CBESS Discussion Paper 09-15

The representation of alienable and inalienable rights:
Games in transition function form

by Ben McQuillin and Robert Sugden*

*University of East Anglia

27th August 2009

University of East Anglia

CBESS

Centre for
Behavioural and
Experimental
Social
Science

Abstract

We propose a new type of cooperative game - a game in transition function (TF) form – as a means of representing social decision making procedures that is suitable for the analysis of rights. The TF form is a generalisation of the e¤ectivity function (EF) form, and in particular it tells us (where the EF form does not) about the alienability of a right. We describe procedures for generating a (unique) EF game from a TF game, and for generating a (non-unique) TF game from an EF game. We make some specific proposals about the representation of rights as properties of TF games and comment on some implications about the relationship between rights and Pareto efficiency.

The representation of alienable and inalienable rights:

Games in transition function form.[1]

Ben McQuillin[2]   Robert Sugden[3]

27th August 2009

[2]School of Economics, University of East Anglia, Norwich NR4 7TJ, UK. *b.mcquillin@uea.ac.uk*
[3]School of Economics, University of East Anglia, Norwich NR4 7TJ, UK. *r.sugden@uea.ac.uk*

Abstract

We propose a new type of cooperative game - a *game in transition function (TF) form* - as a means of representing social decision making procedures that is suitable for the analysis of rights. The TF form is a generalisation of the effectivity function (EF) form, and in particular it tells us (where the EF form does not) about the alienability of a right. We describe procedures for generating a (unique) EF game from a TF game, and for generating a (non-unique) TF game from an EF game. We make some specific proposals about the representation of rights as properties of TF games and comment on some implications about the relationship between rights and Pareto efficiency.

1 Introduction

This paper revisits a classic issue in social choice theory: how should the concept of a right (or of individual liberty, or of a 'protected sphere' of individual choice) be represented? Our intention is not to rehearse old controversies, but to propose a new solution to this old problem.

The idea that social choice theory needs to take account of rights originates in Sen's (1970) famous proof of the 'impossibility of a Paretian liberal'. In the two decades after the publication of that paper, there was vigorous debate about the appropriateness of Sen's *preference-based* approach to the formulation of rights. Sen's condition of 'minimal liberalism' represents a right as a relationship between individual preferences and social preferences. Specifically, let x and y be two (social) outcomes – that is, complete descriptions of states of affairs for society as a whole. The idea that $\{x, y\}$ belongs to the protected sphere of some individual i is represented by the requirement that if i has a strict preference between x and y, the social preference between x and y is the same as i's. An alternative *procedural* approach, proposed in slightly different forms by Nozick (1974: 164-166), Sugden (1978, 1985), Gärdenfors (1981) and Gaertner, Pattanaik and Suzumura (1992), represents a right as a relationship between the actions that individuals take within a social decision-making procedure and the outcomes that result from the workings of that procedure. The arguments for and against these rival approaches are well known, and we will not repeat them here.[1] Our paper is premised on the proposition that the procedural approach is a useful way of thinking about rights. Our concern is with the implementation of that approach.

The main features of the procedural approach can be summarised in the following way. Social decision-making is understood as a rule-governed procedure specified in terms of the actions legitimately available to the players and the consequences that result from combinations of such actions; players' preferences play no part in the specification. 'Legitimate' here means simply 'in conformity with the rules' (there is no presumption that the rules are morally justified), and there is no analysis of *enforcement*: we do not specify consequences

---

[1]The arguments for the procedural approach are set out in the papers we have cited. For Sen's responses to these arguments, see Sen (1983, 1992).

resulting from illegitimate actions. There is, in effect, an assumption that individuals conform to the rules. A (putative) right is construed as a condition that a social decision-making procedure may or may not satisfy.[2]

If the procedural approach is to be formalised, a first step is to fix a game-theoretic concept that is to be used to represent social decision-making procedures. Advocates of the procedural approach have focused on two such concepts – the *game form* and the *effectivity function*.[3]

The concept of a game form, first used by Gibbard (1973), belongs to non-cooperative game theory. Essentially, a game form is a non-cooperative game from which all information about players' payoffs has been removed; the consequences of different strategy choices are represented as outcomes rather than as vectors of payoffs. In the context of rights, this feature of game forms is very useful, since it allows social decision-making procedures to be represented in terms of individual actions and outcomes, without making any assumptions about preferences. However, the game form representation of social decision-making procedures has a serious disadvantage: it can contain a large amount of information which, for the analysis of rights, is redundant. For example, consider the case of two individuals bargaining over the price at which one will sell some indivisible good to the other. Suppose we want to represent the right of free contract. Intuitively, we want to say that no trade should take place without the consent of both parties, and that if both agree to trade at any particular price, they should not be prevented from doing so. But now suppose we try to represent this right as a condition to be imposed on game forms. If we are to represent any particular bargaining process by a game form, we will need to specify an entire bargain-

---

[2]Although the procedural approach has most usually been used to represent principles of negative liberty, it can also be used to represent 'positive' or 'effective' freedom. (For example, the right to freedom from hunger might be represented as the condition that each individual has the power to ensure that the outcome of the game is one in which he is not hungry.)

[3]In some analyses, the effectivity function is supplemented by the waiver function (Deb, 1994). The waiver function describes the powers of individuals and coalitions to confer choice opportunities on others. Elsewhere, similar information is preserved by allowing effectivity functions such that a coalition is effective for one set but not for all its supersets (van Hees, 1995). But in contrast to these approaches, our primary concern is with the powers of individuals and coalitions to determine the outcomes of the social decision-making process. We will argue that the effectivity function fails to capture certain features of rights that are relevant to that concern; these features are essentially different from those that are represented by the waiver function. The theoretical concept we propose – that of a 'game in transition function form' – contains information about powers of waiver, but we do not discuss these powers explicitly.

ing 'protocol' – that is, an exhaustive description of the possible offers, counter-offers and messages that the players can make in the process of negotiation. But since many different protocols respect the right of free contract, this level of detail is unhelpful. We seem to need a coarser theoretical framework which takes account only of those features of game forms that are relevant for the analysis of rights.

The concept of an effectivity function (Moulin and Peleg, 1982; Peleg, 1998) has been widely used for exactly this purpose. Gärdenfors (1981) uses essentially this concept in his version of the procedural approach.[4] Sugden (1985) uses the same idea more informally when discussing whether particular game forms respect particular rights. An effectivity function assigns a set of sets of outcomes to each set (or *coalition*) of players. If some set X of outcomes is an element of the set assigned to coalition A, we will say that A has effectivity for X. To say that A has effectivity for X is to say that there is some array of strategies, one for each member of A, such that if these strategies are chosen, the outcome will be some element of X. In discussions of rights, effectivity functions have often been used to represent properties of game forms.

We will argue that, while game forms contain *too much* information for them to work well in representing rights, effectivity functions contain *too little*. An effectivity function specifies what each coalition of players can guarantee by concerted action. But it leaves open significant questions about what is involved when a coalition acts 'in concert'. Most importantly, it leaves open the question of whether, once some coalition has so acted, its individual members remain free to act independently of that coalition, either as single agents or as members of other coalitions. We will argue that an effectivity function therefore closes down the important (and closely related) distinctions between rights that are *alienable* and rights that are *inalienable* and between rights that are and are not *co-exercisable*.

To say that a right is 'inalienable' is to say that an individual, having been endowed with it, cannot divest herself of it. Conversely, and subject to the satisfaction of relevant conditions, a person who holds an alienable right may choose to transfer it to other individuals or collectives. Suppose that some right is initially invested in person i; this right gives

---

[4]Gärdenfors uses the term 'rights system' for this concept. In this paper, we will use 'rights system' more informally, to refer to rights in the real world.

certain powers to the coalition $\{i\}$. If the right is inalienable, those powers will continue to exist, irrespective of the concerted actions of any coalition that contains i. But if instead it is alienable, the concerted action of some such coalition may transfer the right from i, eliminating those powers.

The idea that certain fundamental rights are inalienable has had great significance in political discourse. For example, the signatories of the US Declaration of Independence held it to be self-evident, not only that all men are endowed with the rights to life, liberty and the pursuit of happiness, but also that those rights are 'unalienable'. In making the latter claim, they rejected the proposition that Americans' natural rights to liberty were now vested in the British sovereign. Their loyalist opponents might have invoked Thomas Hobbes's (1651/1962, Chapters 14, 18) argument that individuals' natural rights to liberty are alienable, and are surrendered on the formation of a 'commonwealth' with a sovereign power.

The distinction between alienable and inalienable rights has been central to some influential discussions of Sen's Paretian liberal problem. Intuitively, it is natural to associate Pareto inefficiency with barriers to exchange, and so to suppose that Sen's problem would disappear if rights were tradable. This intuition is explored by Gibbard (1974), who formulates a 'Pareto-consistent libertarian claim'. Gibbard's idea is that 'a person's rights are his to use or bargain away as he sees fit'; his libertarian claim 'permits at least those bargains to which everyone would agree' (p. 397). Gärdenfors (1981) develops this concept of alienability in a game-theoretic framework. We will argue that the distinction between alienable and inalienable rights cannot be adequately expressed in the language of effectivity functions.

