Institutions and their ethical evaluation

Prasanta K. Pattanaik
Department of Economics, University of California
Riverside, CA 92521, USA
prasanta.pattanaik@ucr.edu

Yongsheng Xu
Department of Economics, Georgia State University
Atlanta, GA 30302, USA
yxu3@gsu.edu

This version: 16 October 2017

Abstract. An institution is viewed as an arrangement of the society through which various social states emerge as outcomes from the actions and interactions of individuals belonging to the society. Following an earlier contribution by Pattanaik and Suzumura (1996), we subscribe to the view that the ethical desirability of an institution is based on the goodness of the outcomes emerging from individuals' actions and interactions via the institution as well as the intrinsic attractiveness of the institution itself. In this paper, we explore the structure of an individual's ethical evaluations of institutions. For this purpose, we discuss how social choices are made through institutions given profiles of individual preferences over social states, propose several *a prior* properties of institutions and examine their plausibility, and introduce and discuss a lexicographic maximin rule for evaluating institutions ethically.

Keywords: institution, institutional framework, ethical evaluation, social choice rule, rationality, distribution of power, equilibrium outcome, Pareto optimality, lexicographic maximin rule

JEL Classification Numbers: D02, D63, D71, P00, A13

1. Introduction

The social state[1] that emerges as the outcome for a society is the result of the actions and interactions of numerous individuals and organizations of individuals belonging to the society. These actions and interactions typically take place within a framework of laws, conventions, and norms provided by legal, political, social, and economic institutions. In ethically evaluating the desirability of these institutions for the society, not only do we take into account the goodness or badness of the outcomes that result from the actions taken by individuals given the rules of those institutions, but we also often consider the intrinsic attractiveness of institutions from an ethical point of view. Thus, in justifying the institution of competitive markets, besides appealing to the Pareto optimality of allocations which result from the institution of competitive markets, economists have also invoked the freedom that competitive markets provide to the agents in an economy (see, for example, Buchanan (1986), Friedman (1962), and Sen (1993)). The fact that institutions may have their own intrinsic ethical appeal in addition to their instrumental value in reaching ethically desirable social outcomes makes the structure of overall ethical judgments of institutions somewhat complex. In this paper, we study the structure of such ethical judgments regarding institutions. At the risk of emphasizing the obvious, we would like to clarify that we are not dealing with the problem of how the society should aggregate the different individuals' opinions about alternative institutional frameworks so as to choose a specific institutional framework. Instead, we study the problem of how a given individual – we call her an ethical evaluator (EE) – may approach the task of evaluating those institutions in terms of their ethical desirability.[2]

The problem of evaluating institutions has been studied earlier by Pattanaik and Suzumura (1996). We draw on the insights of Pattanaik and Suzumura; in particular, we use their notion of an extended social alternative, i.e. the notion of a pair the first component of which is an institution and the second

[1] We use the term "social state" (or, equivalently, the term "social alternative") as it is conventionally used in the theory of social choice to denote a complete description of all aspects of the society excepting the decision-making process in the society. Thus a social state does not include any specification of the institutions in the society.

[2] Little (1952) and Bergson (1954) were some of the earliest writers who drew attention to the distinction between the problem of aggregating people's opinions about various social options to reach a social decision and an individual's problem of ethically evaluating the different options confronting a society. Sen (1977) provides what constitutes perhaps the clearest and most precise statement of this distinction and its importance, and Pattanaik (2005) presents a contrast between the views of Little and Bergson and those of Arrow (1951; 1963) on such problems and related issues concerning social welfare and their respective methodological positions in welfare economics.

component of which is a social alternative.[3] But we depart from their basic framework in two significant respects. First, we explicitly introduce into the formal framework the possibility that some individuals may act in a way that is not permissible under the institution prevailing in the society. This is of some importance since, no matter how carefully the institutional rules are crafted, one can hardly ensure that all individuals in the society will always comply with the rules of an institution (see Section 3.1). Second, unlike Pattanaik and Suzumura (1996), who assume that the evaluator predicts the outcome of an institution on the basis of a fixed profile of individual preference orderings over the social alternatives, we argue that the evaluator needs to take into account possible changes in the profile of individual orderings (see Section 4).

The plan of the paper is as follows. In Section 2, we introduce our basic notation. Section 3 introduces the notions of an institution and an institutional framework and discusses how social choices are made through institutions given a profile of individual orderings over the social alternatives. Section 4 discusses how the EE may rank the different possible institutions corresponding to a given physical environment. In Section 5, we discuss a possible way of relaxing the assumption made in previous sections, namely, the assumption that every relevant game has a unique equilibrium outcome. Section 6 discusses certain a priori properties of institutions and their plausibility. We conclude in Section 7.

2. The basic notation

Let $N = \{1, \dots, n\}$ be the set of individuals constituting the society with $\infty > n \geq 2$, and let Y be the set of all conceivable social alternatives, a social alternative being conventionally defined as a complete description of the state of the society except for the process or mechanism through which the social decisions are arrived at. At any given time, the society is faced with a set of physically feasible social alternatives and this set can be different at different points of time. Let $\mathbb{X}$ be a non-empty class of non-empty (finite) subsets of Y. An element of $\mathbb{X}$ is to be interpreted as a set of social alternatives which may come up as the set of all physically feasible social alternatives available to the society at some point

[3] In a related contribution, Pattanaik and Xu (2014) study various types of ethical rules (e.g., consequentialist and non-consequentialist) for evaluating extended social alternatives, though their notion of an extended social alternative has three components: a social alternative, an institution, and a profile of individual strategies.

of time. We shall assume that, at any given time, each individual $i \in N$ has a weak preference relation ("at least as good as"), R_i, defined over Y and that for every $i \in N$, R_i is an ordering, i.e., R_i is reflexive, transitive and connected over Y. P_i will denote the asymmetric factor of R_i. The interpretation that we adopt for an individual's ordering over Y is the Arrovian interpretation (See Arrow, 1951, p. 17) that it reflects "whatever standards" individual i "deems relevant" in ranking the social alternatives. Let $\mathcal{P}$ be the set of all preference orderings over Y. Then, at any given point of time, we have a profile, $(R_i)_{i\in N} \in \mathcal{P}^n$, of individual orderings over Y.

