(\frac{1}{2}P_1 + \frac{1}{2}P_4)}{eu_i(\frac{1}{2}P_2 + \frac{1}{2}P_3)} \geq 1, \ P_{14} \succeq P_{23} \tag{5}$$

In a single episode of group decision making, the group (e.g., perhaps, a legislature) will choose from a small set of pairwise incompatible options (perhaps, bills for combinations of taxation and public expenditure). The set $\mathcal{G}$ of all probability distributions over those options is the common field of

[6] See remark (3) at the end of this section.
[7] $\succeq, \succ$, and $\approx$ are the relations of weak preference, strong preference, and indifference.

the group preference ranking $\succeq_0$ and the individual preference rankings $\succeq_i$ of group options. In Harsanyi's postulates the number 0 represents a group and the numbers $1, \ldots, I$ represent the individuals who make it up.

Harsanyi's postulates. For $i, j = 1, \ldots, I$ and $P, P_i, Q \in \mathcal{G}$:
H_1 All individuals' rankings $\succeq_i$ satisfy $M_1 - M_4$.
H_2 So does the group's ranking, $\succeq_0$.
H_3 *Functionality* : $P \approx_0 Q$ if $P \approx_i Q$ for all i.
H_4 *Uniqueness* : $\exists Q\ \forall i\ \exists P\ \forall j \neq i$ ($P \succ_i Q$ but $P \approx_j Q$).[8]
H_5 *Positivity* : $P \succeq_0 Q$ if $P \succeq_i Q$ for all i and $\succ_i$ for some i.

Harsanyi's Aggregation (= Representation) Theorem: Postulates $H_1 - H_5$ imply the existence of *eu*'s for the preferences $\succeq_0, \succeq_i$ that satisfy the condition $eu_0 = \sum_i eu_i$. These are unique up to a positive affine transformation.

For an accessible explanation of the axioms and a proof of a somewhat stronger form of this theorem, see Weymark (1991) sec. 3.[9]

When is individual i's preference intensity for P_1 over P_2 greater than (or less than, or equal to) individual j's for P_3 over P_4? This is the form that questions of interpersonal comparison of utilities take when individual and group preferences determine only ratios of differences of utilities as in (3) above. These may well be substantive questions, which people do sometimes manage to answer correctly by various devices appropriate to particular persons and their situation.[10] Answers to such questions guide the synthesis of group preferences out of individual ones.

But here we work backwards, from a group preference ranking that all find acceptable as an even–handed aggregation of their various preferences to the interpersonal comparison of individual utility differences which that ranking presupposes. Whether or not the individuals have accurately answered the substantive questions, their group ranking can be analyzed so as to discover what are in effect common judgments, right or wrong, of form "r = the ratio of i's preference intensity for P_1 over P_2 to j's for P_3 over P_4".

The idea is adequately illustrated in the case of a two–person group. Suppose that, somehow or other, individuals 1 and 2 have come to regard a particular preference ranking $\succeq_0$, satisfying $H_1 - H_5$ for the group constituted by the two of them, as an even–handed aggregation of their individual

[8] Harsanyi (1955) does not state H_4 as an axiom, but presupposes it in the first sentence of the proof of his Theorem V. Note that in H_4, P depends on i but Q does not.

[9] In his treatment, Weymark (1991, p. 272) permutes the first two quantifiers in H_4 to obtain a weaker axiom ("Independent Prospects") in which both P and Q depend on i, and which still yields uniqueness.

[10] See Harsanyi (1955, 1990) and Weymark's (1991) counterarguments. See also Jeffrey (1992), chapter 10.

preference rankings, $\succeq_1$ and $\succeq_2$. Then any function eu_0 representing $\succeq_0$ can be used to determine whether or not given functions eu_1, eu_2 representing the personal rankings are interval–commensurate:

Interval Commensuration Revealed Retrospectively.
If $H_1 - H_5$ hold with $I = 2$, then by H_4 there are $P_1, P_2, Q \in \mathcal{G}$ satisfying (a) and (b).

$$\text{(a) } P_1 \succ_1 Q \approx_1 P_2 \qquad \text{(b) } P_2 \succ_2 Q \approx_2 P_1$$

Representations eu_1, eu_2 of $\succeq_1, \succeq_2$ will be called "interval commensurate" iff some (and, so, every) representation eu_0 of $\succeq_0$ satisfies

$$\frac{eu_1(P_1) - eu_1(Q)}{eu_2(P_2) - eu_2(Q)} = \frac{eu_0(P_1) - eu_0(Q)}{eu_0(P_2) - eu_0(Q)} \tag{6}$$

Given conditions (c) and (d), formula (6) follows from conditions (a) and (b):[11]

$$\text{(c) } eu_0(P) = eu_1(P) + eu_2(P)$$
$$\text{(d) } eu_0, eu_1, eu_2 \text{ represent } \succeq_0, \succeq_1, \succeq_2$$

