

## **Interpersonal comparisons of utility and the policy paralysis problem**

**Michael Mandler**

Department of Economics, Harvard University, Cambridge, MA 02138 USA  
(e-mail: mmandler@harvard.edu)

Received: 26 March 1997/Accepted: 16 November 1998

**Abstract.** Several “Paretian” welfare rules are equivalent when policymakers know agents’ characteristics, e.g., a policy is optimal if (a) any other policy making someone better off harms some agent, or (b) it is the maximum of some social welfare function. This paper extends these and other rules to environments where policymakers have a probability distribution over a state space of possible models. Under weak conditions, rule (a), which postulates ex ante preferences for agents, recommends some change from almost every status quo policy. Unfortunately, (a) requires a demanding form of interpersonal welfare comparability. Rule (b) labels all policies optimal if the state space obeys a weak diversity condition. Since the probabilities of states are irrelevant for this result, only a small perturbation of a model with no uncertainty generates policy paralysis.

### **1 Introduction**

Although it is a commonplace that Paretian welfare economics cannot dispense practical policy advice, most theoretical work continues to put policy proposals to a Pareto optimality test. The Pareto approach’s popularity is due to the fact that it offers unambiguous policy recommendations that appear to avoid interpersonal comparisons of welfare. The welfare theorems of general equilibrium theory provide the paradigmatic illustration: the first welfare theorem guarantees that an economy with distortions can achieve Pareto optimality if distortions are eliminated, while the second welfare theorem shows

---

I thank Bob Pollak for an early, helpful discussion. A referee suggested several important changes.

that with appropriate redistributions the optimum can be Pareto-improving relative to the status quo. The view that the Pareto approach is unworkable in practice is evidently at odds with the clear-cut recommendations of the welfare theorems. This paper provides a theoretical framework with which to evaluate this discrepancy. Our aim is to pinpoint where the difficulty with the Pareto approach lies and under what assumptions it can be applied consistently.

A principal drawback of the welfare theorems is that optimal policies vary as a function of the attributes of the model; optimal policies therefore require knowledge of the details of the underlying economy. Are strong policy recommendations still available when a policymaker's ignorance is recognized explicitly? As we will see, it *is* possible in an uncertain environment to define a Pareto ordering that ranks a sufficient number of policies. But the orderings require a system for making interpersonal comparisons of utility. When interpersonal comparisons are systematically eliminated, every policy can be optimal.

The complaint that Paretianism is impractical has a long history, though it has usually been advanced as an applied rather than a theoretical criticism. Many have observed, for instance, that political or institutional considerations restrict the set of available policies; policies instituting Pareto improvements or Pareto-optimal allocations might therefore be infeasible. A second, more theoretical, objection to the Pareto approach, stemming from the theorem of the second best (Lipsey and Lancaster 1956), holds that Pareto-optimal policies can be complicated and counterintuitive, with policy parameters in one area of economic policy affected by all of a model's parameters. Constructing a Pareto-optimal policy therefore seems to require unobtainable information.<sup>1</sup> We examine only this latter complaint, and consider only the pure effect of informational limitations without introducing any exogenous constraints on the government's policy options.

There is an obvious response to the limited-information criticism. Even if every policy choice will with high probability not be Pareto optimal *ex post*, the appropriate standard in the presence of uncertainty is whether or not a policy is optimal *ex ante*. But, as we will see, with an *ex ante* optimality concept, the opposite danger emerges: too many rather than too few policies may be optimal. We therefore gauge the success of an *ex ante* welfare rule by the extent to which it can label sufficiently many policies as suboptimal. If a welfare rule designates all policies or almost every policy as optimal, we say there is *policy paralysis*. Policies will be elements of Euclidean space; "almost every" therefore refers to a set whose complement has measure zero.

Before considering Pareto orderings under uncertainty, recall the two

---

<sup>1</sup> Some of the initial criticisms of the new welfare economics faulted the Pareto criterion for not specifying which Pareto-optimal allocation should be selected (see, e.g., Samuelson 1947, pp. 243–244, or Arrow 1951, p. 37). With time, this charge has seemed less compelling. Indeed the second objection above makes the opposite charge: the Pareto approach offers too few, not too many policy options.

standard methods for avoiding interpersonal comparisons of utility in economic analysis. The first is the standard definition of Pareto optimality: a policy is optimal if there is no alternative making some agents better off and harming no agent. This definition relies on the appeal of policies that agents unanimously approve. The second method, dating to Bergson (1938) and Samuelson (1947), advocates the maximization of social welfare but singles out those optimality conditions which are independent of the weights used to aggregate individual utility functions. These conditions, such as the requirement that marginal rates of substitution be equal, are then identified as free from interpersonal comparisons of utility. Since the two methods often identify the same policies as optimal, it has been common to think of them as different characterizations of a single notion of optimality.

Suppose now that the policymaker faces a probability distribution over a state space of possible models. It will be important that it is the policymaker who is uncertain, not the agents. Some features of the world are unknown to both agents and the policymaker, but other facts, for instance the precise characteristics of individual preferences, while collectively known to the agents, are unknown to the policymaker. We call the latter type of ignorance *policy-making uncertainty*.

How should optimality be defined? First consider (partial) orderings on the set of alternatives. The standard Pareto criterion translates into what we here call the *agent-based* ordering: policy or allocation  $a$  is superior to  $b$  if every agent in expectation weakly prefers  $a$  to  $b$  and at least one agent in expectation strictly prefers  $a$  to  $b$ . The agent-based ordering recommends only those policy changes that no agent would reject *ex ante*. But the agent-based approach suffers from a difficulty. With policy-making uncertainty, there is no *ex ante* vantage point at which agents form preferences or at which their choices can be observed. It is the policymaker, therefore, who must aggregate or “weight” each agent’s potential utility functions. Just as with weights on individual utilities in standard social welfare functions, however, weights on potential utility functions imply a system of interpersonal welfare comparisons. Hence, although the agent-based ordering duplicates the standard Pareto criterion’s reliance on agent unanimity as a policy guide, it does not preserve neutrality with respect to inter-preference comparisons.

A second ordering, in the spirit of the social welfare maximization approach and which we call *utility-independent*, defines  $a$  to be superior to  $b$  if for all utility representations of the different potential preferences (or all systems of weights for different potential utility functions), the resulting sum-of-expected-utilities (or weighted sum-of-expected-utilities) social welfare function weakly recommends  $a$  over  $b$ , and for one set of utility representations  $a$  is strictly recommended over  $b$ . The utility-independent ordering has a distinct argument in its favor: since it is not tied to any method of weighting potential utility functions, it avoids inter-preference comparisons of welfare.

Another category of welfare rule recommends optimal policies, whether or not those policies are ranked relative to the status quo. The above two orderings automatically generate optimality definitions by letting undominated

policies be optimal. In addition, say that a policy is *maximization* optimal if it maximizes a social welfare function corresponding to some choice of utility representations. The maximization of social welfare functions defined in this way is routine in the economics of welfare under uncertainty. See, e.g., Hammond (1983) (where ex post welfare functions correspond to our social welfare functions) and the references provided there.

The question is most important in policy paralysis traditional economic environments, where variables such as commodity taxes and redistributions of wealth are the policy instruments. It is after all the welfare theorems of general equilibrium theory that originally held out the promise that the Pareto approach could neutrally discriminate among standard policy instruments. For reasons we outline in the concluding discussion, however, a formal analysis of traditional economic environments (studied in Mandler 1996) is technically difficult. To keep matters simple, this paper uses a standard social choice setting in which a finite number of options and their probability mixtures are the policy choices; for economic applications, one may think of the pure policy choices as public goods projects. More broadly, the current set of policy choices should be seen as a surrogate for a general Euclidean set of policies. As we will see, a probabilistic setting has the advantage of restricting the class of admissible social welfare functions, thus strengthening our policy paralysis result.

Our first results, in Sect. 2, simply summarize the equivalence of the different descriptions of the Pareto approach in the absence of policy-making uncertainty. The equivalence of the two welfare orderings is a standard result in social choice theory (see, e.g., Sen (1970) and Theorem 1 below). In fact, the equivalence still holds if we additionally require that individuals with identical preferences are assigned the same utility representation and weight in the social welfare function (Theorem 2). This condition requires that economically indistinguishable agents are treated interchangeably and is natural when an agent's index is only a labeling device.<sup>2</sup> Due to the fact that we allow mixtures of pure policies (thus convexifying the set of achievable utilities) and have only a finite number of pure policies, the three optimality concepts also coincide, even with the restriction on utility representations just mentioned (Theorems 1 and 2). As noted earlier, it may be due to such results that it is common to think that the standard Pareto criterion can rely on either a unanimity or a no-interpersonal-comparisons-of-utility defense. Concerning policy paralysis, we show in the certainty case that if the number of pure policy options,  $n$ , is larger than the number of agents,  $m$ , then under a mild regularity condition any policy that places positive probability on each pure option can always be dominated (Theorem 3).