The 'co-exercisability' question is crucial for the analysis of what, for many liberals, is a fundamental property of a free society. Adam Smith (1776/ 1976) and Hayek (1960) are among those who have argued that an individual's freedom in a liberal social order consists, not so much in what she can do in isolation, as in a multiplicity of independent possibilities for voluntary interactions with others.[5] If freedom is conceived in this way and if rights

---

[5]Sugden (2009) develops an analysis of 'opportunity as mutual advantage' which is intended to represent

are to safeguard it, the separate rights of overlapping sets of individuals should be capable of being exercised independently of one another. For example, the right of persons i and j to attend a political meeting together should be exercisable independently of the right of persons i and k to engage in market exchange. When i and j jointly exercise their right of freedom of assembly, they do not thereby constitute themselves as a collective entity with the power to veto voluntary exchanges between i and k. Similarly, if i and k exercise their right to make such an exchange, k should remain free to negotiate further exchanges with l without requiring i's consent. We will show that the language of effectivity functions cannot express the fundamental idea that the exercise of one right does not compromise the exercise of another.

Our response to these problems is to propose a new theoretical framework within which social decision-making procedures can be described and analysed. Formally, we define a new type of cooperative game – a *game in transition function form.*

To understand the intuition behind this proposal, it is important to recognise that an effectivity function can be interpreted as a concept in cooperative game theory (rather than as a property of a non-cooperative game form). Consider the concept of a *characteristic function,* as used in cooperative game theory. A characteristic function assigns to each coalition of players a set of *utility imputations* – that is, a set of utility vectors, such that each vector specifies a utility level for each member of the coalition. To say that some imputation u_A is in the set assigned to some coalition A is to say that A has the collective power to ensure that each member achieves at least the level of utility specified by u_A. One might say that an effectivity function differs from a characteristic function in the same way that a game form differs from a game: in each case, the difference is the substitution of outcomes for utility vectors. In cooperative game theory, it is a standard modelling strategy to represent real-world bargaining problems as *games in characteristic function form* – that is, as games that are fully described by their characteristic functions. Analogously, a 'game in effectivity function form' might be used to represent a system of rights, without presupposing that its effectivity function was derived from some game form.

this form of freedom.

5

Our concept of a game in transition function form generalises that of a game in effectivity function form. More precisely: every game in transition function form induces a unique effectivity function. For every effectivity function there is at least one game in transition function form which induces it, but there may be more than one such game. Intuitively, the transition function form contains information that cannot be retrieved from the effectivity function. This additional information can be used (among other purposes) to answer questions about whether the formation of one coalition constrains the opportunities of its members to participate in other coalitions.

We must stress that we are proposing a new theoretical framework within the general tradition of cooperative game theory. A game in transition function form is not a compact description of relevant properties of some more basic non-cooperative game form. We treat cooperative game theory as a free-standing mode of analysis. We do not share the view of some theorists that the concepts of cooperative game theory are legitimate only to the extent that they are reducible to those of non-cooperative game theory – that non-cooperative game theory describes strategic interaction as it 'really' is, and cooperative game theory is merely a convenient short cut. For us, the two branches of game theory provide *alternative* ways of representing a world that is much more complex than any theoretical model. Our claim is that the concept of a game in transition function form is useful as a means of representing systems of rights that exist in the real world, or which feature in real political debate. Its domain of applicability is not restricted to any particular class of rights, but extends to any analysis of rights that follows the procedural approach.[6]

2   What effectivity functions do *not* tell us about rights

We begin by defending our claim that certain significant features of rights systems are not represented in effectivity functions. We do this by considering three stylised examples of social decision-making problems. In each case, we compare two alternative specifications

---

[6]The present paper focuses on the game theoretic representation of rights, but we also hope that the concept of a game in transition function will prove to be a useful addition to cooperative game theory generally.

of rights. In a normative perspective which values individual liberty, these specifications are significantly different; but they are described by the same effectivity function. Our descriptions of the three problems are deliberately informal. Recall that our concern is with how real-world rights systems are best represented in social choice theory. Thus, as far as possible, our examples are described in ordinary language, without presupposing any particular modelling framework.

Example 1 *The Proselytising Mineowner*

This example is loosely based on a passage in Hayek's *Constitution of Liberty* (1960, pp. 135-137). Hayek is presenting a conception of liberty as freedom from coercion. He asks whether an individual i is subject to coercion when a necessity of life can be bought only from a single seller j, or when i's only source of subsistence is the sale of his labour and j is the only buyer. Hayek's answer to this question is not entirely clear, but he seems to want to distinguish between the case in which j merely sets a non-discriminating profit-maximising price, to which i must adapt as best he can, and the case in which j uses her market power as a means of controlling i's actions in other spheres of life. Hayek argues that in the second case, i is subject to coercion:

> There are, undeniably, occasions when the condition of employment creates opportunity for true coercion. In periods of acute unemployment the threat of dismissal may be used to enforce actions other than those originally contracted for. And in conditions such as those of a mining town the manager may well exercise an arbitrary and capricious tyranny over a man to whom he has taken a dislike. (pp. 136-137)

Whatever one makes of this answer, Hayek has surely identified a problem for any normative theory of liberty. Does the right of free contract allow one person to make her consent to trade with another person conditional on the latter's acting in specific ways in areas of life that are private to him? For example, may the owner of Hayek's mine make it a condition of employment that her workers attend the church that she deems right for them?

In our example, there are just two actors, the employer (e) and the worker (w). For the purposes of our discussion, outcomes differ on just two dimensions - whether or not w works for (and is paid by) e, and whether of not w attends the services of a particular church. In the status quo position, w and e have no contractual obligations to one another; w is unemployed and does not attend church. There are four relevant outcomes: unemployment and non-attendance (UN), unemployment and attendance (UA), employment and non-attendance (EN), and employment and attendance (EA). Now consider two alternative specifications of rights.

Specification 1.1

1. If w is unemployed, he enters employment if and only if w and e both choose that this should be the case; this employment contract, if made, is annulled if and only if both agents so choose.

2. If w is unemployed, he attends church if and only if he chooses to do so.

3. If w is employed, the employment contract must take one of two forms: either it specifies that w attends church (A), or it specifies that he does not (N). Each form is operational if and only if both agents consent. After the employment contract has been made, switches between the A and N forms can be made if and only if both agents so choose.

4. The status quo is UN.

Specification 1.2

1. As in clause (1) of Specification 1.2.

2. Whether w is employed or unemployed, he attends church if and only if he chooses to do so.

3. The status quo is UN.

As far as the employment relation is concerned, both specifications respect the right of free contract as this is normally understood: w sells his labour to e if and only if both agents

consent to the exchange. Further, both specifications respect w's freedom of worship *if he is not employed*. However, they differ with respect to the terms on which the two rights can be exercised together. In Specification 1.1, w's right to freedom of worship is conditional on his using (or not using) the right of free contract in a particular way. If he enters into an employment contract with e, his previously individual right to freedom of worship is transformed into a right that can be exercised only jointly by the two agents together. To put this another way, in entering the employment relation, w alienates his individual right to freedom of worship. In Specification 1.2, w's right to freedom of worship is unconditional; it remains operative irrespective of how he chooses to use the right of free contract.

One does not have to take any firm position about which of the two specifications is 'more liberal' to recognise the normative significance of the question of whether the right to freedom of worship is alienable or inalienable. Thus, we claim, an adequate formal representation of rights should distinguish between Specifications 1.1 and 1.2. But, in the case of our story, the effectivity function does not. Both specifications imply *exactly the same* effectivity function. This function is fully described by the following properties:[7] $\{w\}$ has effectivity for $\{UN\}$ and $\{UA\}$ (i.e. w has the power to remain unemployed, and if he remains unemployed, to attend church or not); $\{e\}$ has effectivity for $\{UN, UA\}$ (i.e. e has the power not to hire w); and $\{w, e\}$ has effectivity for $\{EN\}$ and $\{EA\}$ (i.e. w and e have the joint power to bring about any outcome).

Example 2 *The Subjection of Women*

In our second story, there are two individuals, a man (m) and a woman (f). The woman has an inheritance which can be invested in two ways, which we will call cautious and rash. At the start of the story, m and f are single and f's inheritance is invested cautiously (an outcome denoted by SC). There are three other possible outcomes in which, respectively, m and f are single and the inheritance is invested rashly (SR), they are married and

---

[7] Throughout this Section, we take it as given that effectivity functions satisfy three standard properties of internal consistency, which we will define more formally in Section 4.1. The first is that if some coalition A has effectivity for some set X of outcomes, every superset of A also has effectivity for X. The second is that if A has effectivity for X, it also has effectivity for every superset of X. The third is that no two disjoint coalitions have effectivity with respect to the same set of outcomes (other than the set which contains all possible outcomes).

the inheritance is invested cautiously (MC), and they are married and the inheritance is invested rashly (MR). At the time John Stuart Mill wrote *The Subjection of Women* (1869/ 1988), English laws of property and marriage effectively implied the following specification of rights for this case:[8]

Specification 2.1

1. If m and f are single, they marry if and only if they both choose to do so.

2. If m and f are married, they remain married irrespective of their subsequent choices.

3. If f is single, she chooses whether her inheritance is invested cautiously or rashly; if m and f are married, m chooses whether f's inheritance is invested cautiously or rashly.