A *physical environment* (PE) for the society at any given time is an $(n+1)$-tuple $e \equiv (X, (S_i)_{i\in N})$, where: (i) $X \in \mathbb{X}$ is the non-empty set of all physically feasible social alternatives, and (ii) for all $i \in N$, S_i is the non-empty (finite) set of feasible strategies of i. The class of all possible physical environments will be denoted by $\mathcal{E}$; we shall assume $\mathcal{E}$ to be a finite class. S will denote $\times_{i\in N} S_i$. The elements of S will be denoted by s, s', etc. For every non-empty subset N' of N, $S_{N'}$ will denote $\times_{j\in N'} S$ and S_{-i} will denote $S_{N-\{i\}}$ for all $i \in N$.

3. Institutions and social choice

3.1. Institutions

Given $e \equiv (X, (S_i)_{i\in N}) \in \mathcal{E}$, an *institution* is defined as a $(2n+3)$- tuple $(N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$ where: (i) for all $i \in N$, T_i is a non-empty subset of S_i, T_i being interpreted as the set of all feasible strategies of i, which are also admissible (or, equivalently, permissible); and (ii) $g: S \longrightarrow X$ is an outcome function which specifies a social state in X as the outcome for every n-tuple of feasible strategies of the individuals in the society. Intuitively, given a PE, an institution specifies the admissible (or, equivalently,

permissible) strategies for each individual in the society[4] and specifies an outcome for every n-tuple of feasible strategies. For every non-empty subset N' of N, $T_{N'}$ will denote $\times_{j\in N'}\ T_j$ and T_{-i} will denote $T_{N-\{i\}}$ for all $i \in N$. For each $e \in \mathcal{E}$, let $\mathcal{M}_e$ be the (finite) set of all feasible institutions. An *institutional framework* is a functional rule J which specifies exactly one institution $M \in \mathcal{M}_e$ for every $e \in \mathcal{E}$. For all $e \in \mathcal{E}$ and for all institutional frameworks, J and J', we say that J and J' are e-variants if and only if $J(e) \neq J'(e)$ and $J(e') = J'(e')$ for all $e' \in \mathcal{E} - \{e\}$.

An institution, as we conceive it, reflects the collection of legal and social rules, regulations, conventions, and norms that govern what the individuals may or may not do and determines the outcomes of the strategies which the individuals may adopt. We would like to make a few remarks to clarify the concept of an institution. First, not only does the function g figuring in an institution M specify a social alternative for every n-tuple of feasible strategies, all of which are admissible, but it also specifies an outcome for every n-tuple feasible strategies some of which may not be admissible. For example, the outcome specified by g for an n-tuple of strategies some of which are not admissible may involve prescribed penalties for individuals adopting inadmissible strategies. This may be contrasted with the type of construction, often used to model the notion of individual rights[5], which involves the specification of permissible strategies of every individual without any explicit introduction of feasible but not permissible strategies, and an outcome only for each n-tuple of permissible strategies. While such a construction may be adequate for articulating the notion of individual rights, in modeling institutions it seems intuitively necessary also to specify what happens when some individuals choose to adopt strategies which are not admissible under the rules/ norms reflected in the institution under consideration. Second, legal entities such as the state are implicitly embedded in an institution in the

[4] Once the set, S_i, of all feasible strategies and the set, T_i, of all admissible and feasible strategies are specified for an individual i, the set of feasible but inadmissible strategies of i is, by implication, known to be $S_i - T_i$.

[5] See, for instance, Gaertner, Pattanaik, and Suzumura (1992) and Pattanaik and Suzumura (1996).

forms of: (i) the implied partition of S_i into the set of admissible strategies, T_i, and the set of inadmissible strategies, $S_i - T_i$, for each individual i in the society, and (ii) the outcome function. For example, the strategy "giving a hate speech", though feasible, can be made inadmissible, and if an individual uses this strategy, the individual may face a penalty under the outcome function g. Similarly, the strategy involving "non-payment of income tax" may not be an admissible strategy in an institution, and the outcome function g of the institution can stipulate something like "failure to pay one's income tax may result in one's imprisonment or seizure of one's properties." The role of the state in these examples is reflected in two features of the underlying institution, namely, the specification of admissible and inadmissible strategies for individuals and the specification of the outcomes corresponding to different n-tuples of feasible strategies. Finally, we are implicitly assuming that the rules, norms and conventions involved in an institution are coherent. This is not always the case. It is not difficult to find examples of societies where "honor-killing" and discrimination based on caste or ethnicity may be inadmissible under the laws of the country but are encouraged by the social norms and conventions. We shall ignore such cases of inconsistency between the different types of norms and rules prevailing in the society; alternatively, our notion of an institution may be thought of as referring only to the legal rules and conventions of the society, which are more likely to be coherent. For all institution $(N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, $g(S)$ and $g(T)$ denote, respectively, the sets $\{x \in X: x = g(s)$ for some $s \in S \equiv \times_{i\in N} S_i\}$ and $\{x \in X: x = g(t)$ for some $t \in T \equiv \times_{i\in N} T_i\}$.

3.2. Institutions and social choice

The process through which the society chooses a social alternative is visualized as follows. Suppose the society has an institutional framework J and the physical environment of the society is given by $(X, (S_i)_{i\in N})$. Given the PE $e = (X, (S_i)_{i\in N})$, J specifies an institution $M = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$. Given M, we have a game form $G = (N, X, (S_i)_{i\in N}, g)$, where N is the set of players, X is the set of all feasible social alternatives figuring in e and $(S_i)_{i\in N}$ is the profile of sets of feasible strategies figuring in

e. The game form G will be called the game form associated with the institution M. Now suppose $(R_i)_{i\in N}$ is the profile of individual preference orderings.[6] Then we have a game $(G, (R_i)_{i\in N})$, which will be called the game associated with the institution M and the preference profile $(R_i)_{i\in N}$. To proceed from this game to a social outcome, we need to introduce the notion of an equilibrium. Suppose α-equilibrium is some notion of equilibrium which is considered appropriate for predicting the outcome of the game. Let $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ be the set of all $x \in X$ such that $x = g(s)$ for some α-equilibrium s of the game $(G, (R_i)_{i\in N})$. In general, depending on the notion of α- equilibrium, there may not exist an α-equilibrium of the game $(G, (R_i)_{i\in N})$, in which case $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ will be empty, or there may be more than one α-equilibrium with non-identical outcomes, in which case $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ will have more than one element. For convenience of exposition, unless specified otherwise, we shall assume that, for every $e \in \mathcal{E}$, every $M \in \mathcal{M}_e$, and every $(R_i)_{i\in N} \in \mathcal{P}^n$, the game associated with M and $(R_i)_{i\in N}$ has a single equilibrium outcome[7].