---

<sup>2</sup> In addition, some (e.g., Arrow 1951, p. 31) have argued as a matter of principle that the utility representations used in welfare functions should depend only on ordinal preferences.

Section 3 lays out the model with policy-making uncertainty and demonstrates that the agent-based ordering remains well-behaved. Additionally, the recommendations of the agent-based ordering that are made when there is no policy-making uncertainty are robust to the introduction of a small amount of uncertainty. We also investigate the extent of the interpersonal comparisons needed by the agent-based approach. We find that even mild and plausible restrictions on the degree of interpersonal comparability are inconsistent with the agent-based assumption that each agent has a complete *ex ante* ordering.

Section 4 defines the utility-independent and maximization optimality concepts under policy-making uncertainty. We show that even if identical potential agents are represented by the same utility function, any policy is optimal as long as we are free to choose  $n$  of agents in the support of the distribution of preferences (Theorem 6). The choices of the  $n$  agents can be made robustly (they can be arbitrary elements of open sets) and the choices can be made independently of the other preferences in the support. Furthermore the probabilities of the states in which the  $n$  agents appear, although positive, can be arbitrarily small. Thus, beginning with no uncertainty, any model can be perturbed robustly, by adding states with arbitrarily small probabilities, in such a way that any policy is the maximum of some social welfare function or, equivalently, is optimal in the utility-independent sense.

Our basic conclusion, therefore, is that in order to make a Pareto-like welfare rule workable in an uncertain environment, extensive interpersonal comparisons of welfare are necessary. The viability of the agent-based ordering in a sense validates common practice; most models (outside of the mechanism design and implementation literatures) begin by presupposing complete preferences for agents over policy choices. When the policymaker and the agents have the same information, there is nothing problematic in this modeling strategy; but with policy-making uncertainty, the use of *ex ante* preference orderings requires detailed welfare and utility comparisons.

We have mentioned two restrictions on how social welfare functions are constructed: only additive social welfare functions are considered and identical agents are represented by the same utility function. The first requirement is justified by the Harsanyi (1955) theorem stating that when agents have von Neumann-Morgenstern preferences social welfare orderings can be represented by functions that are additively separable in agent utilities. The second requirement is appropriate when an agent's index communicates no substantive information to the policymaker. More importantly, restrictions on the set of social welfare functions *strengthen* the policy paralysis conclusion: if the policy paralysis obtains with a restricted class of functions, it also obtains without the restrictions. Our aim, therefore, is to impose as many plausible restrictions as possible.

We have described Sects. 2 through 4. The concluding discussion in Sect. 5 considers the implications of the policy paralysis result, the mechanism design literature, and the pertinent differences between economic and social choice environments. We also sketch a unifying mathematical framework that encompasses our various policy criteria.

## 2 Policy analysis with no policy-making uncertainty

Let  $A = \{a_1, \dots, a_n\}$  denote a finite set of pure policy alternatives and  $Q = \Delta(A) = \{q = (q_1, \dots, q_n) \in \mathbf{R}_+^n : \sum_{i=1}^n q_i = 1\}$  denote the corresponding set of mixed policy choices. A policy is *strictly mixed* if  $q \gg 0$ . Each individual  $j \in J = \{1, \dots, m\}$  is described by a preference relation over  $Q$  denoted  $R_j$ . An individual  $j$ 's strict preference and indifference relations are defined in the standard way and are denoted  $P_j$  and  $I_j$ . We assume, for all  $j \in J$ , that  $R_j$  can be represented by a von Neumann-Morgenstern utility function. That is, there is a nonempty set of functions,  $U_j$ , such that  $u_j : A \rightarrow \mathbf{R}$  is an element of  $U_j$  if and only if  $\sum_{i=1}^n q_i u_j(a_i) \geq \sum_{i=1}^n q'_i u_j(a_i) \Leftrightarrow q R_j q'$ . To economize on notation, let  $E(q, u_j)$  signify  $\sum_{i=1}^n q_i u_j(a_i)$ .

An *unrestricted assignment* of utilities is a vector of functions  $u = (u_1, \dots, u_m)$  where for each  $j \in J$ ,  $u_j \in U_j$ . Corresponding to each assignment is a (additive) social welfare function from  $Q$  to  $\mathbf{R}$  defined by  $E(q, \sum_{j \in J} u_j)$ . As mentioned in the introduction, Harsanyi (1955) provides compelling axioms that imply that social preference relations can be represented by functions that are additive in agent utilities. The key assumption in Harsanyi is that social preference relations obey the von Neumann-Morgenstern axioms; since our social choices are lotteries, the von Neumann-Morgenstern axioms are applicable and plausible. Of course, it is the social preference relation underlying  $E(\cdot, \sum_{j \in J} u_j)$  that matters; we could use any monotonic transformation of  $E(\cdot, \sum_{j \in J} u_j)$  instead. Note that since multiplying each of the  $m$  utility functions in an assignment by an arbitrary positive number yields another assignment, the set of all assignments generates not only the rankings of  $Q$  representable by unweighted sums of utilities but also the rankings representable by weighted sums.

There are several possible ways of defining a Pareto-based welfare rule. A first category employs (partial) binary orderings over policies.

**Definition 1.** *In a model with no policy-making uncertainty, policy  $q$  is superior to  $q'$  by the agent-based ordering if, for all agents  $j$ ,  $q R_j q'$ , and if, for some  $j$ ,  $q P_j q'$ . Policy  $q$  is superior to  $q'$  by the utility-independent ordering if for all possible unrestricted assignments of utilities,  $E(q, \sum_{j \in J} u_j) \geq E(q', \sum_{j \in J} u_j)$ , and for some assignment,  $E(q, \sum_{j \in J} u_j) > E(q', \sum_{j \in J} u_j)$ .*

Transparently, if  $q$  is superior to  $q'$  by the agent-based ordering,  $q$  is also superior to  $q'$  by the utility-independent ordering. To argue in the other direction, suppose  $q$  is superior to  $q'$  by the utility-independent ordering but not by the agent-based ordering. Then either  $q' I_j q$  for all agents  $j$  or there is some agent  $h$  with  $q' P_h q$ . In the former case, for any assignment of utilities,  $E(q', \sum_{j \in J} u_j) = E(q, \sum_{j \in J} u_j)$ , contradicting the superiority of  $q$  by the utility-independent ordering. In the latter case, pick an arbitrary assignment of utilities,  $u$ , and note that  $E(q', u_h) > E(q, u_h)$ . By fixing  $\alpha_h$  and setting  $\alpha_j$ , for  $j \neq h$ , near 0, there is a  $\alpha \gg 0$  such that  $E(q', \alpha \cdot u) > E(q, \alpha \cdot u)$ . Using the assignment  $(\alpha_1 u_1, \dots, \alpha_m u_m)$ , we have a contradiction of the superiority of  $q$  by the utility-independent ordering. Thus, in the current context, the standard

Paretian requirement that a policy change make no agent worse off is identical to requiring that policy changes be social welfare-improving for all possible interpersonally comparable utility functions.

A second type of welfare rule identifies a set of optimal policies without stipulating any ranking of suboptimal policies. Two definitions of optimality are readily derived from the above two orderings. A policy  $q$  is *agent-based* (resp. *utility-independent*) *optimal* if there is no  $q' \in Q$  superior to  $q$  by the agent-based (resp. utility-independent) ordering. Given the above comments, these concepts clearly coincide. A third definition is also common; say that  $q$  is *maximization optimal* if for some assignment of utilities  $u$  and all  $q' \in Q$ ,  $E(q, \sum_{j \in J} u_j) \geq E(q', \sum_{j \in J} u_j)$ . It is immediate that if  $q$  is maximization optimal then  $q$  is optimal by the other two definitions. The reverse implication also holds in this model, as stated in the following theorem summarizing the results so far.

**Theorem 1.** *In the absence of policy-making uncertainty, the agent-based and utility-independent orderings are equivalent, and the agent-based, utility-independent and maximization definitions of optimality are equivalent.*

We omit the remainder of the proof (showing that if  $q$  is agent-based optimal then  $q$  is maximization optimal) which is largely routine. But a word of warning: the equivalence of the three optimality definitions need not hold when  $Q$  is not generated from a finite set of pure policies. The social welfare function maximized at an agent-based optimum could require some subset of individual utility functions to be multiplied by a zero weight; these utility functions would not constitute an assignment of utilities in our sense.