4. The status quo is SC.

Mill argued that the legal status of a married woman, as laid down by provisions such as clause (3), was effectively that of a slave (pp. 31-32). Among his proposed reforms was the following rule: 'Whatever would be the husband's or wife's if they were not married, should be under their exclusive control during marriage' (p. 50).[9] This reform would give:

Specification 2.2

1–2. As in clauses (1)–(2) of Specification 2.1.

3. Whether f is single or married, she chooses whether her inheritance is invested cautiously or rashly.

4. The status quo is SC.

Despite the normative significance of Mill's proposal, both specifications imply the same effectivity function for our story. The effectivity function implied by both specifications is

---

[8]At the time, legal divorce was possible but only under very stringent conditions and at considerable expense. We simplify slightly by treating marriages as indissoluble.

[9]Mill adds the clause that this rule 'need not interfere with the power to tie up property by settlement, in order to preserve it for children' (p. 50).

defined by the following properties: $\{f\}$ has effectivity for $\{SC\}$ and $\{SR\}$ (i.e. f has the power to remain single, and if she remains single, can choose how her inheritance is invested); $\{m\}$ has effectivity for $\{SC, SR\}$ (i.e. m has the power to remain single); and $\{m, f\}$ has effectivity for $\{MC\}$ and $\{MR\}$ (i.e. m and f together have the power to bring about any outcome).

Example 3 *The Discriminatory Vendor*

In our third story, there are three individuals i, j and k. At the start of the story, each individual owns one unit of a different indivisible good: i is the *initial owner* of good I, j of good J, and k of good K. On the assumption that there are no unilateral transfers, there are six possible outcomes of exchange: IJK, IKJ, JIK, JKI, KIJ and KJI, where IJK denotes 'i possesses I, j possesses J, k possesses K', and so on; the status quo position for the story is IJK. An exchange that involves only two individuals' holdings, for example a move from IJK to JIK, will be called *bilateral*; the individuals whose holdings are affected (i and j in this case) are the *parties* to this exchange. An exchange that involves three parties (for example, a move from IJK to JKI) is *trilateral*. Here are two possible specifications of rights to exchange:

Specification 3.1

1. A bilateral exchange takes place if and only if both parties and *the initial owners of the relevant goods consent*. (For example, given that IJK is the status quo position, a move from IKJ to KIJ requires the consent of i and j as parties to the exchange, and of i and k as the initial owners of the two goods that are being exchanged.)

2. A trilateral exchange takes place if and only if all three parties consent.

3. The status quo is IJK.

Specification 3.2

1. A bilateral exchange takes place if and only if both parties consent.

2. A trilateral exchange takes place if and only if all three parties consent.

3. The status quo is IJK.

These two specifications give different packages of rights to the initial owner of a good. Both specifications give this person the right to retain possession if he so chooses, and the right to exchange this good with any willing person or persons. The difference concerns what happens after such an exchange has been made. Under Specification 3.1, the initial owner retains a veto on any subsequent exchanges. Under Specification 3.2, the initial owner's rights in the good - including the right to transfer ownership to others - are transferred in their entirety to the new possessor.

To see the potential significance of this difference, suppose that i thinks that people of k's class, caste, religion or gender should not possess good I. Under Specification 3.2, the only way in which i can guarantee that good I does not reach k is by refusing to trade at all; if i wants to possess good J, she must give up her power of veto over subsequent transfers of good I. Thus, if j and k both prefer JKI to JIK, i has no realistic prospect of achieving JIK. Under Specification 3.1, in contrast, i can credibly try to hold out for JIK by offering to trade bilaterally with j while refusing to consent to any other trades involving good I.

Despite this difference, both specifications imply the effectivity function that has the following properties: $\{i\}$ has effectivity for $\{IJK, IKJ\}$, with $\{j\}$ and $\{k\}$ having symmetrical effectivities; $\{i, j\}$ has effectivity for $\{IJK\}$ and $\{JIK\}$, with $\{i, k\}$ and $\{j, k\}$ having symmetrical effectivities; and $\{i, j, k\}$ has effectivity for $\{JKI\}$ and $\{KIJ\}$.

These three examples have a common structure. First, consider the case of the Proselytising Mineowner. The status quo is UN. Under Specification 1.2, a move from EN to EA can be made with the consent of w alone, while under Specification 1.1, this move requires the consent of both w and e. However, in order for EN to be reached from the

12

status quo, w and e must first act together (exercising their joint right to make a contract of employment). Thus, the idea that w alone has power over the move from EN to EA – the idea that is expressed in Specification 1.2 but not in Specification 1.1 – implies that w retains the power to act independently of e after w and e have acted together.

A similar analysis applies to Subjection of Women. Here the status quo is SC. The two specifications differ in respect of the power to move from MC to MR. Under Specification 2.1 this move can be made with the consent of m alone, while under Specification 2.2 it can be made with the consent of f alone. However, in order for MC to be reached from the status quo, m and f must first act together (exercising their joint right to marry). The idea that m alone or f alone has power over the move from MC to MR implies that one of the two individuals retains the power to act independently of the other after they have both acted together.

The case of the Discriminatory Vendor is a little more complicated. Here the status quo is IJK. The difference between the two specifications is exemplified by their different treatments of the move from JIK to JKI. Under Specification 3.2 this move can be made with the consent of j and k, while under Specification 3.1 it can be made only with the consent of all three individuals. However, in order for JIK to be reached from the status quo, i must give up his initial holding of good I. That implies either some exchange between i and j, which (under either specification) requires at least the consent of i and j, or some exchange between i and k, which requires at least the consent of i and k.[10] Thus, JIK can be reached only if, at some stage, either i and j act together or i and k act together. So the idea that j and k together have power over the move from JIK to JKI – the idea that is expressed by Specification 3.2 but not by Specification 3.1 – implies that at least one of j and k has retained the power to act independently of i after having acted jointly with her.

It seems that the effectivity function is unable to represent the idea that, after a coalition of players has acted jointly, its members remain free to act independently of that coalition. Or, to put this another way, it is unable to represent the idea that the rights of overlapping sets of individuals can be exercised independently of one another. We suggest that what is

---

[10] JIK can be reached from IJK without *direct* trade between i and j. For example, suppose i first trades with k (leading to KJI), then k trades with j (leading to KIJ), then i trades with k.

needed is a theoretical framework in which every 'move' is represented explicitly, and the powers of each coalition are described separately for each move. Our concept of a game in transition function form does exactly this.

3 Games in Transition Function Form

In this Section we introduce our concept of a game in transition function form (or a *TF game*). The TF game form generalizes the effectivity function form in a way that enables us to preserve the distinction between the separate specifications in each of the examples above (and in a host of other possible situations) while maintaining a cooperative game theoretic perspective on the underlying social choice problem; we wish to suggest that it preserves enough information to make proper assessment of the rights characteristics of the underlying situation.

For any set X we define $2^X \equiv \{Y \subseteq X : Y \neq \emptyset\}$ and $\pi(X) \equiv \{(x,y) : x \in X, y \in X \setminus \{x\}\}$.

Definition 1 (A Game in Transition Function Form) *A game in transition function form (or a TF game, generally Γ) is a 6-tuple comprising a set of players (N), a set of social states (σ), a status quo ($q \in \sigma$), a set of social alternatives or outcomes (ω), a mapping (O) from the set of social states to the set of outcomes, and a mapping (generally T) from 2^N to $2^{\pi(\sigma)}$ with the property $\forall A, B \in 2^N, A \subseteq B, (x,y) \in T(A) \rightarrow (x,y) \in T(B)$.*

We interpret T (which we call the *transition function*) in the following way: if an ordered pair of social states (x,y) is an element of $T(A)$ then the players in A can - if other players let the social state rest at x - collectively cause the social state to become y. In other words, A can *effect a transition* from x to y. The only formal constraint on T entails that a set of players (a *coalition*) can effect any transition that can be effected by one or more of its subsets. Additionally, we often assume *transitivity*, by which we mean that a coalition can always effect two consecutive transitions if its complement cannot alter the intermediate

social state. Formally, we say that T is *transitive* if and only if $\forall A \in 2^N$,

$$(x,y) \in T(A), (y,z) \in T(A), \{(y,w)_{w \in \sigma}\} \cap T(N \setminus A) = \{(y,y)\} \to (x,z) \in T(A).$$

O (which we call the *outcome function*) has the interpretation '$O(z)$ is the outcome that prevails in state z'. In other words, $O(z)$ is the *outcome associated with* social state z. In cases where O is bijective we shall abuse notation by treating $O(z)$ as identical to z and then by treating (N, q, ω, T) as an adequate specification of the TF game.

It is often convenient to illustrate a game in transition function form using a directional and labelled multigraph in which social states are represented by vertices labelled by their associated outcomes and in which possible transitions are represented by edges labelled by sets of players that can effect them. We tend to omit the edges that can be deduced from existing edges, our formal constraint on the transition function, and transitivity if the transitivity assumption is apparent or stated. The status quo is represented by a shaded vertex.

3.1 Examples

The formalism above, and also our method of illustration, can be exemplified using the three stories in our introduction.