Note that differences of form $eu_j(P) - eu_j(Q)$ are not uniquely determined by the corresponding relation $\succ_j$, but ratios of such differences *are*—e.g., as on the right–hand side of (6).[12] Then in view of (6) the ratio of differences for $j = 1, 2$ (i.e. a ratio of interval commensurate preference intensities) is fixed by certain group preference intensities, and thus, in view of Marschak's representation theorem, by the group's preference ranking.[13]

We conclude this section with three remarks:

(1) Of course questions of interpersonal comparison are idle if Harsanyi's aggregation theorem is vitiated by an *ex ante* impossibility theorem, as some would seem to think;[14] but it is not so. On the contrary, Harsanyi's method

[11] *Proof.* By (a), (b), (d) the denominator on the left of (6) is non–null. Now operate on the right: First apply (c) to the four eu_0 terms; by (a) and (b) we may now substitute $eu_1(Q)$ for $eu_1(P_2)$ and $eu_2(Q)$ for $eu_2(P_1)$; after cancelling the $\pm eu_2(Q)$ terms in the numerator and the $\pm eu_1(Q)$ terms in the denominator, equation (6) becomes an identity.

[12] The social preference ranking determines eu_0 uniquely up to an affine transformation $eu_0 \mapsto a \cdot eu_0 + b$ with $a > 0$, and the value of the right–hand side of (6) is unaffected by any such transformation because we can drop $b - b$ from the numerator and the denominator, after which the a's in the numerator cancel those in the denominator.

[13] By confining this commensuration technique to consecutive pairs $(1, 2), \ldots (I - 1, I)$ of individuals, Harsanyi's aggregation result might be obtained with H_4 weakened to this: $\forall i = 1, \ldots, I - 1[\exists P\, \exists Q(P \succ_i Q \text{ but } P \approx_{i+1} Q) \text{ and } \exists P\, \exists Q(P \succ_{i+1} Q \text{ but } P \approx_i Q)]$.

[14] Broome (1991, pp. 160, 201) *seems* to be saying that Harsanyi's scheme is vitiated in that way, but this impression is created by his broad use of the term "Harsanyi's theorem" not only for Harsany's own aggregation theorem (above), but for variants of it in which the *vN–M* framework is replaced by frameworks like those of Savage and Bolker–Jeffrey, in which personal probabilities figure alongside utilities.

of utility aggregation is immune to *ex ante* impossibility theorems simply because, as we have observed, it is neither *ex ante* nor *ex post.*

(2) The object of the *vN–M* and Marschak axiomatic treatments of preference was to counter the view that game theory's cardinal concept of utility was metaphysical nonsense. Since there were no such qualms about the long–run frequency view of cardinal *probability*, von Neumann and Morgenstern adopted that view in their exposition (p. 19):

> "Probability has often been visualized as a subjective concept, more or less in the nature of an estimation. Since we propose to use it in constructing an individual, numerical estimation of utility, the above view of probability would not serve our purpose. The simplest procedure is, therefore, to insist upon the alternative, perfectly well founded interpretation of probability as frequency in long runs. This gives directly the necessary numerical foothold.[2]
>
> [2]If one objects to the frequency interpretation of probability then the two concepts (probability and preference) can be axiomatized together. This too leads to a satisfactory numerical concept of utility which will be discussed on another occasion."

But what made Harsanyi adopt the vN–M framework was no commitment to a long run frequency view of probability; rather, it was his view of probability as (in von Neumann and Morgenstern's words, above) "a subjective concept, more or less in the nature of an estimation." Harsanyi was that sort of subjectivist well before Savage showed how personal probabilities of events can be recovered from personal *eu*'s—i.e., ultimately, from personal preferences among gambles on those events. From the start, Harsanyi took it for granted that your expectations concerning random variables would be represented by probability–weighted means in which the probabilities are "subjective", representing your own uncertain judgments.[15] He could use the *vN–M* utility theory without the sorts of qualms mentioned in the unkept promise made in their footnote 2, above—a promise that Savage later made good.[16] The *vN–M* theory provided Harsanyi with a random variable u that could be combined with personal probabilities, exogenous to that theory, to yield exogenous personal *eu's*. It was Ramsey (1931) and Savage (1954) who provided decision theories with endogenous personal probabilities as well as utilities.

[15]In this sense of the term, Carnap (1945, 1950, 1962) was also a subjectivist. Like Carnap, Harsanyi took the legitimate source of the differences between different people's "subjective" probability judgments to be differences in the data on which those judgments are based.

[16]Savage (1954) points out that Ramsey (1931) had made the promise good decades earlier.

(3) We *form* our preference ranking of acts under uncertainty by judging the probabilities and utilities of the possible outcomes of those acts as best we can. From this constructive point of view it is our probability and utility judgments that determine our *eu's*, and our *eu's* that determine our preferences. This way of forming preferences has been tuned up over the past three centuries and more. A high–tech version can be found in Raiffa's 1968 "How to Think" book for MBA's. And a low–tech version had the place of honor at the end Arnauld's 1662 "How to Think" book for the innumerate:

> "To judge what one must do to obtain a good or avoid an evil, it is necessary to consider not only the good and the evil in itself, but also the probability that it happens or does not happen; and to view geometrically the proportion that all these things have together."