So far we have allowed individuals with identical preferences to be represented by distinct utility functions. This practice violates the spirit of the utilitarian/social-welfare-maximization tradition when individuals, in addition to having the same preferences, are indistinguishable in all respects. In the Harsanyi (1953) model, for example, the social welfare function is derived from the hypothetical choice situation of having to pick an  $a_i \in A$  while imagining that one could be any of the  $m$  agents, each with probability  $1/m$ . If an agent's index communicates nothing about his or her well-being, then a Harsanyi decision maker would presumably be indifferent between being agent  $j$  with some preference relation and being  $h \neq j$  with the same preference relation. We therefore require that agents with the same preferences be assigned the same utility representation. Define a *restricted assignment* of utilities to be a vector of functions  $u = (u_1, \dots, u_m)$  where (a) for each  $j \in J$ ,  $u_j \in U_j$ , and (b) for all  $j, h \in J$ ,  $R_j = R_h$  implies  $u_j = u_h$ . A *strong utility-independent ordering* (and an associated definition of optimality) and *strong maximization optimality* are defined relative to the set of restricted assignments. Let the modifier *weak* refer to welfare rules that arise from unrestricted assignments. The following theorem indicates that the strong welfare concepts coincide with the weak concepts; as a consequence, *Pareto optimality* has an unambiguous meaning when there is no policy-making uncertainty.

**Theorem 2.** *The strong and weak utility-independent orderings are equivalent and the strong and weak maximization definitions of optimality are equivalent.*

*Proof.* The equivalence of the two orderings and that strong optimality implies weak optimality are either immediate or slight variants of above arguments. Suppose that  $q'$  is optimal by the weak maximization definition; let  $u$  be one of the assignments for which  $E(\cdot, \sum_{j \in J} u_j)$  reaches a maximum at  $q'$ . For each agent  $j$  let  $J(j) \subset J$  be defined by the condition  $h \in J(j)$  if and only if  $R_h = R_j$ . For all  $j$ , let  $\bar{u}_j = (1/\#J(j)) \sum_{l \in J(j)} u_l$ . Since for all  $q$ ,  $E(q, \sum_{j \in J} \bar{u}_j) = E(q, \sum_{j \in J} u_j)$ ,  $q'$  must be optimal by the strong maximization definition using the restricted assignment  $(\bar{u}_1, \dots, \bar{u}_m)$ . ■

We turn to policy discrimination. Let us say that a model and welfare criterion exhibit *policy paralysis* if all or almost every policy is optimal. Without a sufficiently diverse set of policies, even the most discriminating welfare ordering can classify a large proportion of policies as optimal. To construct a precise definition of diversity, pick an arbitrary assignment of utilities  $u$  and consider the set of feasible utilities,  $F_u = \text{co}\{w \in \mathbf{R}^m : w = (u_1(a_i), \dots, u_m(a_i)), i = 1, \dots, n\}$ , where  $\text{co } S$  indicates the convex hull of a set  $S$ .

**Definition 2.** *The diversity condition is satisfied if, for some assignment  $u$ , the set  $F_u$  has a nonempty interior relative to  $\mathbf{R}^m$ .*

Note that if for some assignment  $u$ ,  $F_u$  has nonempty interior then for any other assignment  $u'$ ,  $F_{u'}$  also has nonempty interior. The diversity condition implies that  $n > m$ . Furthermore, if the vectors,  $(u_1(a_i), \dots, u_m(a_i))$ ,  $i = 1, \dots, n$ , do not lie in the same  $(m - 1)$ -dimensional affine subspace of  $\mathbf{R}^m$ , then the diversity condition holds. Consequently, if  $n > m$ , then for almost every specification of agent utility functions (each seen as an element of  $\mathbf{R}^n$ ), the diversity condition will be satisfied.

How likely is it that  $n > m$ ? Although economic environments differ in important respects from social choice frameworks (see Sect. 5), the discrepancies do not arise if the policymaker can directly choose allocations of commodities and no two agents have the same preference relation. In an exchange economy with  $\ell$  commodities, the pertinent set of policies is the frontier of the set of allocations. This set has dimension  $\ell(m - 1)$ , which is larger than  $m$  under the mild restriction that  $m \geq 2$ ,  $\ell \geq 2$ , and  $\ell + m \geq 5$ .

The following theorem shows that policy paralysis does not occur when the diversity condition is satisfied and there is no policy-making uncertainty. In fact, the Pareto criterion shows a substantial capacity to discriminate among policies.

**Theorem 3.** *If the diversity condition is satisfied and there is no policy-making uncertainty, almost every policy is Pareto suboptimal and there is an optimal policy that is Pareto superior to any suboptimal policy.*

*Proof.* We show that if  $q$  is strictly mixed, then  $q$  is suboptimal. For any policy  $q$ ,  $(E(q, u_1), \dots, E(q, u_m)) \in F_u$ . Furthermore, if  $\bar{q}$  is strictly mixed, it is easy to show, given the diversity condition, that  $(E(\bar{q}, u_1), \dots, E(\bar{q}, u_m)) \in \text{int } F_u$ . Consequently, for  $\varepsilon > 0$  sufficiently small,  $(E(\bar{q}, u_1) + \varepsilon, \dots, E(\bar{q}, u_m) + \varepsilon) \in F_u$ . Since for each  $w \in F_u$ , there exists a  $q$  such that  $w = (E(q, u_1), \dots, E(q, u_m))$ , there is a policy  $q'$  such that  $(E(q', u_1), \dots, E(q', u_m)) = (E(\bar{q}, u_1) + \varepsilon, \dots, E(\bar{q}, u_m) + \varepsilon)$ . Any strictly mixed  $\bar{q}$  is therefore suboptimal. To find an optimal policy that is Pareto superior to an arbitrary suboptimal policy  $q^*$ , maximize the continuous function  $E(q, \sum_{j \in J} u_j)$  over the compact subset of  $Q$  defined by the constraints  $E(q, u_j) \geq E(q^*, u_j)$ ,  $j \in J$ . Any  $q$  in the nonempty set of solutions to this problem is optimal and Pareto superior to  $q^*$ . ■

Note that the proof shows that for each of the suboptimal policies there is an optimal policy leaving each agent strictly better off.

The diversity condition is not necessary for the conclusion that every mixed policy is suboptimal. To take the simplest example, suppose that  $n \geq 2$  and that each agent strictly prefers pure policy  $a_1$  to  $a_2$ . Any policy with  $q_2 > 0$  (and hence any mixed policy) is dominated by a policy that shifts some probability weight from  $a_2$  to  $a_1$ . Yet, if  $n \leq m$ , the diversity condition cannot be satisfied. On other hand, as we will see in Sect. 4, when  $n \leq m$  there are robust cases where every policy is optimal.

Summing up: if  $n > m$ , policy paralysis generically obtains, while if  $n \leq m$ , both policy paralysis and effective policy discrimination can robustly occur.

### 3 Policy-making uncertainty and agent-based welfare rules

We now reexamine the ability of the Pareto criterion to rank policies when the policymaker is uncertain about the underlying model. Note that by thinking of a policy  $a_i$  as a lottery over more primitive prizes, Sect. 2 can permit the policymaker and agents to face uncertainty (above and beyond the uncertainty that stems from mixing policies). But in Sect. 2 the policymaker must be at least as well informed as any of the agents. We now allow the policymaker to face an additional uncertainty that the agents collectively do not experience.

We preserve from Sect. 2 a fixed set of pure policies  $A$ , the set of mixed policies  $Q = \Delta(A)$ , and the number of agents  $m$ . Now, however, the policymaker has a finite state space,  $\Omega = \{\omega_1, \dots, \omega_s\}$ . Each state specifies von Neumann-Morgenstern preferences over  $Q$  for each of the  $m$  agents. Denote the preferences of agent  $j$  in state  $\omega_k$  by  $R_j(\omega_k)$ . When convenient, we refer to the  $R_j(\omega_k)$  as  $j$ 's ex post preference relations. We assume that  $R_j(\omega_k)$  obeys the assumptions imposed on preferences in Sect. 2. The policymaker also begins with a list of probabilities for the  $s$  states, denoted by  $\bar{\pi} = (\bar{\pi}_1, \dots, \bar{\pi}_s) \gg 0$  with  $\sum_{k=1}^s \bar{\pi}_k = 1$ . Using familiar arguments, this description can, in addition to simple uncertainty about the preferences of agents, also incorporate uncertainty about how actions translate into outcomes.