Example 1 *(The Proselytising Mineowner)*

Specification 1.1 $\left(\Gamma^{1.1}\right)$

$\Gamma^{1.1} = (N, \sigma, q, \omega, O, T)$ where $N = \{e, w\}$, $\sigma = \omega = \{UA, UN, EA, EN\}$,

$q = UN$, O is bijective, and T is defined by

$\qquad T(\{w\}) = \{(UA, UN), (UN, UA)\}$

$\qquad T(\{e\}) = \emptyset$

$\qquad T(\{e, w\}) = \pi(\sigma)$

Specification 1.2 $(\Gamma^{1.2})$

$\Gamma^{1.2} = (N, \sigma, q, \omega, O, T\prime)$ where $N = \{e, w\}$, $\sigma = \omega = \{UA, UN, EA, EN\}$,

$q = UN$, O is bijective, and $T\prime$ is defined by

$\quad T\prime(\{w\}) = \{(UA, UN), (UN, UA), (EA, EN), (EN, EA)\}$

$\quad T\prime(\{e\}) = \emptyset$

$\quad T\prime(\{e, w\}) = \pi(\sigma)$

The difference between our two specifications shows up in the difference between T and $T\prime$. Specifically, $T\prime$ but not T assigns to $\{w\}$ the ability to effect transitions between EA and EN. $\Gamma^{1.1}$ and $\Gamma^{1.2}$ are illustrated in Figure 1.

Figure 1. The Proselytising Mineowner

(LHS shows $\Gamma^{1.1}$; RHS shows $\Gamma^{1.2}$; both transition functions are transitive)

Example 2 *(The Subjection of Women)*

Specification 2.1 $(\Gamma^{2.1})$

$\Gamma^{2.1} = (N, \sigma, q, \omega, O, T)$ where $N = \{m, f\}$, $\sigma = \omega = \{SC, SR, MC, MR\}$,

$q = SC$, O is bijective, and T is defined by

$\quad T(\{m\}) = \{(MC, MR), (MR, MC)\}$

$\quad T(\{f\}) = \{(SC, SR), (SR, SC)\}$

$\quad T(\{m, f\}) = \{(SC, SR), (SR, SC), (SC, MC), (SR, MR), (SC, MR), (SR, MC),$

$\quad\quad\quad\quad\quad\quad (MC, MR), (MR, MC)\}$

16

Specification 2.2 ($\Gamma^{2.2}$)

$\Gamma^{2.1} = (N, \sigma, q, \omega, O, T\prime)$ where $N = \{m, f\}$, $\sigma = \omega = \{SC, SR, MC, MR\}$,

$q = SC$, O is bijective, and $T\prime$ is defined by

$\quad T\prime(\{m\}) = \emptyset$

$\quad T\prime(\{f\}) = \{(SC, SR), (SR, SC), (MC, MR), (MR, MC)\}$

$\quad T\prime(\{m, f\}) = \{(SC, SR), (SR, SC), (SC, MC), (SR, MR), (SC, MR), (SR, MC),$

$\qquad\qquad\qquad (MC, MR), (MR, MC)\}$

In our Subjection of Women story, the difference between specifications lies in the power to effect transitions between MC and MR: T assigns this power to $\{m\}$, while $T\prime$ assigns it to $\{f\}$. Figure 3 illustrates $\Gamma^{2.1}$ and $\Gamma^{2.2}$.

Figure 2. The Subjection of Women

(LHS shows $\Gamma^{2.1}$; RHS shows $\Gamma^{2.2}$; both transition functions are transitive)

Example 3 *(The Discriminatory Vendor)*

Specification 3.1 ($\Gamma^{3.1}$)

$\Gamma^{3.1} = (N, \sigma, q, \omega, O, T\prime)$ where $N = \{i, j, k\}$,

$\sigma = \omega = \{IJK, IKJ, JIK, JKI, KIJ, KJI\}$,

$q = IJK$, O is bijective, and $T\prime$ is defined by

17

$$T\prime(\{i\}) = T(\{j\}) = T(\{k\}) = \emptyset$$
$$T\prime(\{i,j\}) = \{(IJK, JIK), (JIK, IJK)\}$$
$$T\prime(\{i,k\}) = \{(IJK, KJI), (KJI, IJK)\}$$
$$T\prime(\{j,k\}) = \{(IJK, IKJ), (IKJ, IJK)\}$$
$$T\prime(\{i,j,k\}) = \pi(\sigma)$$

Specification 3.2 $(\Gamma^{3.2})$

$\Gamma^{3.2} = (N, \sigma, q, \omega, O, T)$ where $N = \{i, j, k\}$,

$\sigma = \omega = \{IJK, IKJ, JIK, JKI, KIJ, KJI\}$,

$q = IJK$, O is bijective, and T is defined by

$$T(\{i\}) = T(\{j\}) = T(\{k\}) = \emptyset$$
$$T(\{i,j\}) = \{(IJK, JIK), (JIK, IJK), (IKJ, KIJ),$$
$$(KIJ, IKJ), (JKI, KJI), (KJI, JKI)\}$$
$$T(\{i,k\}) = \{(IJK, KJI), (KJI, IJK), (IKJ, JKI),$$
$$(JKI, IKJ), (JIK, KIJ), (KIJ, JIK)\}$$
$$T(\{j,k\}) = \{(IJK, IKJ), (IKJ, IJK), (JIK, JKI),$$
$$(JKI, JIK), (KJI, KIJ), (KIJ, KJI)\}$$
$$T(\{i,j,k\}) = \pi(\sigma)$$

Again, our two specifications entail two different transition functions.
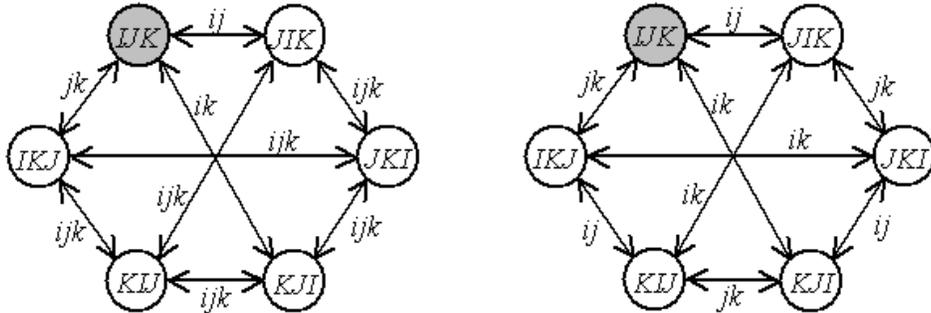
Figure 3. The Discriminatory Vendor

(LHS shows $\Gamma^{3.1}$; RHS shows $\Gamma^{3.2}$; both transition functions are transitive)

3.2 Further Examples

The transition function form enables us to describe a very broad range of social decision making procedures. We will briefly mention two further examples that (though they are not directly relevant to our claims about TF games and rights) help to illustrate its scope. Both examples can be thought of as 'investment games', and the outcomes take the shape of money payoffs. The first game (based on a game proposed by Berg, Dickhaut and McCabe, 1995) entails *trust*. A unit of wealth (initially held by the investor, a) can be transformed into three units only if it is first transferred to an entrepreneur, b; but it is then up to the entrepreneur to decide whether to return a's investment with a share of the profit. The second game simply entails *risk*. Investor a has one unit of wealth, and has the option of investing this in such a way that chance alone will determine whether her wealth is doubled or lost.

Example 4 *(A Trust Game)*

$\Gamma^4 = (N, \sigma, q, \omega, O, T)$ where $N = \{a, b\}$, $\sigma = \{W, X, Y, Z\}$, $q = (1, 0)$,
$\omega = \{(1, 0), (2, 1), (0, 3)\}$, $O = \{(W, (1, 0)), (X, (1, 0)), (Y, (0, 3)), (Z, (2, 1))\}$,
and T is defined by
$$T(\{a\}) = \emptyset, \quad T(\{b\}) = \{(X, Y), (X, Z)\}$$
$$T(\{a, b\}) = \{(W, X), (X, Y), (X, Z)\}$$

Example 5 *(A Risk Game)*

$\Gamma^5 = (N, \sigma, q, \omega, O, T)$ where $N = \{a, \rho\}$, $\sigma = \omega = \{1, 2, 0\}$, $q = 1$,
O is bijective and T is defined by
$$T(\{a\}) = \{(1, 2), (1, 0)\}, \quad T(\{\rho\}) = \{(2, 0), (0, 2)\}$$
$$T(\{a, \rho\}) = \pi(\sigma)$$

19
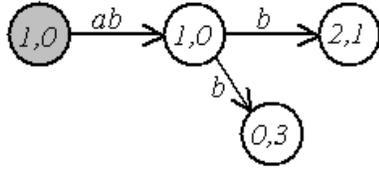
Figure 4. A Trust Game   Figure 5. A Risk Game
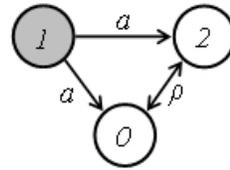
Example 4 differs from our previous examples in two important respects. First, *the transition function is not transitive*: (W, X) and (X, Z) are elements of $T(\{a, b\})$, and $\{a, b\}$ can veto any transitions out of state X, but still (W, Z) is not an element of $T(\{a, b\})$. If (W, Z) *were* an element of $T(\{a, b\})$ the game would be very different: the players would have a straightforward transition from the status quo to their mutual advantage. As it is, a's readiness to leave state W must be be doubted. Second, *the outcome function is not bijective*: there are two social states (W and X) that share the same outcome, but which must be viewed as distinct because they entail different transition rights. Though the outcome functions in each of our three main examples are bijective, in other cases they will commonly not be.

