Hindawi Publishing Corporation
Game Theory
Volume 2014, Article ID 742508, 11 pages
http://dx.doi.org/10.1155/2014/742508

Research Article

Renegotiation Perfection in Infinite Games

Julian C. Jamison

Office of Research, Consumer Financial Protection Bureau, 1700 G Street NW, Washington, DC 20552, USA

Correspondence should be addressed to Julian C. Jamison; julison@gmail.com

Received 25 August 2013; Accepted 19 December 2013; Published 26 February 2014

Academic Editor: X. Henry Wang

We study the dynamic structure of equilibria in game theory. Allowing players in a game the opportunity to renegotiate, or switch to a feasible and Pareto superior equilibrium, can lead to welfare gains. However, in an extensive-form game this can also make it more difficult to enforce punishment strategies, leading to the question of which equilibria are feasible after all. This paper attempts to resolve that question by presenting the first definition of renegotiation-proofness in general games. This new concept, the renegotiation perfect set, satisfies five axioms. The first three axioms—namely Rationality, Consistency, and Internal Stability—characterize weakly renegotiation-proof sets. There is a natural generalized tournament defined on the class of all WRP sets, and the final two axioms—External Stability and Optimality—pick a unique "winner" from this tournament. The tournament solution concept employed, termed the catalog, is based on Dutta's minimal covering set and can be applied to many settings other than renegotiation. It is shown that the renegotiation perfection concept is an extension of the standard renegotiation-proof definition for finite games, introduced by (Benoit and Krishna 1993), and that it captures the notion of a strongly renegotiation-proof equilibrium as defined by (Farrell and Maskin 1989).

1. Introduction

In sharp contrast with the welfare theorems of general equilibrium theory, traditional equilibrium concepts in game theory exhibit no particular tendency to pick out Pareto efficient outcomes. The various Folk theorems suggest that "coordination" may be possible in repeated games but that great many other strategy choices can also be supported as equilibria.[1] This result displays an unfortunate lack of predictive power. However, even were it possible to reach an efficient equilibrium in a one-shot game, repeated games pose another quandary: coordination generally requires the threat of punishments, and these are in turn often inefficient by their very nature. Thus if it is always possible to renegotiate to an efficient equilibrium, punishments may no longer be credible and the original equilibrium itself breaks down. Equilibria that are immune to such problems are called renegotiation-proof. The relationship between renegotiation-proofness and Pareto efficiency is exactly analogous to that between subgame perfection and Nash equilibrium: both require credibility of off-path events.

The motivation and underlying assumption for the renegotiation literature in game theory is that the group of players as a whole can (and hence will) switch to a Pareto superior outcome whenever possible. The process by which this occurs is not explicitly modeled.[2] In this paper we take an axiomatic approach to determine which equilibrium payoffs, the renegotiation perfect set, are possible outcomes in this setting. The first axiom, Rationality, restricts attention to Nash equilibria only. The second axiom, Consistency, requires that all continuation equilibria also be renegotiation perfect. The third axiom, Internal Stability, states that no renegotiation perfect equilibrium should Pareto dominate another. This axiom begins to capture the assumption above, and together these axioms define weakly renegotiation-proof (WRP) equilibria.

This relation based on Pareto dominance defines a generalized tournament structure between WRP sets. Axiom four, External Stability, demands that the set of renegotiation perfect equilibria be comparable to all other WRP sets according to the tournament solution concept. This is stability in the sense that if two sets were not distinguishable, there could

be no rationale for selecting one of them.[3] The final axiom, Optimality, given that everything is now comparable, simply picks out the best element according to the tournament ordering. This is the renegotiation perfect set, a proposed complete solution to the renegotiation problem.

The tournament solution concept used in the definition above is termed the catalog of the tournament. A generalized tournament is an incomplete order, or relation, (perhaps intransitive) with qualitative strengths between elements that are comparable. One element is said to cover another if it beats the other and everything the other beats. Dutta [1] introduces the minimal covering set, such that any other element is covered when added to this set. It extends in a straightforward manner to generalized tournaments, and, when iteratively applied, making use of the additional information conveyed by the relation strengths, yields the tournament catalog as defined here. This is a ranking rather than a set of winners only. It is applicable in situations such as voting, where strengths are determined by number of votes received, and equilibrium selection in normal-form games, where strengths might be derived from relative payoffs.

The question of renegotiation-proofness in games was first raised for the case of infinitely repeated stage games, in the contemporaneous papers of Bernheim and Ray [2] and Farrell and Maskin [3]. Both papers introduced concepts equivalent to weakly renegotiation-proof equilibria in this setting. In fact this terminology is due to the latter paper, which also defines strongly renegotiation-proof equilibria. Unfortunately these do not exist in all games, but when they do they are contained in the renegotiation perfect set. The question of renegotiation-proofness in finite games is less contentious, having been settled by Benoit and Krishna [4] using a backward induction argument. Their definition, Pareto perfection, is also captured by the renegotiation perfect concept, a useful check on its reasonability. The infinite case has proven to be more troublesome due to the inherent circularities involved, and other partial solutions have been proposed in the last decade.[4] An overview of the relevant literature follows in the next section.