### 3.1 The agent-based approach

In the presence of policy-making uncertainty, the agent-based approach supposes that each agent  $j \in J$  faces uncertainty over  $\Omega$  as well as  $A$ . Furthermore, each agent is assumed to possess well-behaved – and in particular complete – preferences under this hypothetical veil of ignorance. The typical ex ante object of choice is  $(q, \pi)$ , a lottery over policies paired with a lottery over arrays of agents' ex post preferences. Let  $R_j^*$  denote  $j$ 's complete ex ante preferences over the set of such lottery pairs, and let  $P_j^*$  and  $I_j^*$  be the derived strict preference and indifference relations. We use  $(a_i, \omega_k)R_j^*(a_{i'}, \omega_{k'})$  to mean  $(q, \pi)R_j^*(q', \pi')$ , where  $(q, \pi)$  assigns probability 1 to  $a_i$  and  $\omega_k$  and  $(q', \pi')$  assigns probability 1 to  $a_{i'}$  and  $\omega_{k'}$ . Similarly, let  $a_iR_j(\omega_k)a_{i'}$  mean  $qR_j(\omega_k)q'$ , where  $q$  assigns probability 1 to  $a_i$  and  $q'$  assigns probability 1 to  $a_{i'}$ .

We assume that each  $R_j^*$  obeys the appropriate expected utility property. If  $v$  is a function from  $A \times \Omega$  to  $\mathbf{R}$ , let  $V(q, \pi, v)$  denote  $\sum_{i=1}^n \sum_{k=1}^s q_i \pi_k v(a_i, \omega_k)$ .

*Assumption 1.* For each  $j$ , there is a nonempty set of functions  $U_j(R_j^*)$  such that  $v_j^* : A \times \Omega \rightarrow \mathbf{R}$  is an element of  $U_j(R_j^*)$  if and only if  $(q, \pi)R_j^*(q', \pi') \Leftrightarrow V(q, \pi, v_j^*) \geq V(q', \pi', v_j^*)$ .

We are assuming, only for simplicity, that the randomization over  $a_i$  is independent of  $\omega_k$ . Given that “nature” chooses  $\omega_k$  while the policymaker randomizes among the  $a_i$ , this restriction is appropriate. A more general model would begin with arbitrary probability measures on  $A \times \Omega$ , which need not equal  $q_i \pi_k$  at each  $(a_i, \omega_k) \in A \times \Omega$ .

Although it does not directly affect the policy paralysis issue, the following consistency condition on the  $R_j^*$  and  $R_j(\omega_k)$  is standard.

*Assumption 2.* For all  $j, \omega_k$ , and  $(a_i, a_{i'})$ ,  $a_iR_j(\omega_k)a_{i'} \Rightarrow (a_i, \omega_k)R_j^*(a_{i'}, \omega_k)$ .

Harsanyi (1977) calls this assumption the *principle of acceptance*; it states that if  $j$  knows ex ante that his preferences will be a particular  $R_j$  then  $j$  will choose ex ante among policies according to  $R_j$ . Assumptions 1 and 2 imply, for all  $\omega_k \in \Omega$  and  $v_j^* \in U_j(R_j^*)$ , that  $v_j^*(\cdot, \omega_k)$  is an expected utility representation of the preferences  $R_j(\omega_k)$ .<sup>3</sup>

A policymaker employing the agent-based approach uses an ex ante preference relation  $R_j^*$  satisfying Assumptions 1 and 2 and the probabilities  $\bar{\pi}$  to assess  $j$ 's welfare. The use of  $\bar{\pi}$  is without loss of generality since probabilities linearly transform state-dependent utilities. A policymaker can therefore believe that each  $j$  would hypothetically assign a distinct vector of subjective probabilities  $\pi^j \gg 0$  to  $\Omega$ . More precisely, suppose the policymaker attributes to  $j$  the ex ante preference relation  $R_j^*$  and the probabilities  $\pi^j$ , and let

<sup>3</sup> There are other plausible conditions on the relationship between  $R_j^*$  and the  $R_j(\omega_k)$ . Given our discussion of how welfare functions should weight agents with identical preferences, it is natural to assume, for all  $j, a_j, \omega_k$ , and  $\omega_{k'}$ , that  $R_j(\omega_k) = R_j(\omega_{k'}) \Rightarrow (a_j, \omega_k)I_j^*(a_i, \omega_{k'})$ .

$V(q, \pi^j, v_j^*) = \sum_{i=1}^n \sum_{k=1}^s q_i \pi_k^j v_j^*(a_i, \omega_k)$  be an ex ante utility function for  $j$ . Since  $\sum_{i=1}^n \sum_{k=1}^s q_i \pi_k^j v_j^*(a_i, \omega_k) = \sum_{i=1}^n \sum_{k=1}^s q_i \bar{\pi}_k (\pi_k^j / \bar{\pi}_k) v_j^*(a_i, \omega_k)$  and, for each  $\omega_k$ ,  $(\pi_k^j / \bar{\pi}_k) v_j^*(a_i, \omega_k)$  represents the preferences  $R_j(\omega_k)$ , there exists an ex ante preference relation  $\bar{R}_j^*$  satisfying Assumptions 1 and 2 that, when combined with  $\bar{\pi}$ , provides the same ranking over policies as  $R_j^*$  combined with  $\pi^j$ .<sup>4</sup>

If the policymaker is willing to posit ex ante preferences, the agent-based ordering of the previous section can be extended to the current environment.

**Definition 3.** Policy  $q$  is superior to  $q'$  by the agent-based ordering in a model with policy-making uncertainty if  $(q, \bar{\pi})R_j^*(q', \bar{\pi})$  for all agents  $j$  and if, for some  $h$ ,  $(q, \bar{\pi})P_h^*(q', \bar{\pi})$ .

Correspondingly, policy  $q$  is agent-based optimal if there does not exist a  $q'$  superior to  $q$  by the agent-based ordering. Setting  $s = 1$ , it is clear that these concepts generalize the agent-based welfare rules of Sect. 2.

Using these definitions, Theorem 3, appropriately interpreted, applies. For each agent  $j$ , choose any  $v_j^* \in U_j(R_j^*)$  and define  $F_{v^*} = \text{co}\{w \in \mathbf{R}^m : w = (\sum_{k=1}^s \bar{\pi}_k v_1^*(a_i, \omega_k), \dots, \sum_{k=1}^s \bar{\pi}_k v_m^*(a_i, \omega_k)), i = 1, \dots, n\}$ . If  $F_{v^*}$  satisfies the diversity condition (has nonempty interior) and  $\bar{q}$  is strictly mixed, there is an agent-based optimal policy that improves on  $\bar{q}$ . As in Sect. 2, the diversity condition will typically be satisfied when  $n > m$ .

The agent-based approach has the advantage that policy changes recommended when the model is known with certainty will survive the addition of a small amount of uncertainty. Formally, if  $(R_1, \dots, R_m)$  is a model with no policy-making uncertainty, let a *perturbation* of the model be a  $(R_1^*, \dots, R_m^*, \Omega, \bar{\pi})$ , where  $R_j(\omega_1) = R_j$  for each  $j$  and where  $\bar{\pi}_1$  is near 1. Suppose the certainty model begins with a status quo policy  $\bar{q}$  such that there exists a  $q'$  with  $q' P_j \bar{q}$  for all  $j$  (as, for example, when  $\bar{q}$  is strictly mixed and  $F_u$  has nonempty interior). Now perturb the model to  $(R_1^*, \dots, R_m^*, \Omega, \bar{\pi})$ . It is then immediate that there is some  $\varepsilon > 0$  such that if  $\bar{\pi}_1 > 1 - \varepsilon$  then  $q'$  is also superior to  $\bar{q}$  by the agent-based definition in the perturbed model. In other words, if we adopt the agent-based approach – and thus impose the ex ante preferences  $(R_1^*, \dots, R_m^*)$  – then a sufficiently small amount of uncertainty will not lead the policymaker to retract a strict Pareto improvement. Note that the  $\varepsilon$  above is a function of the preference relations  $(R_1^*, \dots, R_m^*)$ . If we instead began with a  $\Omega$  and  $\bar{\pi}$ , with  $\bar{\pi}_1$  arbitrarily near 1, it is *not* the case that  $q' P_j(\omega_1)\bar{q}$  for each  $j$  implies, for all possible  $(R_1^*, \dots, R_m^*)$ , that  $q'$  is agent-based superior to  $\bar{q}$ .