Representation theorems are analytical, not constructive: given a fully formed preference ranking that satisfies the axioms, they assure us of the existence of *eu* functions that represent the ranking, and of the uniqueness of those representations up to a positive linear transformation. But of course we do not have fully formed preference rankings over all the prospects that interest us. (If we did, we could simply read the solutions to our decision problems off them.) The problem in decision making is the constructive one of forming or discovering preferences we can live with. From the analytical point of view taken in representation theorems it is true enough that an *eu* function is a mere representation of a given preference ranking. But from the point of view of decision makers it is their preference rankings that merely represent their *eu* functions, which in turn merely reflect their probabilities and utilities.

3 Aggregation ex ante

We now turn to frameworks for preference in which actual personal probabilities play a role—in particular, the Savage framework in the present section, and the Bolker–Jeffrey framework in sec. 4. In the *vN–M* framework numerical probabilities of lottery outcomes are specified explicitly, and actual personal probabilities play no role. In the new frameworks personal probabilities play a central role, and are recoverable from the given preference ranking if it satisfies the relevant axioms. Here are thumbnail sketches of the two frameworks:

Savage. Preference is a relation between "acts". Acts are represented by functions f, each of which assigns to each possible "state of nature" s a definite "consequence" $f(s)$. If the act is betting \$10 on Bluebell to win, then we have

$$f(s) = \text{ "be \$10 richer" if Bluebell wins in state } s$$

$f(s) =$ "be \$10 poorer" if Bluebell does not win in state s

The expected utility $eu(f)$ of an act f is the mean value of $u(f(s))$ for all states of nature s, weighted with the individual's personal probability distribution P over the states of nature. Savage's representation theorem guarantees the existence of functions u and P which together represent the preference ranking in the sense that act f is preferred to act g if and only if $eu(f)$ is greater than $eu(g)$.

Bolker–Jeffrey. Here preference is a relation between "events" A (i.e., between the same things to which probabilities are attributed), and utilities $u(s)$ are attributed to states of nature s. Performing an act is a matter of making some particular event true, e.g., the event of betting \$10 on Bluebell to win. Given a utility function u and a probability function P, the "desirability" $des(A)$ of an event A is defined as the mean value of $u(s)$ for all states of nature s, weighted with the conditional probability distribution $P(-|A)$. According to Bolker's representation theorem truth of event A is preferred to truth of event B if and only if the desirability of A is greater than that of B.[17]

Desirability can be defined as conditional expectation of utility,[18] $des(A) = E(u|A) = \int_A u\, dP(-|A)$. In the discrete case, where the set S of states of nature is finite or countably infinite, the integral becomes a sum:

$$des(A) = \sum_{s \in A} u(s) P(\{s\}|A) \tag{7}$$

Example: "Dessert?" Consider Alice's problem of deciding whether to say "Yes" or "No" in answer to this question. She is sure that dessert would turn out to be chocolate ice cream (c), vanilla ice cream (v) or pie (p), i.e., $Dessert = \{c, v, p\}$—but she does not know which.

Data: For these possibilities her probabilities conditionally on *Dessert* are $P_{Alice}(\{c\}|\{c,v,p\}) = P_{Alice}(\{v\}|\{c,v,p\}) = \frac{1}{8}$ and $P_{Alice}(\{p\}|\{c,v,p\}) = \frac{3}{4}$, and her utilities are $u_{Alice}(c) = 68$, $u_{Alice}(v) = -100$, $u_{Alice}(p) = 16$. For the remaining possibility, *None ("n"),* her utility is $u_{Alice}(n) = 0$.

[17] For accessible overviews of the theory see Bolker (1967), Jeffrey (1983), and Broome (1990). For important modifications of the theory see Joyce (1992) and Bradley (1997).

[18] Bolker's (1965, 1966, 1967) representation theorem guarantees existence of a function *des* representing preference between elements of a Boolean algebra—but on assumptions under which the algebra cannot be a field of sets (of "states"). Under those assumptions the function *des* is not the conditional expectation of any function $u(s)$. But of course existence of such a representation when those assumptions hold does not imply non–existence when they do not. Jeffrey (1992, chapter 15) recasts Bolker's theorem in a form applicable to Boolean algebras of sets of states—algebras on which $des(A)$ can be defined as $E(u|A)$ after all. (The gimmick is like the one Kolmogorov [1948, 1995] uses to transform fields of sets on which probability measures exist into Boolean algebras of the sort postulated in Bolker's theorem.)