One final point, concerning nomenclature, is worth making before proceeding with the paper. The renegotiation literature has perhaps been somewhat unfortunately named, in that it is not a theory of renegotiation but rather of the implications deriving from the possibility of renegotiation. This is the reason for using the generic term renegotiation-proofness: it is in fact a theory of nonrenegotiation. Indeed, since all of the games considered have complete information and no stochastic element, all actions are correctly foreseen and hence no actual renegotiation should ever take place. Other roughly synonymous terms for the idea of stability with respect to the constant possibility of unanimous deviation to an alternate and preferred feasible equilibrium are dynamic consistency and Pareto perfection. The moniker "renegotiation perfect" will be reserved for the specific definition developed in this paper.[5]

This outline of the paper is immediately followed by Section 2, which presents a more detailed look at the previous literature in the field, including where the present paper lies in relation to it. Section 3 then discusses generalized tournaments and defines the catalog solution concept. Section 4 develops the central axiomatic formulation of renegotiation perfection and states results. Section 5 briefly concludes, and finally an appendix collects proofs of the various propositions.

2. Literature

As stated previously, the idea behind the renegotiation question is that somehow the players can get together between actions in the game. At this point they are all aware of what strategy choices are directed by the current equilibrium, but we assume that it is possible for them to switch to different strategies if they can believably agree to do so. This can be because of explicit opportunities to communicate,[6] or it can be due to more indirect machinations. In any case, the vital problem is that of determining believability, which has three attributes: first, that the new choices constitute an equilibrium; second, that the associated payoffs Pareto dominate the payoffs expected according to the current equilibrium (else why switch?); and third, that the proposed new equilibrium be immune to the same sort of deviations (else how is it credible?). In infinitely repeated games, this introduces an obvious circularity and so the problem is born. Note that this discussion is implicitly assuming a kind of stationarity, in that the solution to the problem will be a single set of renegotiation-proof equilibria to be considered feasible at any time. This set depends only on the structure of the game and not, for example, on the history of play.[7] In particular, no preference is given to the current equilibrium, unlike in several of the papers discussed below.[8] The justification for this approach is that an equilibrium concept is being defined, rather than say a process. One would not expect the set of Nash equilibria in identical subgames to depend on the history of actions so far.

One paper that does discuss an explicit model of communication is Blume [5], which also considers bargaining over continuation payoffs. The focus in this paper is over the bargaining outcomes rather than the renegotiation question. It treats primarily the case of finite games, although the model is briefly extended to cover the infinite case as well. Unfortunately, existence cannot be guaranteed in the infinite setting, which is a common drawback arising from the circular nature of the problem. It also differs from the present paper by imposing costs on bargaining. Another paper looking at bargaining instead of Pareto jumps between equilibria is that of Abreu et al. [6]. This approach is better suited to some situations than to others, and naturally more assumptions need to be made about the explicit nature of the process. The analysis here is more general (apart from spelling out the details of the bargaining game) but also more difficult, tending to yield more limited results. Something like it, however, would be an important addition to the model of the present paper. Finally, this is related to the coalition-proofness concept, developed by Bernheim et al. [7], in which some of the same questions arise as in renegotiation models.

Returning to the slightly simpler, baseline case of unanimity (Pareto dominance) and stationarity ("let bygones

be bygones"), we start with finitely repeated games. Here the definitive paper is Benoit and Krishna's [4], which gives a full characterization of the Pareto perfect equilibria and discusses behavior as the time horizon lengthens indefinitely. Just as with subgame perfection, the definition proceeds backward step by step, recursively requiring efficiency subject to the constraints imposed by the set of feasible continuation equilibria. These equilibria turn out to be neither generically efficient nor generically inefficient within the entire SPE set. Renegotiation perfection, which applies to all finite games, coincides with Pareto perfection in finitely repeated games. In this sense, it is an extension of the finite definition to the infinite case. A scattering of other works has been done on renegotiation-proofness in finite games; for instance, Wen [8] centers on finite games with more than two players.

The literature on renegotiation-proofness in infinitely repeated games began with the simultaneous (and neighboring) papers of Bernheim and Ray [2] and Farrell and Maskin [3], predating and stimulating the work on finite games. The increase in difficulty for the infinite case arises from the fact that it is necessary to determine the set of renegotiation-proof equilibria at the same time that the set of credible continuation equilibria is determined. If in a particular game the question of feasibility had been previously resolved for all equilibria but one, the problem would be trivial for that one, by comparing it with the feasible equilibria and checking that its continuation equilibria were all feasible. But to do this for all equilibria simultaneously is more troublesome.

Bernheim and Ray [2] define internally consistent equilibria to be those such that the equilibrium and all of its continuation equilibria are mutually Pareto incomparable. They then strengthen this to the concept of consistent sets, which always exist. Unfortunately there may be multiple consistent sets, and elements of one may dominate elements of another, so it is not clear how to interpret the choice of any particular one of them. This is a case for which the stationarity assumption will not be satisfied.