### 3.2 Partial comparability

The heart of the agent-based approach is its stipulation of a complete ex ante preference relation for each individual. Instead of remaining agnostic about

<sup>4</sup> This conclusion would be false if the assumption mentioned in the previous footnote were also imposed.

an agent  $j$ 's welfare when ex post preferences conflict, the ex ante relation  $R_j^*$  posits at least a weak preference judgment between any pair of policy options.

What types of interpersonal comparisons does the completeness of  $R_j^*$  imply? This question is motivated in part by Sen's (1970) argument that complete social preference relations do not always require "unit" comparability of individual utility functions, wherein social preferences are constructed from a single relative weighting of individual utilities. As Sen shows, multiple relative weights on utilities can sometimes generate the same social preference ordering.<sup>5</sup> Insofar as ex ante preference relations do not hinge on onerous levels of interpersonal comparability, the agent-based position gains credibility as a philosophically consistent extension of the Paretian position. Unfortunately, Sen's condition for a social preference relation to be generated from multiple utility weights implies that social preferences regard no two social choices as indifferent. In our context of choice under uncertainty, any von Neumann-Morgenstern preference relation will violate this requirement if  $A$  contains three or more elements. Sen's results therefore cannot be applied.

We begin by defining different types of comparability. For each  $\omega_k$ , fix an arbitrary expected utility representation of  $R_j(\omega_k)$ , say  $\bar{v}_j(\cdot, \omega_k)$ , and identify with each  $b \equiv (b_1, \dots, b_s) \in \mathbf{R}_{++}^s$  the ex ante utility  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$ . Since we hold  $\pi$  fixed throughout at  $\bar{\pi}$ , we suppress  $\pi$  in the definition of  $j$ 's welfare; given that  $\bar{v}_j(\cdot, \omega_k) = \bar{\pi}_k v_j(\cdot, \omega_k)$  for some utility  $v_j(\cdot, \omega_k)$ , this is a notational convention. In order to avoid multiple  $b$ 's representing the same relative weights on ex post preferences we restrict ourselves to  $A_{++}^{s-1} \equiv \{b \in \mathbf{R}_{++}^s : \sum_{k=1}^s b_k = 1\}$ . Given a  $B \subset A_{++}^{s-1}$ , we consider  $j$ 's welfare to be at least as high at  $q$  as at  $q'$  if and only if for all  $b \in B$ ,  $V(q, b, \bar{v}) \geq V(q', b, \bar{v})$ . Let us then say that the *aggregation ordering induced by  $B$* ,  $R(B)$ , weakly prefers  $q$  to  $q'$ .

We can now provide formal definitions of comparability in terms of subsets of  $A_{++}^{s-1}$ .

**Definition 4.** *Unit comparability holds when  $B \subset A_{++}^{s-1}$  consists of a single element. Noncomparability holds when  $B = A_{++}^{s-1}$ . Partial comparability holds when  $B$  is neither unit nor noncomparable.*<sup>6</sup>

Unit comparability requires that only one set of weights on ex post preferences be used in  $j$ 's ex ante utility function. In line with the classical utilitarian position, the cardinal utility "units" of ex post preference relations are fixed. Noncomparability in contrast places no restrictions on how ex post preferences are aggregated. Partial comparability stakes out a middle ground.

<sup>5</sup> Our topic is different from Sen's in that we are constructing a single agent's preferences rather than a social preference ordering. But formally the two problems are identical.

<sup>6</sup> Our definitions differ slightly from Sen (1970) where comparability is first defined in terms of utility representations, and then linked to the coefficients that scale utility functions. Also, Sen defines partial comparability so as to include unit comparability as a special case.

We assume henceforth that  $s \geq 2$ ; otherwise, no issue of utility weighting arises and unit and noncomparability coincide.

If the  $R_j(\omega_k)$  differ and  $B$  is the noncomparability subset, then  $R(B)$  obviously cannot be complete. But this does not imply that unit comparability is necessary for the completeness of  $R(B)$ . To see a simple example, restrict the set of choices to  $A$  rather than  $Q = \mathcal{A}(A)$ , and suppose that there exists a  $b \in B$  such that  $\sum_{k=1}^s b_k \bar{v}_j(a_i, \omega_k) \neq \sum_{k=1}^s b_k \bar{v}_j(a_{i'}, \omega_k)$  for all distinct  $a_i, a_{i'} \in A$ . Clearly, if  $B$  is a set of sufficiently small but nonzero radius, then  $R(B)$  is a complete ordering. The assumption of unit comparability is thus needlessly strong in this case: under the right form of partial comparability, all of the resulting ex ante utility function for  $j$  rank the pure policy options in the same way.

If  $Q$  is the available set of policy options, when do partial comparability subsets induce complete aggregation orderings?<sup>7</sup> That is, given an ex ante utility function for  $j$  with weights  $b$ , when do there exist alternative weights inducing the same ordering? For an arbitrary  $b$ , let  $B(b)$  denote the set of relative utility weights that induce the same ordering as  $b$ :  $B(b) = \{ \hat{b} \in \mathcal{A}_{++}^{s-1} : R(\{ \hat{b} \}) = R(\{ b \}) \}$ .

For any integer  $r > 0$ , let  $\mathbf{1}^n$  denote the vector of  $n$  1's.

**Definition 5.** *An agent  $j$ 's preferences are regular if and only if  $\min[n, s + 1]$  vectors in  $\{ \bar{v}_j(\cdot, \omega_1), \dots, \bar{v}_j(\cdot, \omega_s), \mathbf{1}^n \}$  are linearly independent.<sup>8</sup>*

**Theorem 4.** *For any  $b$ : (1) if  $n > s$  and preferences are regular, then  $B(b)$  is unit comparable, (2) if  $n < s$ , or if  $n = s$  and  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$  is not constant, then  $B(b)$  is not unit comparable.*

*Proof.* First assume  $n > s$ . We show there is no distinct pair  $(b, b')$  such that  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$  and  $\sum_{k=1}^s b'_k \bar{v}_j(\cdot, \omega_k)$  represent the same ordering. If there were such a pair,  $\sum_{k=1}^s b'_k \bar{v}_j(\cdot, \omega_k)$  must be an increasing affine transformation of  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$ . Equivalently, there must be a  $\alpha \in \mathbf{R}_{++}^{s-1} \setminus \{ \mathbf{1}^{s-1} \}$  such that  $b_1 \bar{v}_j(\cdot, \omega_k) + \sum_{k=2}^s \alpha_k b_k \bar{v}_j(\cdot, \omega_k)$  is an increasing affine transformation of  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$ . Observe that the equation

<sup>7</sup> This question is not answered by the related issue of whether there is a unique affine combination of individual von Neumann-Morgenstern utilities yielding a given social welfare function (see Fishburn (1984) and Coulhon and Mongin (1989)). The difference is that, following Sen, we are interested in whether multiple relative weights on the  $\bar{v}_j(\cdot, \omega_k)$  generate the same ex ante ordering, not the same ex ante utility function. In this vein, observe that affine independence (Coulhon and Mongin's necessary condition for uniqueness of utility weights) is not necessary for there to be a single  $b$  generating a given ex ante ordering. For example, suppose  $n = 3$  and  $s = 2$ , and let  $j$ 's utility for the three pure policies be  $(3, 2, 1)$  at  $\omega_1$  and  $(1, 2, 3)$  at  $\omega_2$ . Then  $b_1 = b_2 = 1/2$  is the unique  $b \in \mathcal{A}_{++}^{s-1}$  yielding the ex ante ordering that ranks all policies indifferent.

<sup>8</sup> Given that any expected utility representation of each  $R_j(\omega_k)$  is an increasing affine transformation of any other representation, regularity does not depend on which representation is denoted  $\bar{v}_j(\cdot, \omega_k)$ .

$$b_1 \bar{v}_j(\cdot, \omega_1) + \sum_{k=2}^s \alpha_k b_k \bar{v}_j(\cdot, \omega_k) = \beta \sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k) + \gamma \mathbf{1}^n. \quad (3.1)$$

is satisfied at  $\alpha = \mathbf{1}^{s-1}$ ,  $\beta = 1$ ,  $\gamma = 0$ . Regularity implies that  $b_2 \bar{v}_j(\cdot, \omega_2), \dots, b_s \bar{v}_j(\cdot, \omega_s), \sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$ , and  $\mathbf{1}^n$  are linearly independent. Hence, since  $n \geq s + 1$ , there is no  $\alpha \neq \mathbf{1}^{s-1}$  satisfying (3.1).