Solution: As Alice sees it, the states of nature form the set $S = \{c, v, p, n\}$ and the event *Dessert* has desirability $des_{Alice}(\{c, v, p\}) =$

$$\sum_{s\in\{c,v,p\}} u_{Alice}(s)P_{Alice}(\{s\}|\{c, v, p\}) = 68(\tfrac{1}{8}) - 100(\tfrac{1}{8}) + 16(\tfrac{3}{4}) = 8.$$

Then since $des_{Alice}(\{n\}) = u_{Alice}(n) = 0 < 18$, Alice does want dessert: $\{c, v, p\} \succ_{Alice} \{n\}$. Similar calculations show that she prefers pie to ice cream: $des_{Alice}(\{p\}) = u_{Alice}(p) = 16 > des_{Alice}(\{c, v\}) = -16$. Note that until she makes her decision, Alice's probability for dessert will be strictly between 0 and 1, e.g., $P_{Alice}(Dessert)$ might be 1/2, or 7/10, or whatever. But the actual value makes no difference to her decision, since the probabilities of interest are all conditional on *Dessert*, and we suppose (see Jeffrey 1996) that those remain constant as the unconditional probability of *Dessert* varies.

Where Savage assigns probabilities to events independently of what act is being performed, Bolker and Jeffrey assign conditional probabilities to events given acts. (Since acts are not events for Savage, these conditional probabilities make no sense for him.) The Bolker–Jeffrey framework allows probabilities to be updated either by observation or by decision: the updated unconditional probability will be the prior conditional probability given the event observed or chosen. But in the Savage framework choice of an option cannot affect probabilities. Note, too, that Savage's treatment is problematical in cases where it is important to consider players' probabilities for other players' performing various acts, as in interactive decision theory (= game theory).

Two Dismal Possibility Theorems. Here we note two specifications of the generic *ex ante* possibility theorem indicated in sec. 1. The species is Mongin's (1995) modification of the Savage framework—a modification in which an additional postulate assures σ-additivity of the probability measure.

Let $\succeq_i$, u_i, and P_i be individual i's preference relation, utility function, and probability function. Mongin adopts analogs of Harsanyi's "Pareto" conditions H_3 (functionality) and H_5 (positivity). To give these postulates material to work on he adds an assumption of diversity (linear independence) of the various individuals' probabilities or utilities. Either assumption implies the following condition, which is an analog of H_4:

> **Independence.** Each individual i has some preference $f \succ_i g$ where all others are indifferent: $f \succ_i g$, but $f \approx_j g$ if $j \neq i$.

Finally, Mongin postulates a minimal *Agreement* condition:

> **Agreement.** There exist consequences c_1, c_0 such that all individuals i assign higher utility to the former: $u_i(c_1) > u_i(c_0)$.

Mongin uses the term "overall dictator" for an individual whose probabilities and preference intensities are the same as Society's. Of course, such individuals need not really be dictators—e.g., they might be immensely public-spirited citizens, or ones whose personal attitudes are somehow formed by the same causes as the group's; or the "dictator" might be chosen by lot, or by vote; or the coincidence might be the result of blind chance. As Hylland and Zeckhauser (1979) point out, real dictatorship would be a property of the preference aggregation scheme, W, i.e., the property of assigning a particular individual's preferences to society regardless of what probabilities and utilities the others may have. But anyway it would be a very restrictive possibility theorem that implied the existence of Mongin's "dictators".

Below, Mgn_1 and Mgn_2 are weaker consequences of Mongin's main possibility results.[19] "Positivity" is the analog of H_5 (i.e., the group prefers f to g if some member does and none prefer g to f), and "Functionality" is the analog of H_3. In Mgn_2 we use the terms "diverse" and "clone" as follows:

Probability clones are individuals with identical probability functions.

Utility clones are individuals with affine equivalent utility functions.

Diversity of the individuals' probability functions means that all are distinct and none are weighted averages of others.

> **Mgn$_1$** : In the modified Savage framework with functionality and positivity there will be an overall "Dictator" if no individual probability or utility function is a linear combination of others.

> **Mgn$_2$** : In the modified Savage framework, Positivity and Agreement together imply (1) and (2):
> (1) If the probability functions are diverse, all are utility clones.
> (2) If not all are probability clones, some are utility clones.

Politics makes strange bedfellows. Results like Mgn_1 and Mgn_2 may seem less disturbing—only to be expected—in the light of the well known fact that unanimity about the relative ranking of two options may be based on quite incompatible assessments of probability or utility. Raiffa (1968, p. 230) offers a simple, striking example, with two options (a_1, a_2), two states of nature (θ_1, θ_2), and a pair of experts, Alice and Bob, who are indifferent between the options for very different reasons: Alice assigns probabilities $.2, .8$ to θ_1, θ_2 and utilities $1, 0, .5, 1$ to $a_1\theta_1, a_2\theta_1, a_1\theta_2, a_2, \theta_2$, while Bob assigns probabilities $.8, .2$ and utilities $.5, 1, 1, 0$ to the same states and act-state pairs. These experts have the same expected utilities (.6 for a_1, .8 for a_2) but for precisely opposite reasons. As Raiffa argues, such examples cast doubt on the seemingly ineluctable functionality principle, H_3. This idea is pushed further in the next section, under "flipping".