3.3. Institutional frameworks and social choice rules

It may be tempting to interpret an institutional framework J as a social choice rule as it is usually understood in the theory of social choice, but one needs to be cautious about this. For our purpose, we define a *social choice rule* as a function, F, which, for every $(X, (R_i)_{i\in N)}) \in \mathbb{X} \times \mathcal{P}^n$, specifies exactly one non-empty subset $F(X)$ of X. Note that, if an institutional framework, J, is to be interpreted as a social choice rule as defined above, then, given J, the social outcome under J must depend exclusively on the set of feasible social alternatives and the profile of individual orderings over the set of all social

[6] Strictly speaking, we should consider $R_1|X$, $R_2|X, \dots, R_n|X$, i.e., the restrictions of $R_1, R_2, \dots, R_n$ to X as the relevant preferences of the players in this game, but, to reduce the notational burden, we write the preferences in this game as $R_1, R_2, \dots, R_n$ rather than as $R_1|X$, $R_2|X, \dots, R_n|X$ at the cost of slight looseness.

[7] Note that the game may have multiple equilibria but a single equilibrium outcome if the outcomes for all the equilibria are identical.

alternatives. Thus, it can be easily seen that the following property, Invariance, is a necessary and sufficient condition for us to be able to interpret J as a social choice rule.

Invariance: Let $e = (X, (S_i)_{i\in N})$ and $e' = (X', \ (S'_i)_{i\in N})$ be any two elements of $\mathcal{E}$ such that $X = X'$. Let $G = (N, X, \ (S_i)_{i\in N}, g)$and $G' = (N, X, \ (S'_i)_{i\in N}, g')$ be the game forms associated with $J(e)$ and $J(e')$, respectively. Then, for all $(R_i)_{i\in N} \in \mathcal{P}^n$, the equilibrium outcome of the games $(G, (R_i)_{i\in N})$ and $(G', \ (R_i)_{i\in N})$ must be the same.

This is an exceptionally restrictive condition. It basically requires that, if the physical environments e and e' are such that $X = X'$, then, for every given preference profile $(R_i)_{i\in N}$, the outcomes emerging from the institutions $J(e)$ and $J(e')$ must be the same no matter what the strategy sets and the outcome functions in the two games $(G, (R_i)_{i\in N})$ and $(G', \ (R_i)_{i\in N})$ associated, respectively, with $J(e)$ and $J(e')$, may be.

To see how an institutional framework J may violate Invariance, consider a two-person society $N = \{1, 2\}$, and two physical environments, $e = \big(\{\{x, y, z\}, (S_1, S_2)\big)$ and $e' = \big(\{x, y, z\}, (S'_1, S'_2)\big)$. Let the institutions $J(e)$ and $J(e')$ and the preference profile (R_1, R_2) be such that, given (R_1, R_2), the games $(G, (R_1, R_2))$ and $(G', \ (R_1, R_2))$, associated with $J(e)$ and $J(e')$, respectively, are as follows:

Table 1

$(\boldsymbol{G}, (\boldsymbol{R_1}, \boldsymbol{R_2}))$

		Individual 2	
		s_2	s_2'
	s_1	x	y
		(6, 12)	(10, 9)
Individual 1	s_1'	y	z
		(10, 9)	(8, 7)
	s_1''	z	z
		(8, 7)	(8, 7)

$(\boldsymbol{G}', (\boldsymbol{R_1}, \boldsymbol{R_2}))$

		Individual 2	
		s_2	s_2'
	s_1	x	y
		(6, 12)	(10, 9)
Individual 1	s_1''	z	z
		(8, 7)	(8, 7)

(For convenience, we have used the payoffs of the players to represent their respective orderings. The payoffs of the players for each social alternative are given by the pair of numbers below that alternative, the first number being 1's payoff and the second number being 2's payoff.)

Suppose Nash equilibrium is the relevant notion of equilibrium. y is the outcome for the sole Nash equilibrium, (s_1', s_2), of the game $(G,\ (R_1, R_2))$ associated with $J(e)$ and (R_1, R_2), while z is the outcome of the sole Nash equilibrium, (s_1'', s_2) of the game $(G', (R_1, R_2))$ associated with $J(e')$ and (R_1, R_2). Since the set of feasible outcomes in both cases is $\{x, y, z\}$, it is then clear that J violates Invariance and cannot be interpreted as a social choice rule. In fact, it can be checked that, in our example, J violates even the following condition which is much weaker than Invariance.

Restricted Invariance: Let $e = (X, (S_i)_{i\in N})$ and $e' = (X, (S_i')_{i\in N})$ be two elements of $\mathcal{E}$ such that $[S_i' \subseteq S_i$ for all $i \in N]$ and $[S_i' \subset S_i$ for some $i \in N]$. Let $G = (N, X, (S_i)_{i\in N}, g)$ and $G' = (N, X, (S_i')_{i\in N}, g')$ be the two game forms associated with the institutions $J(e)$ and $J(e')$, respectively, with: $[g(s) = g'(s)$ for all $s \in S']$ and $[g(S) = g'(S')]$. Then, for all $(R_i)_{i\in N} \in \mathcal{P}^n$, the equilibrium outcome of the games $(G, (R_i)_{i\in N})$ and $(G', (R_i)_{i\in N})$ must be the same.

4. The ethical ranking of institutions for a given physical environment

We shall assume that the EE has an ethical ordering ("at least as good as"), [$\geq$], defined over the set of all possible institutional frameworks, with [$>$] and [=] being, respectively, the asymmetric and symmetric parts of [$\geq$]. We shall assume that

(1) For all $e \in \mathcal{E}$ and all institutional frameworks $J, J', \bar{J}$, and $\bar{J}'$, if [J and J' are e-variants and so are $\bar{J}$ and $\bar{J}'$] and [$J(e) = \bar{J}(e)$ and $J'(e) = \bar{J}'(e)$], then [J [$\geq$] J' if and only if $\bar{J}$ [$\geq$] $\bar{J}'$].

Given (1), it is easy to see that, for every $e \in \mathcal{E}$, [$\geq$] induces an ordering $\trianglerighteq_e$ over $\mathcal{M}_e$, the class of all possible institutions corresponding to e in the following fashion:

(2) for all $M, M' \in \mathcal{M}_e$, $M \trianglerighteq_e M'$ if and only if J[$\geq$] J' for all e-variant institutional frameworks J and J' such that $J(e) = M$ and $J'(e) = M'$.