In Example 5, though the underlying story only entails a single player, the TF game entails a second player (ρ) to be thought of as 'nature' or 'chance'; the use of a 'chance' player considerably expands the range of circumstances that can be accommodated by the TF game form.

4   Transition Functions and Effectivity Functions

In this Section we formalise the relationship between the transition and effectivity function forms. The transition function is a generalisation of the effectivity function in the sense that each TF game has a unique effectivity function (EF) representation, while a game in effectivity function form (or *EF game*) generally has many TF representations.

4.1 Games in Effectivity function form

Definition 2 (A Game in Effectivity Function Form) *A game in effectivity function form (or EF game, generally $\mathcal{G}$) is a 3-tuple comprising a set of players (N), a set of social outcomes (ω), and a function (generally E) from 2^N to the power set of all sets of social outcomes ($2^{(2^\omega)}$) with the following properties:*

1. *$\forall A, B \in 2^N, A \subseteq B, \omega\prime \in E(A) \rightarrow \omega\prime \in E(B)$*

2. *$\forall A \in 2^N, \omega\prime \in E(A), \omega\prime \subseteq \omega\prime\prime \rightarrow \omega\prime\prime \in E(A)$*

3. *For any partition $\{A, B, C, \ldots\}$ of N, if $\omega\prime, \omega\prime\prime, \omega\prime\prime\prime, \ldots$ are elements of $E(A), E(B), E(C), \ldots$ respectively, then $\bigcap_{\omega\prime, \omega\prime\prime, \omega\prime\prime\prime, \ldots} \neq \emptyset$.*

The interpretation of E is that if $\omega\prime \in E(A)$ then the players in A can collectively ensure that the outcome lies in $\omega\prime$.

4.2 The EF representation of a TF game

A game in effectivity function form can be used to summarise a strategic game form; in this case we generally say $\omega\prime \in E(A)$ if the players in A have a joint strategy such that all strategies chosen by $N\backslash A$ lead to an outcome in $\omega\prime$.[11] A game in effectivity function form can be similarly used to summarise a game in transition function form; in this case we shall say $\omega\prime \in E(A)$ if the players in A can, in finite time and regardless of transitions effected by $N\backslash A$, effect a transition of the social state permanently into some $\sigma\prime \subseteq \sigma$, where the outcomes associated with social states $\sigma\prime$ are all contained in $\omega\prime$. We formalize this idea by defining an *EF representation of a TF game*.

Given any TF game $(N, \sigma, q, \omega, O, T)$ let $S_{\pi(\sigma)}$ be the set of identity mappings s on σ. For any set of ordered pairs of social states $\pi\prime \subseteq \pi(\sigma)$ We define $S_{\pi\prime} \subseteq S_{\pi(\sigma)}$:

$$S_{\pi\prime} \equiv \left\{ s \in S_{\pi(\sigma)} : \forall x \in \sigma, s(x) \neq x \rightarrow (x, s(x)) \in \pi\prime \right\}.$$

---

[11] Effectivity as described here is sometimes referred to as α-*effectivity*. Other interpretations of E include β-*effectivity* whereby we say $Z \in E(A)$ if for any strategy chosen by $N\backslash A$ the players in A have a joint strategy that leads to an outcome in Z.

Thus, for any $A \in 2^N$, $S_{T(A)}$ can be thought of as the *strategy set* for A. A strategy s for coalition A assigns to every social state, x, another, $s(x)$, with the interpretation that A will effect a transition from x to $s(x)$ whenever it gets the chance to do so. But a strategy is only possible if for every social state x either $x = s(x)$ or else the pair $(x, s(x))$ belongs to $T(A)$; $S_{T(A)}$ is the set of strategies for which this is true.

For any $S_{\pi\prime} \subseteq S_{\pi(\sigma)}$ and any $\sigma\prime \subseteq \sigma$ let $D^1_{S_{\pi\prime}}(\sigma\prime) \equiv \bigcup_{s \in S_{\pi\prime}} \bigcup_{x \in \sigma\prime} s(x)$, let $D^t_{S_{\pi\prime}}(\sigma\prime) \equiv D^1_{S_{\pi\prime}}(D^{t-1}_{S_{\pi\prime}}(\sigma\prime))$, and let $D^\infty_{S_{\pi\prime}}(\sigma\prime) \equiv \lim_{t \longrightarrow \infty} D^t_{S_{\pi\prime}}(\sigma\prime)$.

For a single strategy, s, and a single social state x, $D^1_{\{s\}}(\{x\}) = s(x)$. For any $A \in 2^N$, $D^1_{S_{T(A)}}(\{x\})$ and $D^t_{S_{T(A)}}(\{x\})$ denote respectively the sets of social states that can be reached from x in a single transition and in t transitions effected by A. And for a set of social states $\sigma\prime$, $D^\infty_{S_{T(A)}}(\sigma\prime)$ denotes the set of states that can be reached from some social state in $\sigma\prime$ in any number of transitions effected by A.

Finally, for any $\sigma\prime \subseteq \sigma$ and any $\pi\prime, \pi\prime\prime \subseteq \pi(\sigma)$ we define the following sequence:

$$\left\{ (\sigma\prime)^{S_{\pi\prime}, S_{\pi\prime\prime}}_t \right\}^\infty_{t=0} \equiv \left\{ \sigma\prime, D^\infty_{S_{\pi\prime}}(\sigma\prime), D^1_{S_{\pi\prime\prime}}(D^\infty_{S_{\pi\prime}}(\sigma\prime)), D^\infty_{S_{\pi\prime}}(D^1_{S_{\pi\prime\prime}}(D^\infty(\sigma\prime))), \ldots \right\}.$$

i.e.

$$(\sigma\prime)^{S_{\pi\prime}, S_{\pi\prime\prime}}_t \equiv \begin{cases} (\sigma\prime) & \text{where } t = 0 \\ D^\infty_{S_{\pi\prime}}((\sigma\prime)^{S_{\pi\prime}, S_{\pi\prime\prime}}_{t-1}) & \text{where } t \text{ is any positive, odd integer} \\ D^1_{S_{\pi\prime\prime}}((\sigma\prime)^{S_{\pi\prime}, S_{\pi\prime\prime}}_{t-1}) & \text{where } t \text{ is any positive, even integer} \end{cases}$$

Informally this means that for any $A \in 2^N$ and for any $s \in S_{T(A)}$, the sequence $\left\{ (\{q\})^{S_{T(N \setminus A)}, \{s\}}_t \right\}^\infty_{t=0}$ comprises *first* the singleton whose element is the status quo, q, *second* the set of social states (including q) that can be reached in any number of transitions effected by A's complement from q, *third* the set of social states that arises after a single transition according to A's strategy s from the second set, *fourth* the set of social states that can be reached in any number of transitions effected by A's complement from the third set, and so on. If and only if there is a strategy $s \in S_{T(A)}$ and a set of social states $\sigma\prime \subseteq \sigma$ such that (i) for any t greater than some $t\prime$, $(\{q\})^{S_{T(N \setminus A)}, \{s\}}_t \subseteq \sigma\prime$ and (ii) the outcomes

22

associated with social states in $\sigma\prime$ all lie in $\omega\prime \subseteq \omega$, then A has effectivity for $\omega\prime$. It is this idea that underlies our definition of the EF representation of a TF game.

Definition 3 (The EF representation of a TF game) (N, ω, E) *is the EF representation of* $(N, \sigma, q, \omega, O, T)$ *if and only if for any* $A \in 2^N$,

where $F(A) \equiv \left\{ \sigma\prime \subseteq \sigma : \exists s \in S_{T(A)}, \exists t\prime \in \mathbb{Z}_+, t > t\prime \rightarrow (\{q\})_t^{S_{T(N \setminus A)}, \{s\}} \subseteq \sigma\prime \right\}$,

$$E(A) = \left\{ \omega\prime \subseteq \omega : \exists \sigma\prime \in F(A), \bigcup_{y \in \sigma\prime} \{O(y)\} \subseteq \omega\prime \right\}.$$

It should be noted that, by definition, each TF game has exactly one EF representation and moreover it is straightforward - provided only that σ is finite - to write a finite algorithm that finds it (see the Appendix).

4.3 Examples

Using Definition 3 we can derive formal EF representations of our main example TF games. As we have anticipated in our discussion already, in each case the EF representation is identical across the two specifications. Throughout, we define the effectivity function E on (N, ω) in terms of its *basis*, $\overline{E}$, with $\forall A \in 2^N$, $E(A) = \{\omega\prime \subseteq \omega : \exists \omega\prime\prime \in \overline{E}(A), \omega\prime\prime \subseteq \omega\prime\}$.