[19] See Mongin's (1995) observation 1 on p. 341, and proposition 7 on pp. 343-4.

4 Aggregation ex post

The strange bedfellows phenomenon may be seen as a warning against muddling judgments of fact and value, and as a call to take the *ex post* stance, in which members' *eu*'s are not directly aggregated, but are first analyzed into probabilities and utilities—which are aggregated separately into group probabilities and group utilities, and only then recombined into group expected utilities.

It should be noted that this stance, with its rationale, was forcefully enunciated by Raiffa (1968) 30 years ago in sections 12 and 13 of his classical text, *Decision Analysis*—e.g, on "The Problem of the Panel of Experts" (pp. 232-233):

> "If I were solely responsible as the decision maker, I should want to probe the opinions of my experts to assess my own utility and probability structure. I should try to keep my assessments for utilities separate from my assessments for probabilities, and I should try to exploit such common agreements as independence.[20] Wherever possible, I should want to decompose issues to get at basic sources of agreement and disagreement. I should compromise at the primitive levels of disagreement and adopt points of common agreement as my own, so long as these common agreements were not compensating aggregates of disagreements. I should do so knowing full well that I might end up choosing an action which my experts would say is not as good as an available alternative. Throughout this discussion, of course, I am assuming that I do not have to worry about the viability of my organization, its morale, and so on."

There, too, he reports a result of Zeckhauser's that would be published 11 years later (Hylland and Zeckhauser 1979) in a somewhat different version:

> "Richard Zeckhauser has proved a mathematical theorem that states this result:
> "No matter what procedure you use for combining the utility functions and for combining the probability functions, so long as you keep these separate and do not single out one individual to dictate the group utility and probability assignments, then you can concoct an example in which your experts agree on which act to choose but in which you are led to a different conclusion." (Raiffa 1968, p. 230)

[20] Convex combinations of i.i.d. distributions are not generally i.i.d., so averaging such distributions would not be a way of preserving common agreement on independence. (To preserve independence one could form the average of the individuals' i. i. d. distributions and use that as the 1–shot probability of an i. i. d. group distribution.)

Raiffa (1968, pp. 233-237) explores the tension this theorem reveals between the following two conditions.

> **Reification**: "the group members should consider themselves as constituting a panel of experts who advise the organizational entity: they should imagine the existence of a higher decision–making unit, the organization incarnate, so to speak, and ask what *it* should do. Just as it made sense to give up Pareto optimality in the problem of the panel of experts, it likewise seems to make sense in the group decision problem." (Raiffa, pp. 233-234)
> **Pareto Optimality**: The group prefers one prospect to another if some members do and none have the opposite preference.

We now introduce a new problem for *ex post* aggregation:[21]

> **Flipping:** In *ex post* aggregation, utility and probability profiles for individuals exist relative to which group preference between some pair of options reverses endlessly as the analysis is refined, even though all individual preferences remain unchanged.

Note how this relates to the result of Hylland and Zeckhauser. They use the *ex ante* Pareto condition in an *ex post* framework, i.e., a framework in which probabilities and utilities are aggregated separately. We use the *ex post* Pareto condition (i.e., on utilities, not expected utilities) in an *ex post* framework. Thus fliping is a problem inherent in the *ex post* approach: Unlike the Hylland–Zeckhauser (1979) result, it does not depend on the tension between *ex ante* standards and *ex post* aggregation.[22]

The flipping phenomenon is illustrated by the following example, which we formulate here in the Bolker–Jeffrey framework sketched in sec. 3 above.[23] In the example, initial group desirabilities $12, 0$ of two options (*dessert*, *none*) change to $-8, 0$ upon closer examination of the first option, and change back to $12, 0$ upon still closer examination. The group desirabilities can flip because the individuals have opposed probabilities and differently opposed utilities, somewhat as in the "politics makes strange bedfellows" example, but here with the opposed tendencies overbalancing in opposite directions at each stage of refinement.

[21] Here we illustrate the problem for a particular aggregation rule, i.e., straightforward averaging of probabilities and summing of utilities. But the problem can arise for any *ex post* Pareto optimal aggregation rule.

[22] See Hylland–Zeckhauser (1979, pp. 1325–6). Their axioms 2 and 3 stipulate *ex post* aggregation of individual probabilities p^k and utilities u^k. Their axiom 5 is a weak *ex ante* Pareto optimality condition: "If $E(a_m|p^k, u^k) > E(a_i|p^k, u^k)$ for all k, then a_i is not an element of the choice set."

[23] I.e. the simplest framework for the purpose. A treatment in a modification of the Savage framework will be published elsewhere.