We shall now explore the structure of $\trianglerighteq_e$ where $e = (X, (S_i)_{i \in N})$ is some given PE. Since, given e and a profile of individual preference orderings over social alternatives, the society chooses a social alternative through an institution in $\mathcal{M}_e$, it is natural that the goodness or badness of the social outcomes, which may emerge from the institution and different preference profiles, will enter into a person's judgment of the goodness or badness of the institution. But we also often attach intrinsic ethical significance to certain features of specific institutions. Given a physical environment, an institution which gives greater freedom to individuals in terms of permissibility of their feasible strategies may lead to the choice of a social outcome that is, in terms of every individual's well-being, inferior to the social alternative that would be chosen if there was in place a different institution giving less freedom to every individual. But the fact that the former institution gives more freedom to all individuals in the sense we have explained may be considered to be a consideration of intrinsic moral significance. Pattanaik and Suzumura (1996) introduced the notion of an extended social alternative to analyze how an individual may combine these two aspects of institutions in ethically evaluating alternative combinations of institutions and social outcomes. We adopt their notion of an extended social alternative but our use of it in modeling how the EE may ethically rank the different institutions corresponding to a given physical environment differs from Pattanaik and Suzumura's (1996) use in two distinct ways. First, implicit in Pattanaik and Suzumura's (1996) analysis, there seems to be the assumption that, given an institution, no individual adopts a strategy that is impermissible for her under that institution. As we have noted earlier, this assumption can hardly be considered "realistic". Second, Pattanaik and Suzumura (1996) assume that, in considering the outcome that may result from a given institution, the EE considers a fixed profile of individual preferences; presumably, it is the preference profile prevailing at the time when the EE evaluates the institutions under consideration. We believe that this assumption is restrictive in many ways and needs to be modified. People's preferences over social alternatives often fluctuate. It is, however, difficult to change institutions frequently without incurring significant social and economic

costs. Therefore, in considering the outcomes that may emerge from an institution, not only does one need to consider the outcomes which emerge from the institution, given the preferences prevailing at the time when one assesses the institution, but one also needs to consider the outcomes that will emerge from the institution if the individuals' preferences change in some way after the institution is established. Should there be a law protecting people's freedom of speech, with prescribed penalties for persons who seek to interfere with other people's free expression of their opinions? Clearly, in considering the outcomes that may result if there is no such law, it would seem to be rather inadequate to argue that, since the current preferences of all individuals in the society happen to be quite liberal, there will be no interference with people's freedom to express their opinions even in the absence of any laws to protect people's freedom of speech. If there is a chance that people's preferences may become less liberal, then one should also consider the outcomes that may result from such less liberal preferences in the absence of any law protecting freedom of speech. In general, the possibility of changes in people's preferences over social alternatives needs to be taken into account in considering the outcomes of institutions.

Given e and the individual preferences, $(R_i)_{i\in N}$, let $\Omega(e,(R_i)_{i\in N})$ be the set of all $(M,x)\in \mathcal{M}_e \times X$ such that x is an α-equilibrium of the game $(N,X,(S_i)_{i\in N},g,\ (R_i)_{i\in N})$. $\Omega(e,(R_i)_{i\in N})$ is then the set of all institution-outcome pairs (M,x) such that (M,x) is physically feasible and the outcome x can be achieved through the institution M if individual preferences over the social alternatives happen to be $(R_i)_{i\in N}$. Suppose the EE has an ethical ordering $\succcurlyeq$ over the set, H, of all pairs (M',x') such that, for some $e'=(X',(S_i')_{i\in N})$, $(M',x')\in \mathcal{M}_{e'}\times X'$. In that case, one can plausibly say that, for every $(R_i)_{i\in N}$, if the evaluator knows that the profile of individual preference orderings over Y is $(R_i)_{i\in N}$ and will remain fixed at $(R_i)_{i\in N}$, then her ethical ranking $\trianglerighteq_e$ over $\mathcal{M}_e$ will satisfy the following condition:

(3) for all $M,M'\in \mathcal{M}_e, M\trianglerighteq_e M'$ iff there exist $x,x'\in X$ such that $(M,x),(M',x')\in \Omega(e,(R_i)_{i\in N})$ and $(M,x)\succcurlyeq (M',\ x')$.

(3) essentially says that, given e and the fixed preference profile $(R_i)_{i\in N}$, the ethical evaluator will consider an institution $M \in \mathcal{M}_e$ to be ethically at least as good as another institution $M' \in \mathcal{M}_e$ if and only if the evaluator considers (M, x) to be ethically at least as good as (M', x'), where x and x' are the social outcomes which result from M and M', respectively, when $(R_i)_{i\in N}$ is the profile of individual preference orderings.

The difficulty of using (3) to establish the link between $\trianglerighteq_e$ and $\succcurlyeq$, however, is that typically we do not face the problem of evaluating institutions with a known and fixed profile of individual preference orderings over social alternatives. As we noted earlier, even when one knows the individuals' preferences, $(R_i)_{i\in N}$, over social alternatives at the time of evaluating alternative institutions, one cannot ignore the possibility of future changes in these preferences. One possible way of taking this into account is as follows.

Suppose we have a given $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$ and a given $M \in \mathcal{M}_e$. Further, suppose that we have a non-empty finite subset $\mathcal{D}$ of $\mathcal{P}^n$ with the following interpretation: $\mathcal{D}$ is the set of all $(R_i)_{i\in N} \in \mathcal{P}^n$, such that the evaluator believes $(R_i)_{i\in N}$ to be a preference profile which has some chance of materializing (the case where $\mathcal{D}$ is a singleton, as well as the case where $\mathcal{D} = \mathcal{P}^n$, is obviously a special case). Then consider the set, A_M, of all $x \in X$ such that, for some $(R_i)_{i\in N} \in \mathcal{D}$, $(M, x) \in \Omega(e, (R_i)_{i\in N})$. It is clear that A_M is the set of all possible outcomes that may arise from the institution depending on the preference profile in $\mathcal{D}$ that may materialize. Thus, as the evaluator sees it, the institution M is associated not necessarily with a unique outcome but with an uncertain prospect A_M, interpreted as the set of all possible outcomes that may arise if the institution M is chosen from $\mathcal{M}_e$. In light of this, how the evaluator will rank the different institutions in $\mathcal{M}_e$ will depend on whether the evaluator views the uncertainty involved as probabilistic or non-probabilistic and the type of decision rule for choice under uncertainty that she considers to be appropriate. We consider here one possible decision rule that the evaluator may follow. If the evaluator treats the uncertain prospects associated with alternative