Example 1 *(The Proselytising Mineowner)*

The EF representation of $\Gamma^{1.1}$ and of $\Gamma^{1.2}$ is

$\mathcal{G}^1 = (N, \omega, E)$ where $N = \{e, w\}$, $\omega = \{UA, UN, EA, EN\}$, and E is defined by

$\overline{E}(\{w\}) = \{\{UN\}, \{UA\}\}$
$\overline{E}(\{e\}) = \{\{UN, UA\}\}$
$\overline{E}(\{e, w\}) = \{\{UN\}, \{UA\}, \{EN\}, \{EA\}\}$

Example 2 *(The Subjection of Women)*

The EF representation of $\Gamma^{2.1}$ and of $\Gamma^{2.2}$ is

$\mathcal{G}^2 = (N, \omega, E)$ where $N = \{m, f\}$, $\omega = \{SC, SR, MC, MR\}$, and E is defined by

$$\overline{E}(\{m\}) = \{\{SC, SR\}\}$$

$$\overline{E}(\{f\}) = \{\{SC\}, \{SR\}\}$$

$$\overline{E}(\{m, f\}) = \{\{SC\}, \{SR\}, \{MC\}, \{MR\}\}$$

Example 3 *(The Discriminatory Vendor)*

The EF representation of $\Gamma^{3.1}$ and of $\Gamma^{3.2}$ is

$$\mathcal{G}^3 = (N, \omega, E) \text{ where } N = \{i, j, k\}, \omega = \{IJK, IKJ, JIK, JKI, KIJ, KJI\},$$

and E is defined by

$$\overline{E}(\{i\}) = \{\{IJK, IKJ\}\}$$

$$\overline{E}(\{j\}) = \{\{IJK, KJI\}\}$$

$$\overline{E}(\{k\}) = \{\{IJK, JIK\}\}$$

$$\overline{E}(\{i, j\}) = \{\{IJK\}, \{JIK\}\}$$

$$\overline{E}(\{i, k\}) = \{\{IJK\}, \{KJI\}\}$$

$$\overline{E}(\{j, k\}) = \{\{IJK\}, \{IKJ\}\}$$

$$\overline{E}(\{ijk\}) = \{\{IJK\}, \{IKJ\}, \{JIK\}, \{JKI\}, \{KIJ\}, \{KJI\}\}$$

We noted in Section 2 that in each of our main Examples (1-3) our two *specifications* entail transition functions that differ only in transitions that can be effected by a coalition A, from a social state x that cannot be reached from the status quo in any number of transitions effected by A or $N \setminus A$. Put differently, the differences lie only in transitions that can be effected by coalitions that can come about only when earlier coalitions disband. The following proposition states that in all such circumstances, the EF representation makes no distinction between the two TF games.

Proposition 1 $(N, \sigma, q, \omega, O, T)$ *and* $(N, \sigma, q, \omega, O, T\prime)$ *will have the same EF representation if* $\forall A \in 2^N$,

$$\left\{ (x, y) \in T(A) : x \in D^{\infty}_{S_{T(A)} \cup S_{T(N \setminus A)}}(\{q\}) \right\} = \left\{ (x, y) \in T\prime(A) : x \in D^{\infty}_{S_{T\prime(A)} \cup S_{T\prime(N \setminus A)}}(\{q\}) \right\}.$$

Proof. $D^{\infty}_{S_{T(A)} \cup S_{T(N \setminus A)}}(\{q\})$ denotes the set of social states that can be reached in an unlimited number of transitions effected by A or $N \setminus A$. Note that $\forall A \in 2^N, \forall s \in S_{T(A)}, \forall t \in$

$\mathbb{Z}_+, (\{q\})_t^{S_{T(N \setminus A)}, \{s\}} \subseteq D^\infty_{S_{T(A)} \cup S_{T(N \setminus A)}}(\{q\})$. So

$$\left\{(x, y) \in T(A) : x \in D^\infty_{S_{T(A)} \cup S_{T(N \setminus A)}}(\{q\})\right\} = \left\{(x, y) \in T\prime(A) : x \in D^\infty_{S_{T\prime(A)} \cup S_{T\prime(N \setminus A)}}(\{q\})\right\}$$

$$\rightarrow \forall s \in S_{T(A)}, \exists s\prime \in S_{T\prime(A)}, \forall t \in \mathbb{Z}_+, (\{q\})_t^{S_{T(N \setminus A)}, \{s\}} = (\{q\})_t^{S_{T\prime(N \setminus A)}, \{s\prime\}}. \qquad \blacksquare$$

4.4 TF representations of an EF game

Theorem 1 *Given any EF game (N, ω, E), and any $z \in \omega$, provided that $\{\{x\} : x \in \omega\} \subseteq E(N)$, there is some TF game on (N, ω) that has (N, ω, E) as its EF representation, and z as its status quo outcome.*

Proof. Theorem 1 is an immediate consequence of Proposition 2 (below). $\qquad \blacksquare$

Proposition 2 sets out a *method* for deriving a TF game with any given status quo outcome and any given EF representation.[12] We use $\Phi(N)$ to denote the set of all partitions of N.

Proposition 2 *Given (N, ω, E), and given $z \in \omega$,*

if $\sigma = \{(\phi, I, \omega\prime, x) : \phi \in \Phi(N), I \in \phi, \omega\prime \subseteq \omega, x \in \omega\prime\}$,

if $q = (\{N\}, N, \omega, z)$,

if $O : \sigma \rightarrow \omega$ is defined using $\forall(\phi, I, \omega\prime, x) \in \sigma, O(\phi, I, \omega\prime, x) = x$,

and if $T : 2^N \longrightarrow 2^\pi$ is defined using

$$\forall A \in \ 2^N, \forall(\phi, I, \omega\prime, x), (\phi\prime, J, \omega\prime\prime, y) \in \sigma,$$

$$((\phi, I, \omega\prime, x), (\phi\prime, J, \omega\prime\prime, y)) \in T(A) \leftrightarrow$$

$$(either \ \phi = \phi\prime, I = J, \omega\prime = \omega\prime\prime, A \subseteq I,$$

$$or \ \exists B \subseteq A, \exists \omega\prime\prime\prime \in E(B), J = I \setminus B, \phi\prime = \{J, B\} \cup \phi \setminus \{I\}, \omega\prime\prime = \omega\prime \cap \omega\prime\prime\prime),$$

then (N, ω, E) is the EF representation of the TF game $(N, \sigma, q, \omega, O, T)$ and $O(q) = z$.

---

[12] A TF representation of an EF game for which the proviso in Theorem 1 does not hold would be derived by applying the method set out in Proposition 2 on a modified EF game, $(N \cup \{\rho\}, \omega, E\prime)$ with the properties $\forall A \in 2^N \setminus \{N\}, E\prime(A) = E\prime(A \cup \{\rho\}) = E(A), E\prime(N) = E(N)$ and $E\prime(N \cup \{\rho\}) = \{\{x\} : x \in \omega\}$. ρ here is the player that represents 'chance' as described in section 3.2.

Proof. We first show that if $\omega^* \in E(A)$ then

$$\exists s \in S_{T(A)}, \exists \sigma\prime \subseteq \sigma, \forall t \in \mathbb{Z}_+, t \geqslant 2 \to (\{q\})_t^{S_{T(N \backslash A)}, \{s\}} \subseteq \sigma\prime, \bigcup_{x \in \sigma\prime} \{O(x)\} \subseteq \omega^*.$$

For any $s \in S_{T(A)}$, $(\{q\})_1^{S_{T(N \backslash A)}, \{s\}} \subseteq \{(\phi, I, \omega\prime, x) \in \sigma : A \subseteq I, \omega^* \cap \omega\prime \neq \emptyset\}$. There is then some $s \in S_{T(A)}$ such that $(\{q\})_2^{S_{T(N \backslash A)}, \{s\}} \subseteq \{(\phi, I, \omega^*, x) \in \sigma : A \not\subseteq I\}$ and therefore such that $\forall t \geqslant 2, (\{q\})_t^{S_{T(N \backslash A)}, \{s\}} \subseteq \{(\phi, I, \omega\prime, x) \in \sigma : \omega\prime \subseteq \omega^*\}$.

We then show that if $\omega^* \notin E(A)$ then

$$\forall s \in S_{T(A)}, \forall t \in \mathbb{Z}_+, \exists t\prime > t, \exists x \in (\{q\})_t^{S_{T(N \backslash A)}, \{s\}}, O(x) \notin \omega^*.$$

It suffices to note, for any $s \in S_{T(A)}$:

    1. $\{(\{N\}, N, \omega, x) : x \in \omega\} \subseteq (\{q\})_1^{S_{T(N \backslash A)}, \{s\}}$.

    2. If t is even then $\forall(\phi, I, \omega\prime, x) \in (\{q\})_t^{S_{T(N \backslash A)}, \{s\}}, I \not\subseteq A \to \{(\phi, I, \omega\prime, y) : y \in \omega\prime\} \subseteq (\{q\})_{t+1}^{S_{T(N \backslash A)}, \{s\}}$.