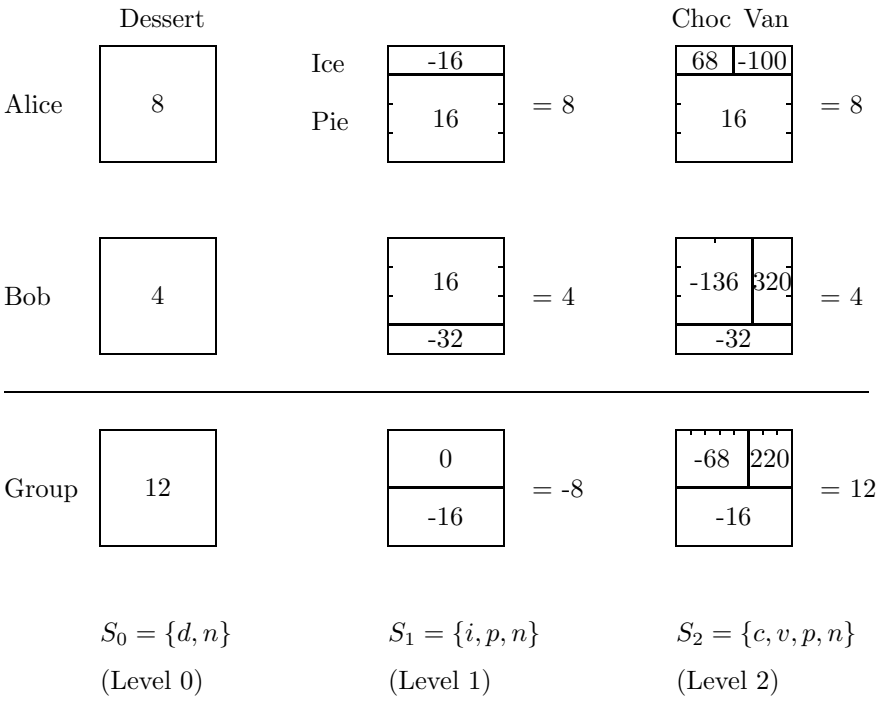


Figure 1: Flipping illustrated by refinements of the "Dessert" option.

Dessert makes strange bedfellows. Alice and Bob are being given a dinner for two in which they must make the same choice from the menu, course by course. Having agreed on all courses so far, they are trying to decide whether or not to have dessert, which the menu lists with no details. Suppose that in fact they both prefer the event *Dessert* to the event *None*, and that on personal desirability scales des_{Alice}, des_{Bob} which they regard as interpersonally commensurate, their desirabilities for *Dessert* are 8 and 4 as shown in Figure 1 (level 0, above the line), and their desirabilities for *None* (not shown in Figure 1) are both 0. Suppose they are sure that dessert will turn out to be *Ice* cream or *Pie*, concerning which their respective commensurate desirabilities are $-16, 16$ for Alice, and $16, -32$ for Bob, as shown above the line at level 1 of Figure 1. Suppose that $P_{Alice}(Ice|Dessert) = 1/4, P_{Alice}(Pie|Dessert) = 3/4$, and that the values for P_{Bob} are just the reverse. These conditional probabilities are represented by the areas of the respective compartments, on a scale where the whole square has area 1. Since $des(A) = E(u|A)$, the desirability of the union of two events that are judged to be incompatible is a weighted average of their

separate desirabilities: *If* $P(A \cap B) = 0$, *then*

$$des(A \cup B) = des(A)P(A|A \cup B) + des(B)P(B|A \cup B) \tag{8}$$

It is easy to verify that with $Dessert = Ice \cup Pie = A \cup B$ this equation, applied to Alice's and Bob's level 1 desirabilities and probabilities, yields their level 0 desirabilities, 8 and 4. And similarly, if both are convinced that Ice would turn out to be $Choc$(olate) or Van(illa), equation (8) delivers their ± 16 level 1 desirabilities for Ice when their probabilities and desirabilities for $Choc$ and Van are as shown at level 2. Then above the line, the three levels of analysis of Alice's attitudes depicted in Figure 1 are mutually consistent, as are the three levels of Bob's.

But *ex post* aggregation of Alice's and Bob's desirabilities by applying the following formula to the numbers shown in Figure 1 yields mutually inconsistent group desirabilities, for the results, shown below the line, exhibit the flipping phenomenon: group desirabilities for *Dessert* flip from 12 to -8 and back again as the aggregation process is applied to finer analyses of the individuals' probabilities and desirabilities.

$$des_{Group}(A) = des_{Alice}(A) + des_{Bob}(A) \tag{9}$$

And it would be straightforward to devise probabilities and utilities for a further stage (say, with $Pie = Apple \cup Banana$) at which group desirability flips back from 12 at stage 2 to -8 at a new stage 3; and one can give an algorithm for continuing the refinements of consistent individual probabilities and utilities so as to carry the $12, -8, 12, -8, \ldots$ flipping process as far as you like—even, endlessly.