institutions as non-probabilistic uncertain prospects and has strong aversion to risk, then she may follow the lexicographic maximin principle in ethically ranking the different institutions in $\mathcal{M}_e$. Let h be a real-valued function from H to the set (0,1] such that h represents $\succcurlyeq$ (recall that H is the set of all pairs (M', x'), such that, for some $e' = (X',\ (S_i')_{i\in N}) \in \mathcal{E}$, $M' \in \mathcal{M}_{e'}$, and $x' \in X'$, and $\succcurlyeq$ is the EE's ordering over H). For every $M \in \mathcal{M}_e$, let the elements of $\{M\} \times A_M$ be named as $a_1, a_2, \ldots$, and $a_{|A_M|}$ in such a way that $h(a_1) \leq h(a_2) \leq \cdots \leq h(a_{|A_M|})$. Let $v(M)$ denote the $|X|$-dimensional vector, $(h(a_1), h(a_2), \ldots, h(a_{|A_M|}), \overbrace{0, 0, \ldots, 0}^{|X|-|A_M| \text{ times}})$, of real numbers. Then we say that the EE's ordering $\trianglerighteq_e$ over $\mathcal{M}_e$ is based on the lexicographic maximin principle if and only if both (4) and (5) below hold for all $M, M' \in \mathcal{M}_e$:

(4) if for all $a \in \{M\} \times A_M$ and for all $a' \in \{M'\} \times A_{M'}$, $[a \sim a']$ then $[M \trianglerighteq_e M' \text{ and } M' \trianglerighteq_e M]$;

(5) if not $[a \sim a']$ for some $a \in \{M\} \times A_M$ and some $a' \in \{M'\} \times A_{M'}$, then $M \trianglerighteq_e M'$ iff $v(M) \geq_L v(M')$, where $\geq_L$ is the (usual) lexicographic ordering over the $|X|$-dimensional real space.

While the lexicographic maximin principle constitutes just one possible route that the EE may follow in ethically ranking alternative institutions corresponding to a given physical environment, it makes sense if EE's focus is on avoiding, as far as possible, the ethically bad institution-outcome pairs that may arise from the choice of an institution.

5. The ethical ranking of institutions when there are multiple equilibrium outcomes or no equilibrium outcome

In this section, we briefly indicate a possible way of extending our preceding analysis (see Section 4) of the ethical ranking of institutions to include cases where the game corresponding to an institution and a preference profile may have multiple α-equilibrium outcomes or no α-equilibrium outcome.

As before, for every $e = (X,\ (S_i)_{i\in N}) \in \mathcal{E}$, every $M = (N, X,\ (S_i)_{i\in N},\ (T_i)_{i\in N}, g) \in \mathcal{M}(e)$, and every profile, $(R_i)_{i\in N}$, of individual preferences, $G = (N, X,\ (S_i)_{i\in N},\ g)$ is the game form associated with M and $(G, (R_i)_{i\in N})$ is the game associated with G and $(R_i)_{i\in N}$. Also, as before, $\mathcal{D}$ is the set of all preference profiles $(R_i)_{i\in N}$ in $\mathcal{P}^n$, such that the evaluator believes $(R_i)_{i\in N}$ has any chance of materializing. We consider the ethical ranking, $\trianglerighteq_e$, over $\mathcal{M}_e$ when, for some $(R_i)_{i\in N} \in \mathcal{D}$, $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ may contain several outcomes or $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ may be empty. For every $e = (X,\ (S_i)_{i\in N}) \in \mathcal{E}$ and every $M = (N, X,\ (S_i)_{i\in N},\ (T_i)_{i\in N}, g) \in \mathcal{M}(e)$, let A_M be defined as follows:

(6) if $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ is non-empty for every $(R_i)_{i\in N} \in \mathcal{D}$, then A_M is the set of all $x \in X$, such that $x \in E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ for some $(R_i)_{i\in N} \in \mathcal{D}$;

and

(7) if $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ is empty for some $(R_i)_{i\in N} \in \mathcal{D}$, then $A_M = g(S)$.

Given the definition of A_M via (6) and (7), the specification of $\trianglerighteq_e$, over $\mathcal{M}_e$ is now analogous to the specification of $\trianglerighteq_e$, over $\mathcal{M}_e$ given by (4) and (5) in the preceding section.

The intuition underlying (7) is that, for every $(R_i)_{i\in N} \in \mathcal{D}$, if $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ is empty, then the EE considers $(\mathrm{G}, (R_i)_{i\in N})$ to be a game which may yield any element of $g(S)$ as the outcome. One can think of possible ways of refining this intuition. Consider one such refinement. Given the game $(\mathrm{G}, (R_i)_{i\in N})$ associated with M and $(R_i)_{i\in N}$, for every $i \in N$, let S_i^* be the set of undominated strategies of i, i.e., S_i^* is the set of all $s_i \in S_i$ such that there does not exist s_i' satisfying the following condition: $[g(s_i', s_{-i})\ R_i\ g(s_i, s_{-i})$ for all $s_{-i} \in S_i]$ and $[g(s_i', s_{-i})\ P_i\ g(s_i, s_{-i})$ for some $s_{-i} \in S_i]$. One can replace (7) by

(7′) if $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ is empty for some $(R_i)_{i\in N} \in \mathcal{D}$, then $A_M = g(\times_{i\in N} S_i^*)$.

The intuition underlying (7′) is that, for every $(R_i)_{i\in N} \in \mathcal{D}$, if $E_\alpha(\mathrm{G}, (R_i)_{i\in N})$ is empty, then the EE assumes that an outcome corresponding to any n-tuple of undominated strategies of the game $(\mathrm{G}, (R_i)_{i\in N})$, but no other social alternative, may emerge as the outcome of the game. One can define A_M via (6) and (7′) instead of defining it via (6) and (7) and then use (4) and (5) to generate $\trianglerighteq_e$ over $\mathcal{M}_e$.

6. Some properties of institutional frameworks

What a priori properties should one postulate for institutional frameworks? In this section, we discuss this issue at some length. Throughout this section, we shall assume that J is a given institutional framework and we shall consider various properties of J.

6.1. "Rationality" of institutional frameworks

In the literature on social choice, various rationality conditions, such as the Weak Axiom of revealed Preference and Sen's Condition α, are often postulated for social choice rules. As we have seen in Section 3.3, an institutional framework may yield different outcomes even if the set of feasible social alternatives and the profile of individual orderings remain fixed; therefore, it does not make much sense to impose the familiar rationality properties for social choice rules on institutional frameworks in general. But even institutional frameworks, which satisfy Invariance and hence can be interpreted as social choice rules, can violate some of the most basic rationality properties that one can formulate for an institutional framework. Consider the following property.