Next assume  $n < s$ . Beginning with an arbitrary  $b$ , we show that there is a  $b' \neq b$  such that  $\sum_{k=1}^s b'_k \bar{v}_j(a_i, \omega_k)$  is an increasing affine transformation of  $\sum_{k=1}^s b_k \bar{v}_j(a_i, \omega_k)$ . Observe that the set of  $(\alpha, \beta, \gamma)$  solving (3.1) has dimension of at least  $s + 1 - n \geq 2$ . Suppose that all such solutions have  $\alpha = \mathbf{1}^{s-1}$ . The matrix  $[\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k) \mathbf{1}^n]$  has at least rank 1, however, and so there is at most a 1 dimensional set of  $(\beta, \gamma)$  consistent with  $\alpha = \mathbf{1}^{s-1}$ , a contradiction.

Finally, assume  $n = s$ . For arbitrary  $b$ , there is now at least a 1 dimensional set of  $(\alpha, \beta, \gamma)$  solving (3.1). Suppose again that all solutions have  $\alpha = \mathbf{1}^{s-1}$ . Since  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$  is nonconstant,  $\sum_{k=1}^s b_k \bar{v}_j(\cdot, \omega_k)$  and  $\mathbf{1}^n$  are linearly independent; consequently, there is a unique  $\beta$  and  $\gamma$  consistent with  $\alpha = \mathbf{1}^{s-1}$ , a contradiction. ■

Since we take policy-making uncertainty to be substantial, the case  $s \geq n$  is, if anything, the norm. The agent-based approach therefore in no sense relies on an assumption of unit comparability, or, put differently, on a commitment to a fixed system of interpersonal welfare comparisons. On the other hand, the types of partial comparability consistent with ex ante preference completeness are highly circumscribed. When  $s \geq n$ , the proof above suggests that the multiple utility weights that induce a single ex ante preference usually lie in a lower dimensional subspace of the set of possible utility weights. This is indeed the case. To induce complete preferences, therefore, partial comparability sets must be chosen carefully. As a consequence, the most plausible forms of partial comparability are incompatible with ex ante preference completeness. If a policymaker can only approximately identify the elements of  $B$ , then any  $b^*$  sufficiently close to a  $b \in B$  should also be in  $B$ . Partial comparability sets should therefore be open relative to  $\Delta_{++}^{s-1}$ , or at least contain open sets.

**Theorem 5.** *If  $B$  contains a nonempty open set,  $n \geq 3$ , and preferences are regular, then  $R(B)$  is incomplete.*

*Proof.* Given that sets open relative to  $\Delta_{++}^{s-1}$  have dimension  $s - 1$ , it is sufficient to show that if  $B$  contains a subset of dimension  $s - 1$ , then  $R(B)$  is incomplete. Given regularity, the maximal set of  $(\alpha, \beta, \gamma)$  satisfying (3.1) is a linear variety of dimension  $\max [0, s + 1 - n]$ . Since  $n \geq 3$ ,  $s + 1 - n < s - 1$ . Consequently, the set of  $\alpha$  satisfying (3.1), which has the same dimension as the maximal set  $B$  such that  $R(B)$  is complete, must have dimension less than  $s - 1$ . Hence, if  $B$  contains a subset of dimension  $s - 1$ ,  $R(B)$  is not complete. ■

To sum up, under trivially weak conditions noncomparability implies that agents cannot be represented by complete ex ante orderings. And although

partial comparability can in principle permit completeness when  $s \geq n$ , plausible forms of partial comparability preclude completeness.

An example illustrates the formal results. Let  $s = n = 3$ , and suppose that some  $j$ 's expected utility at  $\omega_1, \omega_2$ , and  $\omega_3$  can be represented by  $(9, 3, 0)$ ,  $(6, 3, 0)$ , and  $(0, 3, 6)$ , respectively. Each 3-tuple indicates the utility of  $a_1, a_2$ , and  $a_3$ , respectively. Suppose the policymaker posits that  $j$ 's ex ante welfare can be represented by the expected utility function  $(5, 3, 2)$ . These preferences can be induced by assigning the weight  $1/3$  to each state. Since  $s = n$  and ex ante preferences are not constant, Theorem 4 reports that there is a partial comparability subset such that  $R(B)$  is complete. And indeed, each element of  $B' = \{(\lambda, (1 - \lambda)/2, (1 - \lambda)/2) : 0 < \lambda < 1\}$  generates the posited ex ante preference relation. But preferences are regular; it follows that the *only* utility weights that generate the posited ex ante preferences are in  $B'$ . For instance, if  $b$  is an element of  $\{(1/3) + \mu, (1/3) + \mu, (1/3) - 2\mu) : -1/3 < \mu < 2/3, \mu \neq 0\}$  rather than  $B'$ , then  $b$  induces distinct ex ante preferences. Taking  $\mu$  arbitrarily near 0, we confirm that if  $B$  is open and  $(1/3, 1/3, 1/3) \in B$ , then  $R(B)$  is incomplete.

#### 4 The utility-independent and maximization approaches

A utility-independent ordering could be defined so as to maintain equivalence with the agent-based ordering. Given ex ante preference orderings  $(R_1^*, \dots, R_m^*)$  for the  $m$  individuals, define an *assignment of ex ante utilities* as a  $v = (v_1^*, \dots, v_m^*)$  such that  $v_j^* \in U_j(R_j^*)$  for all  $j$ . For each ex ante assignment, a corresponding social welfare function is defined by  $V(q, \bar{\pi}, \sum_{j \in J} v_j^*)$ . We might then let the utility-independent ordering rank  $q$  above  $q'$  if and only if  $V(q, \bar{\pi}, \sum_{j \in J} v_j^*) \geq V(q', \bar{\pi}, \sum_{j \in J} v_j^*)$  for all ex ante utility assignments and  $V(q, \bar{\pi}, \sum_{j \in J} v_j^*) > V(q', \bar{\pi}, \sum_{j \in J} v_j^*)$  for some assignment. Defining utility-independent and maximization definitions of optimality in the obvious way, Theorem 1 holds.

But when a policymaker's uncertainty over the preferences of agents is not experienced by the agents themselves, the above definition of an ex ante assignment restricts utility representations in ways unrelated to the preferences of any agent. (See Sect. 3.2 above on the extent of these restrictions.) The provisional definition of utility independence given above therefore invokes a system of interpreference welfare comparisons. This is an intractable feature of agent-based orderings and their utility-independent equivalents. Consider Harsanyi's (1953) argument that although  $j$  may never have experienced the ex ante vantage point of not knowing what preferences he or she will ultimately have,  $j$  can still imagine choosing under such circumstances. Even granting the legitimacy of the Harsanyi thought experiment, choice under the veil of ignorance constitutes a view of how to make interpersonal comparisons of welfare: a decision-maker must decide whether a policy change creates more well-being for the preference relations who gain than the diminution in

well-being of the preference relations who lose. Any lack of unanimity about the appropriate social welfare function, e.g., a dispute about the capacities of preference relations for satisfaction, will therefore induce disagreement over the possible  $R_j^*$  that  $j$  might have. Consequently, there is no Pareto-like way for the policymaker to choose the  $R_j^*$ .

We therefore turn to a definition of utility independence that avoids the artificial construct of ex ante preferences. In the following definition of an admissible utility assignment, we again require that agents with identical preferences are represented by the same utility function.

**Definition 6.** *An assignment of ex post utilities,  $v = (v_1, \dots, v_m)$ , is admissible if, for all agents  $j$  and  $j'$  and states  $\omega_k$  and  $\omega_{k'}$ ,*

- (1)  $E(q, v_j(\cdot, \omega_k)) \geq E(q', v_j(\cdot, \omega_k)) \Leftrightarrow qR_j(\omega_k)q'$ ,
- (2)  $R_j(\omega_k) = R_{j'}(\omega_{k'})$  implies  $v_j(\cdot, \omega_{k'}) = v_{j'}(\cdot, \omega_{k'})$ .

**Definition 7.** *The policy  $q$  is superior by the utility-independent ordering to  $q'$  in a model with policy-making uncertainty if  $V(q, \bar{\pi}, \sum_{j \in J} v_j) \geq V(q', \bar{\pi}, \sum_{j \in J} v_j)$  for all admissible ex post assignments  $v$  and  $V(q, \bar{\pi}, \sum_{j \in J} v_j) > V(q', \bar{\pi}, \sum_{j \in J} v_j)$  for some admissible assignment.*

Correspondingly, a policy  $q$  is utility-independent optimal if and only if there does not exist a  $q'$  superior to  $q$  by the utility-independent ordering. Finally,  $q$  is maximization optimal if there is some admissible assignment  $v$  such that for all  $q'$ ,  $V(q, \bar{\pi}, \sum_{j \in J} v_j) \geq V(q', \bar{\pi}, \sum_{j \in J} v_j)$ . When there is no policy-making uncertainty, these definitions coincide with their Sect. 2 counterparts.