    3. $\forall(\phi, I, \omega\prime, x) \in \sigma, \omega\prime \notin \omega^*, I \not\subseteq A, s(\phi, I, \omega\prime, x) = (\phi\prime, J, \omega\prime\prime, y) \to J \not\subseteq A, \omega\prime\prime \not\subseteq \omega^*$. ∎

4.5 x-conditional EF representations

A natural extension of Definition 3 can be used to designate coalitional effectivities within the social decision *structure* $(N, \sigma, \omega, O, T)$ not only at the status quo q, but also at possible hypothetical alternative social states. We shall refer to these descriptions of hypothetical effectivities as *x-conditional EF representations*. The formal definition of an x-conditional EF representation substitutes a generic social state x for the status quo, q, in the definition of an EF representation.

Definition 4 (The x-conditional EF representation of a TF game) *For any $x \in \sigma$,*
(N, ω, E) is the x-conditional EF representation of $(N, \sigma, q, \omega, O, T)$ iff for any $A \in 2^N$,
where $F(A) \equiv \left\{\sigma\prime \subseteq \sigma : \exists s \in S_{T(A)}, \exists t\prime \in \mathbb{Z}_+, t > t\prime \to (\{x\})_t^{A, s} \subseteq \sigma\prime\right\}$,

$$E(A) = \left\{\omega\prime \subseteq \omega : \exists \sigma\prime \in F(A), \bigcup_{y \in \sigma\prime} \{O(y)\} \subseteq \omega\prime\right\}.$$

26

By Definition 4, the EF representation of a game can also be termed the *status quo conditional* EF representation of the game.

5 The representation of rights

By using games in transition function form, it is possible to represent social decision making procedures in compact and theoretically tractable models. Such models are well suited to the analysis of rights. Specifically, we propose that a right should be defined with respect to some (N, ω); it will designate a subset of the set of all possible TF games on that (N, ω).

We envisage that a right will be expressed as a criterion against which all TF games on (N, ω) can be measured. The types of criterion we have in mind (though we do not wish to rule out others) are *properties of x-conditional EF representations*. The following (1-3) exemplify forms such criteria could take, with respect to TF games Γ on (N, ω). In each case we take A to be a specific member of 2^N and $\omega\prime, \omega\prime\prime$ to be a specific subsets of ω. (Recalling our notation in Section 4.2, where q is the status quo in Γ we shall say that $D^\infty_{S_{T(N)}}(\{q\})$ is the set of *feasible* states of Γ.)

Form 1. If (N, ω, E) is the (status quo conditional) EF representation of Γ, then $\omega\prime \in E(A)$.

Form 2. For any feasible social state x of Γ, if (N, ω, E) is the x-conditional EF representation of Γ, then $\omega\prime \in E(A)$.

Form 3. For any feasible social state x of Γ, if the outcome associated with x belongs to $\omega\prime\prime$ and if (N, ω, E) is the x-conditional EF representation of Γ, then $\omega\prime \in E(A)$.

A criterion taking Form 1 imposes a straightforward property on the EF representation of the TF game, and rights that have this structure can be analysed using EF games. But criteria taking Forms 2 and 3 are properties of x-conditional EF representations of the TF game so that rights with these structures can only be analysed using the additional information about the social decision making procedure that the TF game form preserves.

As an illustration of this modelling strategy, consider the case of the Proselytising Mineowner, as represented in $\Gamma^{1.1}$ and $\Gamma^{1.2}$. Consider the following rights, written as properties of x-conditional EF representations of TF games, Γ, on $(N = \{e, w\}, \omega = \{UN, UA, EA, EN\})$:

R1. If (N, ω, E) is the (status quo conditional) EF representation of Γ, then $\{UN, EN\}$, $\{UA, EA\} \in E(\{w\})$.

R2. If (N, ω, E) is the (status quo conditional) EF representation of Γ, then $\{UN, UA\} \in (E(\{e\}) \cap E(\{w\}))$ and $\{EN, EA\} \in E(\{e, w\})$.

R3. For any feasible social state x, if (N, ω, E) is the x-conditional EF representation of Γ, then $\{UN, EN\}, \{UA, EA\} \in E(\{w\})$.

R4. (i) For any feasible social state x associated with outcomes UN or UA, if (N, ω, E) is the x-conditional EF representation of Γ, then $\{UN, UA\} \in (E(\{e\}) \cap E(\{w\}))$ and $\{EN, EA\} \in E(\{e, w\})$.

(ii) For any feasible social state x associated with outcomes EN or EA, if (N, ω, E) is the x-conditional EF representation of Γ, then $\{EN, EA\} \in (E(\{e\}) \cap E(\{w\}))$ and $\{UN, UA\} \in E(\{e, w\})$.

$R1$ expresses a right of freedom of worship: it states that w alone has the power to bring about an outcome in which he does not attend church, and also has the power to bring about an outcome in which he does. $R2$ expresses a right of freedom of contract: it states that each of w and e has the power to bring about an outcome in which w does not work for e, and that w and e jointly have the power to bring about an outcome in which he does. It is easy to verify that in each of $\Gamma^{1.1}$ and $\Gamma^{1.2}$, $R1$ and $R2$ (which have the structure of Form 1 above) are respected.

However, we may want to ask whether the rights of freedom of worship and freedom of contract can be exercised together. More specifically: if w and e exercise their joint right to make an employment contract, does w retain the right of freedom of worship? If an employment contract is made, the players move from UN to another state (EN or EA).

So, to answer the question, we must define the right of freedom of worship at states other than the status quo and then investigate whether this more general right is respected. $R3$ expresses one natural generalisation of the right of freedom of worship: the condition that, at each possible social state, $\{w\}$ has effectivity for $\{UN, EN\}$ and $\{UA, EA\}$. $R4$ expresses a corresponding generalisation of the right of freedom of contract: that at UN and UA, each of $\{w\}$ and $\{e\}$ has effectivity for $\{UN, UA\}$ and $\{w, e\}$ has effectivity for $\{EN, EA\}$, while at EN and EA, $\{w\}$ and $\{e\}$ have effectivity for $\{EN, EA\}$, and $\{w, e\}$ has effectivity for $\{UN, UA\}$. $R3$ and $R4$ have the same structures as Forms 2 and 3 above.

Given this specification of rights, the question posed in the preceding paragraph can be answered. In $\Gamma^{1.1}$, $R3$ is not respected (although $R4$ is). At EN, $\{w\}$ does not have effectivity for $\{UA, EA\}$, contrary to a requirement of $R3$. Similarly, at EA, w does not have effectivity for $\{UN, EN\}$. In other words, the *inalienable* right to freedom of worship (as represented by $R3$) is not a property of $\Gamma^{1.1}$. In $\Gamma^{1.2}$, in contrast, $R3$ is respected (and so is $R4$).

The Subjection of Women case can be analysed in a very similar way. The rights of m and f with respect to marriage are analogous with those of w and e with respect to employment, except that the marriage contract cannot be terminated even at the demand of both parties. Mill's specification of f's right to choose how to use her property is analogous with w's right to choose whether or not to attend church. In both games, these rights are respected at the status quo, but f's property rights are alienable in $\Gamma^{2.1}$ and inalienable in $\Gamma^{2.2}$.

In the case of the Discriminatory Vendor, the following $R1'$ and $R2'$ express rights of freedom of contract. $R1'$ and $R2'$ are symmetrical between goods, between players and between social states, so that six separate properties $R_{IJK}, R_{JIK}, \ldots etc$ can be each constructed from any other by transpositions of I, J and K. Here, then, we just set out R_{IJK}:

R^1_{IJK}. If IJK is the outcome associated with the status quo and if (N, ω, E) is the (status quo conditional) EF representation of Γ, then

(i) $\{IJK, IKJ\} \in E(\{i\}), \{IJK, IKJ\} \in E(\{j\}), \{IJK, IKJ\} \in E(\{k\})$

(ii) $\{JIK\} \in E(\{i,j\}), \{KJI\} \in E(\{i,k\}), \{IKJ\} \in E(\{j,k\})$

(iii) $\{JKI\}, \{KIJ\} \in E(\{i,j,k\})$.

$R1'$. $R^1_{IJK} \cap R^1_{JIK} \cap R^1_{JKI} \cap R^1_{KJI} \cap R^1_{KIJ} \cap R^1_{IKJ}$.

R^2_{IJK}. For any feasible social state x associated with the outcome IJK, if (N, ω, E) is the
x-conditional EF representation of Γ, then

(i) $\{IJK, IKJ\} \in E(\{i\}), \{IJK, IKJ\} \in E(\{j\}), \{IJK, IKJ\} \in E(\{k\})$

(ii) $\{JIK\} \in E(\{i,j\}), \{KJI\} \in E(\{i,k\}), \{IKJ\} \in E(\{j,k\})$

(iii) $\{JKI\}, \{KIJ\} \in E(\{i,j,k\})$.

$R2'$. $R^2_{IJK} \cap R^2_{JIK} \cap R^2_{JKI} \cap R^2_{KJI} \cap R^2_{KIJ} \cap R^2_{IKJ}$.

Each R_{IJK} has three parts: part (i) is the power of each player alone to ensure he keeps his own possessions, part (ii) is the joint power of each pair of players to make bilateral exchanges, and part (iii) is the joint power of the set of three players to make trilateral exchanges. But $R1'$ is an alienable right, while $R2'$ is inalienable.