Non-reversal: Let $e = (X, (S_i)_{i\in N})$, $e' = (X', (S_i')_{i\in N}) \in \mathcal{E}$, where $X \subset X'$ and let $J(e) = M = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}\ g)$ and $J(e') = M' = (N, X',\ (S_i')_{i\in N}), (T_i')_{i\in N}),\ g')$. Let G and G' be the game forms associated with M and M', respectively. Then, for all $x, y \in X$ and for all $(R_{i\in N})_{i\in N} \in \mathcal{P}^n$, if x is an equilibrium outcome of the game $(G, (R_i)_{i\in N})$ and y is not, then y cannot be an equilibrium outcome of the game $(G', (R_i)_{i\in N})$ when x is not an equilibrium outcome of $(G', (R_i)_{i\in N})$.[8]

Note that Non-reversal is much weaker than the Weak Axiom of Revealed Preference for institutional frameworks and is one of the weakest rationality conditions that one can think of for an institutional framework.

[8] Note that we have stated Non-reversal for the general case where the relevant games are not assumed to have unique equilibrium outcomes; the statement will be somewhat simpler under our simplifying assumption that each game has a unique equilibrium outcome. Also, given the simplifying assumption of a unique equilibrium outcome for every relevant game, Non-reversal coincides with Sen' Condition α, reformulated for institutional frameworks.

Example 2: Let $N = \{1, 2\}$. Let $\mathcal{E}$ be such that, for all distinct $e = (X, (S_i)_{i \in N}), e' = (X', (S_i')_{i \in N}) \in \mathcal{E}, X \neq X'$. Then, it can be checked that every possible institutional framework will satisfy Invariance and, hence, will be a social choice rule.

Consider $e = (X, (S_i)_{i \in N}), e' = (X', (S_i')_{i \in N}) \in \mathcal{E}$ such that $X = \{x, y, z\}$ and $X' = \{y, z\}$. Let G and G' be the game forms associated with $J(e)$ and $J(e')$, respectively. Let the profile of preference orderings be denoted by (R_1, R_2). Suppose the two games $(G, (R_1, R_2))$ and $(G', (R_1, R_2))$ are as follows in Table 2 (as in Example 1, we represent the two orderings by specifying for each social alternative a pair of numbers indicating the payoffs of individuals 1 and 2, respectively, corresponding to that alternative).

Suppose Nash equilibrium is the appropriate notion of equilibrium. The game $(G, (R_1, R_2))$ has the unique Nash equilibrium (s_1, s_2') with the outcome y. The game $(G', (R_1, R_2))$ has two Nash equilibriia, (s_1', s_2') and (s_1'', s_2'), but a unique equilibrium outcome, z. It is clear that J violates Non-reversal.

Table 2

$(\boldsymbol{G}, (\boldsymbol{R_1}, \boldsymbol{R_2}))$

		Individual 2	
		s_2	s_2'
	s_1	x	y
		(6, 2)	(8, 3)
Individual 1	s_1'	y	z
		(8, 3)	(4, 6)
	s_1''	z	z
		(4, 6)	(4, 6))

$(\boldsymbol{G'}, (\boldsymbol{R_1}, \boldsymbol{R_2}))$

		Individual 2	
		s_2	s_2'
	s_1'	y	z
		(8, 3)	(4, 6)
Individual 1	s_1''	z	z
		(4, 6)	(4, 6)

6.2. Distribution of power

One of the considerations which arise in any discussion of the intrinsic moral value attached to an institution is what may be called the distribution of power embedded in it. It is true that there may be difference of opinion about the exact notion of power to be used in this context and possibly greater difference of opinion about whether a particular distribution of power is better than another distribution of power even when people agree about the notion of power to be used. There are, however, some minimal properties, closely linked to the notion of power, of an institutional framework, about which there is likely to be consensus. We now introduce a few such properties of an institutional framework J. Intuitively, these properties can be partitioned into two groups: (i) properties which postulate that individuals or groups of individuals should not have certain types of power, and (ii) properties which postulate that individuals or groups of individuals should have certain types of power.

6.2.1. Properties which rule out certain types of power for individuals or groups of individuals

Non-dictatorship (ND): There does not exist $i \in N$ such that:

(8) for all $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, if $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, then, for all $x \in X$, there exists $t_i \in T_i$ such that $g(t_i, t_{-i}) = x$ for every $t_{-i} \in T_{-i}$.

No one with universal veto (NOUV): There does not exist $i \in N$ such that:

(9) for all $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, if $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, then, for all $x \in X$, there exists $t_i \in T_i$ such that $g(t_i, t_{-i}) \neq x$ for every $t_{-i} \in T_{-i}$.

If J violates Non-dictatorship, then there exists an individual who has the ability to ensure the emergence of any feasible social alternative as the social outcome by suitably choosing one of her permissible strategies, irrespective of what permissible strategies the other individuals may adopt. Similarly, if J violates NOUV, then there exists an individual such that, for every feasible social alternative x, the individual can prevent x from emerging as the social outcome by suitably choosing one of her permissible strategies, regardless what permissible strategies the other individuals may choose to adopt.

It may be noted that NOUV is a much stronger restriction on an institutional framework than ND. ND has the flavor of Arrow's condition of non-dictatorship for his social welfare function. The following condition, AO, which is reminiscent of a well-known condition due to Gibbard (1969), is stronger than both ND and NOUV.

Absence of oligarchy (AO): There does not exist a unique non-empty proper subset, N', of N such that N' satisfies (10) and (11) below:

(10) for all $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, if $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, then, for all $x \in X$, there exists $t_{N'} \in T_{N'}$ such that $g(t_{N'}, t_{N-N'}) = x$ for all $t_{N-N'} \in T_{N-N'}$;

(11) for every $i \in N'$ and for all $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, if $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, then, for all $x \in X$, there exists $t_i \in T_i$ such that $g(t_i, t_{-i}) \neq x$ for all $t_{-i} \in T_{-i}$.

6.2.2. Properties stipulating certain types of power for individuals

In contrast to the ND and NOUV, which stipulate that the institutional framework should not legitimize excessive concentration of power in the hands of any single individual, our next property rules out the existence of any individual who never has a say in social decisions in the sense that, for every physical environment e in $\mathcal{E}$, given $J(e)$, the individual's choice from her set of permissible strategies has no effect on the social outcome no matter what permissible strategies other individuals may adopt.

Absence of marginalized individuals (AMI): There does not exist $i \in N$ satisfying the following condition:

(12) for all $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, if $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, then, for all $t_{-i} \in T_{-i}$, $g(t_i, t_{-i}) = g(t_i', t_{-i})$ for all $t_i, t_i' \in T_i$.