The welfare function  $V(q, \bar{\pi}, \sum_{j \in J} v_j)$  corresponding to an ex post utility assignment  $v$  establishes relative weights for the ex post utility functions. That is, given a base assignment, say  $\bar{v}$ , each  $v_j(\cdot, \omega_k)$  of any ex post assignment  $v$  equals  $b_{kj}\bar{v}_j(\cdot, \omega_k)$  for some  $b_{kj} > 0$ . Furthermore, since there are admissible assignments that place arbitrarily large weight on each  $v_j(\cdot, \omega_k)$ , no  $R_j(\omega_k)$  can be worse off at  $q$  compared to  $q'$  if  $q$  is utility-independent superior to  $q'$ . Consequently, the utility-independent ordering (in the sense of Definition 7) is no stronger than any of the agent-based orderings: if  $q$  is utility-independent superior to  $q'$ ,  $q$  is superior by any of the agent-based orderings to  $q'$ . But the reverse implication need not hold. Policy  $q$  can be agent-based superior to  $q'$  even though one of the ex post preferences of some agent  $j$  prefers  $q'$  to  $q$ ; all that is necessary is that  $R_j^*$  assigns sufficient weight to a  $\bar{v}_j(\cdot, \omega_k)$  that prefers  $q$  to  $q'$ . In this sense, an agent-based ordering can be strictly weaker than the utility-independent ordering.

The question remains whether, relative to most arbitrary status quo policies  $\bar{q}$ , there is nevertheless at least some policy superior to  $\bar{q}$  by the utility-independent ordering. Although our general stance is that under mild conditions all policies are utility-independent optimal, such a conclusion clearly cannot hold universally. If, for instance,  $(1, 0, 0, \dots, 0)P_j(\omega_k)(0, 1, 0, \dots, 0)$  for all  $\omega_k$  and  $j$ , then any  $q$  with  $q_2 > 0$  cannot, by any Pareto-like rule, be optimal. For paralysis to occur, there must be sufficient diversity in the support of the distribution of possible preferences. To make this more precise,

define  $\text{supp } \Omega$  as the set  $\{R : R = R_j(\omega_k) \text{ for some } \omega_k \in \Omega \text{ and some } j \in J\}$ . Also, define a set of  $n$ -tuples of preference relations, say  $\mathfrak{R}$ , to be *open* if and only if for all  $(R_1, \dots, R_n) \in \mathfrak{R}$  there exists an  $\varepsilon > 0$  such that if, for each  $j \in \{1, \dots, n\}$ ,  $u_j$  and  $u'_j$  are expected utility representations of  $R_j$  and  $R'_j$  respectively and  $\max_{a_i \in A, j \in \{1, \dots, n\}} |u_j(a_i) - u'_j(a_i)| < \varepsilon$ , then  $(R'_1, \dots, R'_n) \in \mathfrak{R}$ .

**Theorem 6.** *There exists an open set of preference  $n$ -tuples such that if  $\text{supp } \Omega$  contains an element this set, then every policy  $q$  is optimal by the maximization and utility-independent definitions.*

The significance of Theorem 6 lies partly in the fact the probabilities underlying  $\Omega$  are nowhere mentioned. Hence, no matter how small the probability of the  $n$  elements in the support, the policy paralysis conclusion follows. The openness of the set of preferences implies that the  $n$ -tuples cannot be dismissed as special constructions that should be assigned probability 0. Indeed, as the proof below indicates, allowable preferences are easily and broadly devised.

The logic underlying Theorem 6 is simple. Since there are social welfare functions placing arbitrarily large weight on any ex post preference relation, the utility-independent and maximization definitions of optimality effectively declare a policy  $a$  to be optimal when any other policy leaves some potential agent strictly worse off compared to  $a$ . Thus, even with fewer agents than policy options, the number and diversity of potential agents can lead to policy paralysis – despite Theorem 3.

In order for  $\text{supp } \Omega$  to contain an  $n$ -tuple of preferences, we must have  $ms \geq n$ . When  $ms < n$ , in contrast, the reasoning behind Theorem 3 applies. As long as the set of feasible utilities generated by the  $ms$  ex post preference relations has nonempty interior, which occurs generically when  $ms < n$ , every strictly mixed policy can be dominated by a boundary policy leaving each of the  $ms$  preference relations better off. Consequently, none of the strictly mixed policies can be utility-independent or maximization optimal: policy paralysis does not occur. (Formally, the relevant diversity condition now appears as the requirement that  $\text{co}\{w \in \mathbf{R}^{ms} : w = (v_j(a_i, \omega_1), \dots, v_j(a_i, \omega_s))_{j=1}^m, i = 1, \dots, n\}$  is open.)

Theorem 6 also completes the analysis of the no policy-making uncertainty case studied in Sect. 2. In the absence of policy-making uncertainty ( $s = 1$ ), Theorem 6 implies there are open sets of preferences for which every policy is optimal – whether by the agent-based, utility-independent or maximization definitions – if  $m \geq n$ .

*Proof of Theorem 6.* Pick any nonsingular  $n \times n$  matrix,  $B$ , such that for some real number  $c$ ,  $B(c, \dots, c) = p \gg 0$ . Let  $D = [d_{ij}] = B^{-1}$  and define  $u_j : A \rightarrow \mathbf{R}$ ,  $j = 1, \dots, n$ , by  $u_j(a_i) = d_{ij}$ . Let  $R_1, \dots, R_n$  denote the corresponding preference relations. It follows that, for any pure policy  $a_i$ ,  $\sum_{j=1}^n p_j u_j(a_i) = c$ . Thus, for a model with no uncertainty that consists of only these  $n$  specified agents, any mixed policy  $q$  is optimal by the (weak) maximization definition of Sect. 2 using the assignment  $(p_1 u_1, \dots, p_n u_n)$ . Therefore, given Theorem 1, for

any  $q' \neq q$  it must be that either for some  $j \in \{1, \dots, n\}$ ,  $qP_jq'$ , or for all  $j$ ,  $qI_jq'$ . The latter possibility can be excluded since, given an arbitrary  $n$ -vector  $g$  of utility levels, the nonsingularity assumption implies that there is a unique nonzero  $n$ -vector  $r$  solving the  $n$  equations  $\sum_{i=1}^n r_i u_j(a_i) = g_j, j = 1, \dots, n$ ; thus no two distinct probability vectors  $q$  and  $q'$  deliver the same vector of utilities.

Consequently, if  $\text{supp } \Omega$  contains the  $n$  preference relations just constructed, then for all  $q$  and  $q'$  there must be a state  $\omega_k \in \Omega$  and agent  $j$  such that  $qP_j(\omega_k)q'$ . Hence, by Theorem 1, any  $q$  is optimal by the weak maximization definition in the model consisting (with certainty) of the  $m_S$  individuals with preference relations  $R_j(\omega_k), j \in J, \omega_k \in \Omega$ . Consequently, there exists a utility assignment  $v$  satisfying condition (1) of Definition 6 such that  $E(q, \sum_{j \in J} \sum_{\omega_k \in \Omega} v_j(\cdot, \omega_k)) \geq E(q', \sum_{j \in J} \sum_{\omega_k \in \Omega} v_j(\cdot, \omega_k))$  for all  $q'$ . By defining  $\bar{v}_j(\cdot, \omega_k) = (1/\bar{\pi}_k)v_j(\cdot, \omega_k)$ , we therefore have  $V(q, \bar{\pi}, \sum_{j \in J} \bar{v}_j) \geq V(q', \bar{\pi}, \sum_{j \in J} \bar{v}_j)$  for all  $q'$ . Following the proof of Theorem 2, for each pair  $(j, \omega_k)$ , define  $J(j, \omega_k) \subset J \times \Omega$  by the condition  $(h, \omega_t) \in J(j, \omega_k)$  if and only if  $R_h(\omega_t) = R_j(\omega_k)$ . Letting  $\hat{v}_j(\cdot, \omega_k) = (1/\#J(j, \omega_k)) \sum_{(h, \omega_t) \in J(j, \omega_k)} \bar{v}_h(\cdot, \omega_t)$ ,  $\hat{v}$  is both admissible and  $V(q, \bar{\pi}, \sum_{j \in J} \hat{v}_j) \geq V(q', \bar{\pi}, \sum_{j \in J} \hat{v}_j)$  for all  $q'$ , demonstrating optimality of  $q$  by the maximization and utility-independent definitions of this section.