The flipping problem has another aspect, i.e., inconsistency of group probabilities and desirabilities with formula (8) when group probabilities conditionally on an act-event D (e.g., the event that we have dessert) are obtained by averaging:

$$P_{Group}(A|D) = \frac{1}{2}P_{Alice}(A|D) + \frac{1}{2}P_{Bob}(A|D) \tag{10}$$

Thus, the desirability of Ice at level 1, obtained via equation (9) as the simple sum of Alice's and Bob's level 1 desirabilities for Ice, is inconsistent with the value obtained via equation (8) as the probability–weighted average of the group's level 2 desirabilities for $Choc$ and Van:

$$des_{Group}(Ice) = -16 + 16 = 0 \text{ from (9)}$$

$$des_{Group}(Ice) = \frac{5}{8}(-68) + \frac{3}{8}(220) = 40 \text{ from (8)}$$

But is formula (9) a correct description of *ex post* aggregation? By definition, *ex post* aggregation adds *utilities*, not desirabilities, so that in genuine *ex post* aggregation formula (9) would be replaced by the corresponding formula for utilities:

$$u_{Group}(s) = u_{Alice}(s) + u_{Bob}(s) \tag{11}$$

Can the effect of applying formula (11) be the same as that of applying formula (9) to the desirabilities of the smallest compartments in Figure 1? The answer is "Yes" if we represent the refinement process as applying primarily to the set S of states of nature, and only derivatively to the events, the subsets of S. Thus, at level 0 there are just two states of nature, the state d in which Alice and Bob have dessert, and the state n in which they have none: at level 0 the set of states of nature is $S_0 = \{d, n\}$ as indicated in Figure 1. The set S_1 of states at level 1 is obtained by replacing d by two states: a state i in which the waiter brings ice cream, and a state p in which he brings pie. And similarly S_2 comes from S_1 by replacing i by c (he brings chocalate ice cream) and v (he brings vanilla).

Here we have three Boolean algebras $\mathcal{A}_k$ of subsets of S_k, with $k = 0, 1, 2$. The algebra $\mathcal{A}_k$ contains $2^{(2^{k+1})}$ events, e.g., $\mathcal{A}_0 = \{\emptyset, \{d\}, \{n\}, S_0\}$. In these, $Dessert$ is represented by three different sets: by $\{d\}$ at level 0, by $\{i, p\}$ at level 1, and by $\{c, v, p\}$ at level 2. We shall say that these three are "associated" with each other, in order to indicate that they are all representations of what is informally seen as one and the same event, $Dessert$. In general, any $A \in \mathcal{A}_k$ for $k = 0, 1$ is associated with an $A' \in \mathcal{A}_{k+1}$ defined as follows, where $\{s\} = S_k - S_{k+1}$ and $\{s', s''\} = S_{k+1} - S_k$:[24]

$$A' = (A - \{s\}) \cup \{s', s''\} \text{ if } s \in A, \text{ else } A' = A \tag{12}$$

As an ideal beyond human powers of attainment, one could think of continuing this process of refinement endlessly, specifying not only the ways in which $Dessert$ and $None$ might turn out, but also possibilities about other things one might care about, e.g., the weather tomorrow (and tomorrow, and tomorrow, ...), various people's states of health, and births, deaths, wars, football scores—whatever. The *ultimate* states of nature are the maximal consistent sets of such specifications. From this idealized point of view the elements of S_k for finite k will be pseudo–states, events (sets of ultimate states) masquerading as states.

Where "s" ranges over ultimate states, aggregation via equations (10) and (11) is immune to the flipping phenomenon illustrated by Figure 1, e.g., because the putative utilities $u_{Alice}(p) = 16$, $u_{Bob}(p) = -32$ at level 1 must really be seen as desirabilities $des_{Alice}(Pie) = 16$, $des_{Bob}(Pie) = -32$ of an event Pie; and formula (11) is no warrant for summing desirabilities. But

[24] I.e., s is the element of S_k that is split into two elements s', s'' to produce S_{k+1}.

application of formula (11) to utilities of ultimate states is beyond human powers: this way out "in principle" leaves *ex post* aggregation impossible in practice. One way or the other, *ex post* aggregation looks like a pipe dream.

If the *ex post* approch is ruled out in this way, the *ex ante* approach has its own severe difficulties. In particular, the *ex ante* possibility theorems rule out any version of liberalism that satisfies the following two conditions. (1) Unanimous individual preferences are preserved as group preferences. (2) Diversity is tolerated as part of political reality, or even cherished, as in Mill's *On Liberty*. (By excluding all linear independence of probability measures and of utility functions, the *ex ante* possibility theorems exclude such diversity.) Liberalism that mets these two conditions violates Bayesian rationality of individuals or the group: it requires irrational people or an irrational society.

In closing we recall that flipping does not arise in Harsanyi's aggregation scheme, for the vN–M or Marschak framework attributes no judgmental probabilities to groups or to individuals.[25] From a certain individualistic point of view this opportunity to deny that groups have beliefs (i.e., judgmental probabilities) is most welcome. On that view we may perhaps speak of groups as agents, and even as having aggregate preferences, but on that view groups are not the sorts of things to which beliefs are to be attributed, and so groups are not to be thought of as rational or irrational.