Our next property incorporates the idea that institutions must give some minimal freedom of action to the individuals and must not dictate the social outcome.

Minimal freedom (MF): For all $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, if [$|S_i| > 1$ for all $i \in N$], $|X| > 1$ and $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$, then [$|T_i| > 1$ for all $i \in N$] and $|g(T)| > 1$.

One of the issues which inevitably arise in evaluating institutions is the balance of power between an individual on the one hand and the state or the rest of the society on the other. John Stuart Mill's (1859) famous notion of a protected private sphere of actions for every individual was intended to carve out areas of an individual's life that should be free from any interference from the state or the society in general. Sen's (1970) path-breaking contribution was the first attempt to provide, in social choice theory, a formal formulation of Mill's notion of an individual's right to liberty in private affairs. For reasons extensively discussed in the literature (see, among others, Nozick 1974, Sugden 1985, Gaertner, Pattanaik and Suzumura 1992, and Pattanaik 1996), we believe that, in articulating Mill's intuition regarding an individual's right to liberty in private affairs, the notion of an institution provides a more flexible and less problematic formal framework than the standard analytical apparatus of the theory of social choice, which was used by Sen (1970). Our next property seeks to capture an important feature of rights to liberty in private affairs in the context of institutions.

Protection of personal spheres (PPS): There exist $e \equiv (X, (S_i)_{i\in N}) \in \mathcal{E}$ and an institution $M \in \mathcal{M}_e$ such that $M = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g) \in J(e)$ and, for every $i \in N$, there exists an integer m_i ($m_i \geq 2$) such that, for some partition of $g(T)$ into non-empty subsets $A_{i1}, A_{i2}, \ldots, A_{im_i}$ and some partition of T_i into non-empty subsets T_{i1} and T_{im_i}, the following holds:

(13) for all $k \in \{1, 2, \dots, m_i\}$, if $t_i \in T_{ik}$, then $g(t_i, t_{-i}) \in A_{ik}$ for all $t_{-i} \in T_{-i}$.

To see the intuitive link between an individual's right to liberty in private affairs and PPS, suppose, given a physical environment e, it is possible for every individual to choose to be a Buddhist or a Hindu, or neither a Buddhist nor a Hindu. Suppose the institution $M = J(e)$ recognizes that a person's religion is her private affair and respects every individual's right to liberty in her private affairs. Such liberty for individual i will then allow us to partition T_i into non-empty sets T_{i1} and T_{i2} and T_{i3} and to partition $g(T)$ into non-empty sets A_{i1}, A_{i2}, and A_{i3} such that T_{i1} is the set of all permissible strategies in which i chooses to be a Buddhist, T_{i2} is the set of all permissible strategies in which i chooses to be a Hindu, T_{i3} is the set of all permissible strategies in which i chooses to be neither a Hindu nor a Buddhist, A_{i1} is the set of all social alternatives in $g(T)$ in which i is a Buddhist, A_{i2} is the set of all social alternatives in which i is a Hindu, and, finally, A_{i3} is the set of all social alternatives in which i is neither a Buddhist nor a Hindu. For all $k \in \{1, 2, 3\}$, if i chooses a strategy from T_{ik}, then so long as other individuals adopt permissible strategies, the social outcome will lie in A_{ik}.

6.3. Pareto optimality of equilibrium outcomes

The properties of an institutional framework that we discussed in in Section 6.2 above, have no reference to the individual preferences, the equilibria which emerge from the institutions given the individual preferences over social alternatives, and, finally, the outcomes in these equilibria. Ethical assessment of institutions, however, typically involves consideration of such equilibria and the outcomes corresponding to these equilibria.

First, it may be useful to note that, in some ways, even the intrinsic ethical appeal of an institution may depend on the equilibria of the game associated with the institutions corresponding to the different preference profiles that may arise. An example may help illustrate this point. Suppose an institution gives every individual the right to criticize policies of the government by making it permissible for her to

voice her criticism publicly and by making it impermissible for anybody to penalize any person for criticizing such policies. The institution may even specify some punishment for anyone penalizing or trying to penalize a person for exercising this right. But how do we know to what extent the institution will be able to ensure that this right of individuals will be respected without considering what will be the equilibria of the relevant games corresponding to the different preference profiles that may materialize? Thus, given $e \in \mathcal{E}$, each of two institutions $M = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$ and $M' = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g')$ in $\mathcal{M}_e$ may give this right to all individuals, but, for some preference profile $(R_i)_{i\in N}$, the right of more individuals may be violated in the equilibrium of the game $(N, X, (S_i)_{i\in N}, g, (R_i)_{i\in N})$ than in the equilibrium of the game $(N, X, (S_i)_{i\in N}, g', (R_i)_{i\in N})$ because in the equilibrium of the former game more individuals who criticize the policies of the government will be penalized by other individuals. For example, this may happen if, compared to the outcome function g', the outcome function g stipulates a much more lenient system of punishment for people who adopt impermissible strategies. Thus, even when we are concerned exclusively with the intrinsic value of institutions, it may not be possible to rank two institutions without considering what happens in the equilibria of the games associated with different institutions and preference profiles.

Of course, the instrumental value of an institution cannot possibly be taken into account without considering the equilibria and social outcomes which result from them corresponding to the different possible preference profiles. A familiar restriction usually postulated for outcomes of institutions is that of Pareto optimality implicit in the following property of an institutional framework, J.

Pareto Principle (PP): Let $e \equiv (X, (S_i)_{i\in N})$ be any element of $\mathcal{E}$ and let $J(e) = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g)$. For all $(R_i)_{i\in N} \in \mathcal{D}$, if x is the equilibrium outcome of the game $(N, X, (S_i)_{i\in N}, g, (R_i)_{i\in N})$, then there does not exist $y \in g(S)$ such that yP_ix for all $i \in N$.

Note that the Pareto Principle has been formulated with reference to every profile belonging to $\mathcal{D}$ (i.e., the set of all preference profiles which, according to the EE's beliefs, may arise) rather than the set, $\mathcal{P}^n$,

of all logically possible preference profiles. This is because, by assumption, the EE believes that no preference profile in $\mathcal{P}^n - \mathcal{D}$ will ever materialize and, hence, the preference profiles in $\mathcal{P}^n - \mathcal{D}$ are of no consequence for the EE's evaluation of institutions.