To show that this conclusion holds for an open set of preferences, we need only observe that  $p = B(c, \dots, c)$  is continuous in the components of  $D$ ; thus, for a small enough change in the  $n$  utility functions, we still have  $p \gg 0$  and the construction of an admissible assignment satisfying the maximization and utility-independent definitions of optimality proceeds as above. ■

## 5. Conclusion

Theorem 6 shows that if a policymaker is uncertain about agents' precise characteristics then, under mild restrictions, every policy choice is the maximum of some expected social welfare function. Since policy-making uncertainty is a frequent – even universal – occurrence, this result is disquieting: to discriminate among policies, interpersonal comparisons are necessary.

The agent-based approach also does not provide a trouble-free resolution of the policy paralysis problem: there is no ex ante vantage point at which agents form preferences over the uncertain prospects that policymakers must choose from. Policymakers must therefore assemble ex ante preferences for agents, perhaps along the lines proposed by Harsanyi (1953). Paretian welfare economics is therefore not unworkable in principle; the difficulty is that it must rely on some scheme for making interpersonal comparisons of welfare, thereby undercutting one of Paretianism's prime rationales. Furthermore, the Harsanyi thought experiment of choice behind the veil of ignorance hardly supplies a transparent method of preference aggregation: what meaning can be ascribed to the preference judgments that agents would make when ignorant of what personality they will ultimately have? Not surprisingly, Harsanyi

has been sketchy about how such judgments might be constructed. The need for interpersonal comparisons is therefore a serious setback to the Paretian program.

The agent-based approach may seem to suffer from a second defect in its seemingly contradictory stance towards interpreference comparisons. If a policymaker can make such comparisons, why not simply maximize social welfare instead of instituting agent-based Pareto improvements or optima? Although there is force to this objection, it overlooks the unanimity argument in favor of the agent-based approach. If a government institutes an agent-based improvement, the government can respond to an agent  $j$  who turns out to suffer from the change that ex ante the government had at least thought that  $j$  would benefit. And if the government is almost certain about the characteristics of agents, then with high probability none of the actual existing agents will be worse off. Our results therefore identify the unanimity argument, rather than any putative absence of interpersonal comparisons, as the rationale underlying Paretian policy analysis.

Although the agent-based and utility-independent approaches offer opposed strategies for extending the Pareto criterion to environments with policy-making uncertainty, the two approaches can be placed in a unified conceptual framework. The agent-based approach posits a single ex ante preference relation over policies for each agent; a policy change is weakly agent-based superior if all  $m$  of these preference relations weakly prefer the change. The utility-independent approach, in contrast, effectively requires for each agent that all of the agent's  $s$  ex post preference relations approve a policy change. In fact, the utility-independent approach identifies a broader set of preference relations: the policy changes that are weakly recommended must, for each agent  $j$ , be weakly superior according to all of the infinite number of ex ante preference relations  $R_j^*$  that obey Assumptions 1 and 2 of Sect. 3.1. (The latter are the preference relations generated by the positively weighted sums of utility representations of  $j$ 's ex post preferences.) So, whereas the agent-based approach identifies a single preference relation over policies, the utility-independent approach identifies an infinite set of preference relations.

This observation suggests a mathematical compromise: posit an arbitrary set of preference relations over policies for each agent  $j$  and require that policy changes are approved by each relation in each agent's posited set. If this set contains more than one preference relation but not all of the  $R_j^*$  satisfying assumptions 1 and 2 we have an intermediate policy ordering: such an ordering will, like the utility-independent criterion, weakly recommend all policy changes that make all  $ms$  potential agents weakly better off, but, like the agent-based criterion, it can also recommend policy changes that harm some of the ex post preference relations. Depending on the size of these posited sets of preferences (and on the comparative magnitude of  $ms$  and  $n$ ), policy paralysis may or may not occur. It is interesting to note, however, that when  $n > m$  there typically will exist infinite sets of preference relations for each individual such that almost all policies are dominated by other policies. That is,

for almost all policies  $q$ , there is a policy  $q'$  that is weakly preferred to  $q$  by each preference relation in each individual's posited set and there is at least one preference relation in some individual's set that strictly prefers  $q'$  to  $q$ . These infinite sets can be constructed as follows: begin with a single  $R_j^*$  for each  $j$  such that any strictly mixed policy is strictly dispreferred by these  $m$  relations to some boundary policy. For generic choices of the  $R_j^*$ , all ex ante relations sufficiently close to  $R_j^*$  will continue to judge any strictly mixed policy to be inferior to some boundary policy. Small infinite sets of ex ante preferences can thus rank policies as effectively as the agent-based criterion. Although it is therefore fair to conclude that the agent-based approach imposes a needlessly demanding system of preference aggregation, the above compromise is of limited practicality. Unless each agent's posited set of preferences contains all of the agent's ex post preference relations, interpersonal comparisons will not be eliminated; and when all ex post preferences are included, Theorem 6 indicates that policy paralysis can occur if  $ms \geq n$ . Furthermore, the agent-based practice of assuming that agents and policymakers face the same state space and that each agent is represented by a single complete preference relation yields a simpler and more concise model.

Policy-making under ignorance of agent characteristics has often been analyzed as a problem of mechanism design (see Holmström and Myerson 1983). Mailath and Postlewaite (1990) is particularly relevant to the policy paralysis question. They consider a model where agents have private valuations of a public works project and ask whether there are mechanisms that institute projects that, with high probability, generate a positive total surplus ex post. Loosely speaking, they show that as the number of agents becomes large the probability that such projects are adopted approaches zero if agents are allowed to veto projects that harm them in expectation. In other words, even a seemingly beneficial deviation from the status quo will not occur in large populations. Their conclusion has an obvious kinship to the current paralysis results, but our framework is quite different. First, as argued in the introduction, ex post optimality is too ambitious a goal. We therefore ask which ex ante policy criterion should be used and how the *size* of the set of optimal policies depends on the choice of welfare criterion. Second, since we take the policymaker's information to be fixed, our analysis is in the spirit of social choice theory rather than mechanism design. Our rationale for fixing the policymaker's information is that although there are situations in which policymakers can hope to establish information-extracting mechanisms, there are other circumstances in which policy choices must be made with the information at hand (e.g., prior to the establishment of a mechanism). In addition, although we will not make any formal argument to this end, mechanisms designed for large groups of agents are likely to transmit information imperfectly; policymakers consequently always face at least some residual uncertainty. Theorem 6 is therefore applicable and the paralysis conclusion holds.

Compared to the current social choice setting, the analysis of policy paralysis in standard economic contexts is more difficult. The effect of traditional economic policy instruments (e.g., tax rates and endowment redistributions)

on final allocations can be complicated, and agents with identical preferences over their own consumption bundles have distinct preferences over the profile of all agents' consumption bundles. The latter fact breaks the connection between policies for which any deviation harms some potential agent and the utility-independent and maximization definitions of optimality, thus complicating the analysis of optimality. A companion paper, Mandler (1996), studies the economic setting with private consumption; the conclusions are weaker, but qualitatively similar.

## References

- Arrow K (1951) *Individual Values and Social Choice*, Wiley, New York 2nd edn, 1963
- Bergson A (1938) A reformulation of certain aspects of welfare economics. *Q J Econ* 52: 310–334
- Coulhon T, Mongin P (1989) Social choice theory in the case of Von Neumann-Morgenstern utilities. *Soc Choice Welfare* 6: 175–187
- Fishburn P (1984) On Harsanyi's utilitarian cardinal welfare theorem. *Theory Decis* 17: 21–28
- Hammond P (1983) Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty. In: Pattanaik P, Salles M (eds) *Social choice and welfare*. North-Holland, Amsterdam
- Harsanyi J (1953) Cardinal utility in welfare economics and in the theory of risk-taking. *J Polit Econ* 61: 434–435
- Harsanyi J (1955) Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. *J Polit Econ* 63: 309–321
- Harsanyi J (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press, Cambridge
- Holmström B, Myerson R (1983) Efficient and durable decision rules with incomplete information. *Econometrica* 51: 1799–1819
- Lipsey R, Lancaster R (1956) The general theory of second best. *Rev Econ Stud* 24: 11–32
- Mailath G, Postlewaite A (1990) Asymmetric information bargaining problems with many agents. *Rev Econ Stud* 57: 351–367
- Mandler M (1996) Policy paralysis in general equilibrium theory. Mimeo, Harvard University
- Samuelson P (1947) *Foundations of Economic Analysis*. Harvard University Press, Cambridge
- Sen A (1970) Interpersonal aggregation and partial comparability. *Econometrica* 38: 393–409