References

Arnauld, A. (1662), *La logique, ou l'art de penser*. Paris. Translation (1964), *The Art of Thinking*, Indianapolis: Bobbs–Merrill.

Arrow, K. (1950), A Difficulty in the Concept of Social Welfare. *Journal of Political Economy* **58**, 328-346; reprinted in Arrow and Scitovsky

Arrow, K. (1951, 1963), *Social Choice and Individual Values*. New York: Wiley.

Arrow, K. and Scitovsky, T., eds. (1969), *Readings in Welfare Economics*. Allen and Unwin, London.

Bergson, A. (1938), A Reformulation of Certain Aspects of Welfare Economics. *Quarterly Journal of Economics* **52**, 310-334; reprinted in Arrow and Scitovsky (1969).

Bergson, A. (1948), Socialist Economics. *A Survey of Contemporary Welfare*

[25] It does arise in other schemes that are neither *ex ante* nor *ex post,* e.g., that of Levi (1997, chapter 9), and the pseudo *ex post* scheme illustrated in Figure 1 above.

Economics, H. S. Ellis (ed.), pp. 412-448: Blakiston, Philadelphia.

Bolker, E. (1965), *Functions Resembling Quotients of Measures*. Ph.D. Dissertation, Harvard University.

Bolker, E. (1966), Functions Resembling Quotients of Measures. *Transactions of the American Mathematical Society* **124**, 293-312.

Bolker, E. (1967), A Simultaneous Axiomatization of Subjective Probability and Utility. *Philosophy of Science* **34**, 333-340.

Bradley, R. (1997). *The Representation of Beliefs and Desires within Decision Theory*. Ph. D. Dissertation, University of Chicago.

Broome, J. (1987), Utilitarianism and Expected Utility. *Journal of Philosophy* **84**, 402–422.

Broome, J. (1990), Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism. *Review of Economic Studies* **57**, 477-503

Broome, J. (1991), *Weighing Goods*. Oxford: Basil Blackwell

Carnap, R. (1945), On Inductive Logic. *Philosophy of Science* **12**, 72-97.

Carnap, R. (1950, 1962), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Harsanyi, J. (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy* 63: 309-321; reprinted in Arrow and Scitovsky (1969)

Harsanyi, J. (1990), Interpersonal Utility Comparisons. *Utility and Probability*, J. Eatwell, M. Milgate, and P. Newman (eds.), New York: Norton

Hylland, A. and R. Zeckhauser (1979), The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values. *Econometrica* 47: 1321-1336

Jeffrey, R. (1965, 1983), *The Logic of Decision*. 1st ed., New York: McGraw Hill. 2nd ed., Chicago: University of Chicago Press

Jeffrey, R. (1992), *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Jeffrey, R. (1996), Decision Kinematics. *The Rational Foundations of Eco-*

nomic Behaviour, K. J. Arrow, E. Colombatto, and M. Perlman (eds.), Macmillan (GB) and St. Martin's (USA).

Joyce, J. M. (1992), *The Foundations of Causal Decision Theory*. Ph. D. dissertation, University of Michigan, Ann Arbor.

Kolmogorov, A. N. (1948, 1995), Algèbres de Boole métriques complètes, *IV Zjadz Matemayików Poslkich*, Warsaw (1948) pp. 21-30. Translation: Complete Metric Boolean Algebras, *Philosophical Studies* **77** (1995) 57-66.

Levi, I. (1997), *The Covenant of Reason*. Cambridge: Cambridge University Press.

Marschak, J. (1950), Rational Behavior, Uncertain Prospects, and Measurable Utility. *Econometrica* **18**, 111-141

Mongin, P. (1995), Consistent Bayesian Aggregation. *Journal of Economic Theory* **66**, 313-351

Raiffa, Howard (1968), *Decision Analysis*. Reading, Mass.: Addison–Wesley

Ramsey, F. P. (1931), Truth and probability, in *The Foundations of Mathematics and other Logical Essays*, R. B. Braithwaite (ed.), Kegan Paul. Reprinted in Ramsey (1990).

Ramsey, F. P. (1990), *Philosophical Papers*, D. H. Mellor (ed.), Cambridge University Press.

Savage, L. J. (1954), *Foundations of Statistics*. New York: Wiley

Seidenfeld, T., Kadane, J. B., and Schervish, M. J. (1989), On the Shared Preferences of Two Bayesian Decision Makers. *Journal of Philosophy* **86**, 225-244.

Sen, A. (1970), *Collective Choice and Social Welfare*. San Francisco: Holden–Day.

von Neumann, J. and O. Morgenstern (1944, 1947, 1953), *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Weymark, J. A. (1991), A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism, in *Interpersonal Comparisons of Well–Being*, J. Elster and J. Roemer (eds.), Cambridge: Cambridge University Press.