While PP is a very familiar principle, the intuition underlying its use in the context of our analysis of an EE's evaluation of institutions needs some scrutiny. Consider $e = (X, (S_i)_{i\in N}) \in \mathcal{E}$, $M = (N, X, (S_i)_{i\in N}, (T_i)_{i\in N}, g) \in \mathcal{M}_e$, $(R_i)_{i\in N} \in \mathcal{D}$, $G = (N, X, (S_i)_{i\in N}, g, (R_i)_{i\in N})$, and $x \in X$ such that x is the equilibrium outcome of the game G. Suppose there exists $y \in g(S)$ such that yP_ix for all $i \in N$. Should this, by itself, be necessarily a matter of ethical concern for the EE? The answer to this will depend on the intuitive content of the preference orderings R_i ($i \in N$). Suppose an individual's preference ordering over social alternatives is interpreted as Arrow (1951) interprets it, i.e., as the product of "whatever standards [the individual] deems relevant" (Arrow 1951, p. 17), so that it may reflect the individual's "values" as well as his "tastes" (see Arrow, 1951, 1963, p. 18). Suppose the personal well-being of every individual, as assessed by that individual as well as by the EE, is higher in x than in y, but every individual prefer y to x because of some ethical commitment ("resources must be devoted to prevent tigers from becoming extinct since tigers have as much right to exist as human beings"), which has no relation to their personal well-being[9]. Then it is not clear that the EE should necessarily be concerned about the emergence of x as the outcome of the game G when yP_ix for all $i \in N$. It is intuitively compelling that, in ethically assessing the outcome resulting from the game G, the EE should take into account personal well-being of all individuals but there is no compelling reason for the EE to take into account those individuals' ethical commitments in addition to their personal well-being; the

[9] We are not claiming that a person's ethical concerns can never be linked to her well-being: if a person's concern about the survival of tigers is accompanied by deep empathy for tigers, then what happens to tigers can indeed have important impact on her personal well-being. We are, however, making the much weaker claim that it is possible for an individual to have strong ethical concerns about the extinction of tigers without her having much empathy for tigers and this can lead her to prefer y, where much resource is being devoted to the preservation of tigers, to x, where no resource is devoted to the survival of tigers, though the individual's personal well-being in x is higher than her personal well-being in y. Regarding the assessment of an individual's well-being on the basis of her preferences, Broome (2008, p. 12) observes, "if a person's well-being is judged by her preferences, things will go particularly badly wrong when her preferences are not self-interested but based on her ethical beliefs."

values that the EE will use (and should use) in ethically assessing the outcomes are her own ethical values. In Arrow's problem of aggregating individual preferences so as to reach a social decision, an individual's preferences reflect her opinion, which is assumed to be given and is based on whatever mixture of "tastes" and "values" the individual may use to form that opinion. In the case of the EE's evaluations of institutions, and social alternatives together with institutions, the EE is expected to give her own justification for the values underlying her own evaluations and, in that context, the content of individual preferences is crucial in determining how those preferences should or should not be used for the EE's evaluation.

7. Concluding remarks

In this paper, we have tried to explore the structure of an individual's ethical evaluations of institutions. In the process, we have extended in several ways Pattanaik and Suzumura's (1996) earlier analysis of the same problem in the specific context of individual rights and have discussed several properties of institutional frameworks and their plausibility. Instead of summarizing the main points in the preceding sections, however,we would like to highlight in these concluding remarks one important lacuna in our analysis. Throughout our analysis we have assumed that the profiles of individual preference orderings are given exogenously. Yet one of the crucial features of liberal democratic institutions is that they permit free debates and discussion of public affairs, which often lead to changes in people's preferences over social alternatives. In fact, in a liberal democracy, individuals often devote considerable effort and resource to such debates. Our formal framework cannot accommodate this institutional feature since the preferences which figure in the relevant games in our framework are assumed to be exogenously given. In some ways, this is a limitation which our analysis shares with much of the literature in the theory of social choice and welfare economics. It is nevertheless a problem that our account of social choice through institutions cannot accommodate one of the central and most vital features of liberal democracies, namely, the free debates and discussions that the individuals engage in to influence each other's preferences over social alternatives. Further research is needed to extend the analysis so as to remove this limitation.

References

Arrow, K.J. 1951; 1963 2nd edition, Social Choice and Individual Value, New York: Wiley.

Bergson, A. 1954, On the concept of social welfare, *Quarterly Journal of Economics* 68: 233-52.

Broome, J. 2008, Why economics needs ethical theory. In *Arguments for a Better World*, Vol. I, eds. K. Basu and R. Kanbur, 7–14. Oxford: Oxford University Press.

Buchanan, J. 1986, Liberty, Market and the State, Brighton: Wheatsheaf Books.

Friedman, M. 1962, Capitalism and Freedom, Chicago: University of Chicago Press.

Gaertner, W., P. K. Pattanaik and K. Suzumura 1992, Individual rights revisited. Economica 59: 161–177.

Gibbard, A. 1969, Social choice and the Arrow conditions, mimeo, Harvard University. [Published in Economics and Philosophy 30(3): 269-284, 2014.]

Little, I. M. D. 1952, Social choice and individual values. *Journal of Political Economy* 60: 422-432.

Mill, J.S. 1859, On Liberty. Reprinted in M. Warnock (ed.), John Stuart Mill, Utilitarianism, On Liberty, Essays on Bentham, together with Selected Writings of Jeremy Bentham and John Austin. New York: Penguin Books, 1962.

Nozick, R. 1974, *Anarchy, State and Utopia*. Oxford: Blackwell.

Pattanaik, P.K. 1996, The liberal paradox: some interpretations when rights are represented as game forms, *Analyse & Kritik* (special issue on the Liberal Paradox) 18: 38--53.

Pattanaik, P.K. 2005, Little and Bergson on Arrow's concept of social welfare, *Social Choice and Welfare* 25: 369-379.

Pattanaik, P. K. and K. Suzumura, 1996, Individual rights and social evaluation: a conceptual framework. *Oxford Economic Papers* 48: 194–212.

Pattanaik, P.K. and Y. Xu, 2014, The ethical bases of public policies: a conceptual framework. *Economics and Philosophy* 30(2): 175-194.

Sen, A. K. 1977, Social choice theory: a re-examination. *Econometrica* 45: 53-89.

Sen, A. K. 1970, The impossibility of a Paretian liberal. *Journal of Political Economy* 78: 152–157.

Sen, A. K. 1993, Markets and freedoms, *Oxford Economic Papers* 45: 519-541.

Sugden, R. 1985, Liberty, preference, and choice. *Economics and Philosophy* 1: 185–205.