Now consider $\Gamma^{3.1}$. In this game, $R1'$ is respected but $R2'$ is not. Specifically, though R^2_{IJK} is respected, R^2_{JIK}, R^2_{JKI}, R^2_{KJI}, R^2_{KIJ} and R^2_{IKJ} are not. In $\Gamma^{3.2}$, in contrast, $R1'$ and $R2'$ are both respected: the right of freedom of contract is inalienable.

6 Rights and Pareto efficiency

So far, we have deliberately avoided saying anything about players' preferences. As we explained in Section 1, our concern in this paper is with the specification of rights; and we are following the procedural approach, in which rights are independent of preferences. We believe that the concept of a game in transition function form will have other applications. If the full potential of this concept is to be exploited, it will clearly be necessary to develop solution concepts that use information about preferences. But that is a task for another paper.

However, Sen's impossibility result has been so influential in the literature of rights that many readers will be expecting some discussion of the compatibility or incompatibility of

rights-respecting social decision making procedures with Pareto efficiency. In this Section, we offer some speculations about how the distinction between alienable and inalienable rights might affect the outcomes of games played by rational individuals.

As a starting point for this discussion, it is useful to go back to a seminal contribution to the procedural analysis of rights, in which the idea of alienable rights plays an important role. Gärdenfors (1981) distinguishes between *having* a right and *exercising* it. His analysis of 'having a right' is independent of preferences; it uses a concept of a 'rights system' that is similar to an effectivity function. This is supplemented by a game-theoretic analysis of the exercise of rights by rational players with given preferences. The aim of this latter analysis is 'to give rationality criteria for when an individual or a group ought to exercise a right and when the right ought to be waived' (p. 347). Gärdenfors treats the exercise of rights as a cooperative game, in which the formation of coalitions is 'an essential part' (p. 348). He proposes a principle of rational play which, when combined with the condition that the rights system gives the set of all players effectivity for every non-empty set of outcomes, implies that the outcome of every such game is Pareto-efficient. Gärdenfors presents this result as a resolution of the paradox of the Paretian liberal (p. 342).

The idea that rational play of cooperative games will lead to Pareto-efficient outcomes rests on two implicit assumptions. The first is the assumption that, if the agreements by which coalitions are formed are enforceable, no set of rational players will fail to take advantage of an opportunity to form a coalition that would make all of them better off. The real-world applicability of this principle of rationality is open to question, but the issues involved are orthogonal to the analysis we have presented in this paper. The second and more relevant assumption is that, once a coalition is formed, every subsequent action by each of its members is constrained by the terms of the agreement by which that coalition was formed. If this assumption did not hold, a coalition's collective action in pursuit of its joint interest might be compromised by subsequent unilateral actions by individual members.

We have argued that it is an essential feature of some systems of rights that the formation of a coalition for one purpose (in our model, to achieve a transition from one state to another) does not remove the power of its members to act subsequently, either as individuals or as

members of other coalitions, for other purposes. When rights take this form, there can be no general presumption that the rational exercise of rights will be Pareto-efficient.

For example, take the case of the Proselytising Mineowner. Suppose that w has the strict preference ordering $EN \succ EA \succ UN \succ UA$ while e's preferences are $EA \succ UA \succ UN \succ EN$. Thus, w prefers to work; other things being equal, he prefers not to go to church, but he would be willing to go to church if that was necessary to get a job. Other things being equal, e prefers not to employ w, but she would be willing to do so if that was necessary to get him to go to church. Consider Specification 3.2 in which freedom of worship is inalienable. Given the preferences we have assumed, it seems clear that rational play of this game will result in UN. If w and e acted jointly to make an employment contract, w would then have the power to achieve his most preferred outcome EN. Anticipating this, e can guarantee that the outcome is one she prefers to EN by exercising her right not to employ w; and then the best that w can do is to exercise his right not to go to church. The result is UN, even though both individuals prefer EA.[13] The obstacle to their achieving the outcome they both prefer is the inalienability of the right to freedom of worship.[14]

Thus, our analysis confirms Sen's original insight that respect for rights can conflict with the Pareto principle. More precisely, a social decision making procedure that respects rights may have Pareto-inefficient results, even if every individual acts rationality with the objective of satisfying his or her preferences. That this conflict is unavoidable has been recognised by many advocates of the procedural approach (e.g. Sugden, 1985: 227-228; Gaertner et al, 1992: 161). Many theorists have had the intuition that this conflict arises, at least in part, from constraints on the alienability of rights, but the development of this idea has been hindered by the lack of a theoretical framework within which the distinction between alienable and inalienable rights can be represented. We suggest that the concept

---

[13]Notice that the core of this game is $\{EA\}$. Thus, the failure of rational players to achieve a Pareto-efficient outcome cannot be attributed to the non-existence or non-uniqueness of a core solution. This result suggests that the core is not a satisfactory solution concept for games in transition function form.

[14]More precisely, the inalienablity of the right to freedom of worship is combined with a specification of the right of free contract which requires w's consent for the termination of the employment contract. If e had the unilateral power to terminate the contract, she could threaten to use this power unless w went to church, making EA a possible result of the game. In Specification 1.2, in which the right to freedom of worship is alienable, it seems clear that rational play would lead to EA.

of a game in transition function form fills that gap.

Appendix

An algorithm to find the EF representation of a TF game.

For any A in 2^N we first enumerate $S_{T(A)}$ so that $S_{T(A)} = \{s_1, s_2, \ldots, s_p\}$. The following algorithm then generates $E(A)$.

1.	Let $\sigma_1 = D^\infty_{S_{T(N \setminus A)}}(\{q\})$, let $F(A) = \{\sigma\prime \in 2^\sigma : \sigma_1 \subseteq \sigma\prime\}$
2.	For $i = 1$ to p
3.	Let $t = 2$
4.	Let $\sigma_t = D^1_{\{s_i\}}(\sigma_{t-1})$
5.	Let $\sigma_{t+1} = D^\infty_{S_{T(N \setminus A)}}(\sigma_t)$
6.	If $\sigma_{t+1} \in F(A)$ then go to **9.**
7.	Let $F(A) = F(A) \cup \{\sigma\prime \in 2^\sigma : \sigma_{t+1} \subseteq \sigma\prime\}$, let $t = t + 2$
8.	Go to **4.**
9.	Next i
10.	Let $E(A) = \bigcup_{\sigma\prime \in F(A)} \bigcup_{x \in \sigma\prime} O(x)$

It is easy to check that the algorithm runs in finite time: p cannot be greater than $|\sigma|^{|\sigma|}$ and (due to step 6.) no i loop can run beyond $t = 2^{|\sigma|+1}$. Step 6. uses the following observation: where $t\prime$ and t are both odd, $t\prime > t$, $(\{q\})^{S_{T(N \setminus A)}, \{s\}}_t \subseteq (\{q\})^{S_{T(N \setminus A)}, \{s\}}_{t\prime} \rightarrow \exists t\prime\prime > t\prime, (\{q\})^{S_{T(N \setminus A)}, \{s\}}_{t\prime} \subseteq (\{q\})^{S_{T(N \setminus A)}, \{s\}}_{t\prime\prime}$.

References

Berg, Joyce, John Dickhaut and Kevin McCabe (1995). Trust, Reciprocity and Social History. Games and Economic Behavior 10: 122-42.

Deb, Rajat (1994). Waiver, effectivity and rights as game forms. Economica 61: 167-178.

Gaertner, Wulf, Prasanta Pattanaik and Kotaro Suzumura (1992). Individual rights revisited. Economica 59: 161-177.

Gärdenfors, Peter (1981). Rights, games and social choice. Noûs 15: 341-356.

Gibbard, Allan (1973). Manipulation of voting schemes: a general result. Econometrica 41: 587-601.

Gibbard, Allan (1974). A Pareto-consistent libertarian claim. Journal of Economic Theory 7: 388-410.

Hayek, Friedrich (1960). The Constitution of Liberty. London: Routledge and Kegan Paul.

van Hees, Martin (1995). Rights and Decisions: Formal Models of Law and Liberalism. Dordrecht; London: Kluwer Academic.

Hobbes, Thomas (1651/ 1962). Leviathan. London: Macmillan.

Mill, John Stuart (1869/ 1988). The Subjection of Women. Indianapolis: Hackett.

Nozick, Robert (1974). Anarchy, State, and Utopia. New York: Basic Books.

Peleg, Bezalel (1998). Effectivity functions, game farms, games, and rights. Social Choice and Welfare 15: 67-80.

Sen, Amartya (1970). The impossibility of a Paretian liberal. Journal of Political Economy 78: 152-157.

Sen, Amartya (1992). Minimal liberty. Economica, 59: 139-159.

Smith, Adam (1776/ 1976). An Inquiry into the Nature and Causes of the Wealth of Nations. Clarendon Press.

Sugden, Robert (1978). Social choice and individual liberty. In M.J. Artis and A.R. Nobay (eds) Contemporary Economic Analysis. London: Croom Helm.

Sugden, Robert (1985). Liberty, preference and choice. Economics and Philosophy 1: 213-229.

Sugden, Robert (2009). Opportunity as mutual advantage. Forthcoming in Economics and Philosophy.