Farrell and Maskin [3] introduce the concept of a weakly renegotiation-proof (WRP) set and equilibrium. Their definition is equivalent to internal consistency, and both are equivalent to axioms one through three below when applied to infinitely repeated games.[9] They give a fairly full characterization of WRP sets, including generic intersection with the Pareto frontier, and analyze examples of WRP equilibria in various games. They too impose an external condition, deriving a new concept called strongly renegotiation-proof (SRP). A set is said to be SRP if it is WRP and if no element is dominated by any element of a WRP set. This avoids any circularity but is conceptually an overly strict condition since a WRP set may be excluded on the basis of another WRP set that is in turn also excluded. Indeed, SRP equilibria often fail to exist in practice. Any SRP equilibrium that does exist is also renegotiation perfect, as shown in Section 4 below. So renegotiation perfection is an extension of this idea that exists for all games. This also implies that their characterization results about SRP sets (and WRP sets) can be carried over to the renegotiation perfection context.

Abreu and Pearce [9] and Pearce [10] also study renegotiation in infinite games, but with quite a different approach. They do not require internal stability for an equilibrium to be renegotiation-proof. The reason that this can be consistent is that stricter conditions are placed on proposed deviations than are placed on the current equilibrium. So punishments may still be credible, since it is no longer necessarily possible to return to the original equilibrium. This is of course a departure from the stationarity assumption outlined above, and it is also the case that existence fails here for certain types of games. However, in some situations this provides a truly viable alternate model of renegotiation.

Bergin and MacLeod [11] synthesize much of the literature (up to that point), including both the finite and infinite cases. They provide an almost complete axiomatic framework, in order to relate previous work, which unfortunately but inevitably involves considerable notation. They also have a brief section on the explicit modeling of communication. The only new concept introduced, recursive efficiency, again gives some precedence to the status quo, though it has some other desirable qualities. One reason for the prevalence of such nonstationary approaches is that they help to alleviate the circularity and self-comparison problems.

It is worth pointing out that there is an entire renegotiation literature in contract theory as well. One of the main distinctions between it and the game theory literature, of course, is that the principal is generally choosing the contract. The effect of allowing renegotiation is therefore always negative (from the principal's point of view), even if renegotiation only occurs with the consent of both parties, because avoiding it becomes simply one more constraint in the principal's optimal contract problem. This result is roughly consistent with the game theoretic result that it may be impossible to reach the Pareto frontier if renegotiation is allowed. The contracts literature is often more general along the dimension of allowing asymmetric information, stochastic output processes, and so forth (which implies that renegotiation may actually take place in equilibrium). On the other hand, it tends to be limited to a finite number of periods, usually two, and is thus of only modest help in the present effort.

3. Tournaments

In this section we examine some aspects of tournament theory and define a generalized tournament. Iterative applications of the minimal covering set, along with suitable extensions thereof, allow us to distinguish and then rank elements as finely as possible. One component of generalized tournaments, the relative strength of the relation between elements that are in fact comparable, is naturally helpful in this endeavor. The result is called the catalog of the tournament and is of independent interest as well as of interest to applied work in various fields. Of course, it is developed here specifically in preparation for the analysis of renegotiation in the next section. The efficient reader can focus on Definitions 1–4 and the result of Proposition 11 as the minimally necessary background for the remainder of the paper.

Definition 1. A tournament **T** is a pair (T, R) such that

(a) T is a finite set;

(b) R is a binary relation on T;

(c) $\nexists x \in T$: xRx;

(d) $\forall x, y \in$ either xRy or yRx but not both.

The motivation for the name lies in the original story of a group of contestants, vying for a prize, who meet pairwise with each match producing exactly one winner. A more practical example giving rise to the same situation is a voting model in which there is a selection of alternatives and it is known how each would fare against any other single alternative.[10] In both cases, the problem is usually how to pick a winner or set of winners, and there is a multitude of solution concepts available in the tournament literature. Naturally, if x is a Condorcet winner, that is, if $xRy \forall y \in T/x$, then x is the unique winner according to any of the solution concepts. The difficulty arises when there is no such x.

Definition 2. A generalized tournament **T** is a triple (T, R, λ) such that

(a) R is a binary relation on T;

(b) $\nexists x \in T : xRx$;

(c) λ is a label function from $\{(x, y) \in T^2 \mid xRy\}$ to the real line $\mathfrak{R}$;

(d) if $\{x_n\} \subseteq T$ with $\lim_{n\to\infty} x_n = x$ and $x_n R y \forall n$ (resp., $y R x_n \forall n$), then $x \in T$ and xRy (resp., yRx).

The last part of the definition, which is trivially satisfied if T is finite, is a technical condition that basically says that T should be closed in some topology and the tournament structure (as given by R) should be continuous with respect to that topology. This extends the standard tournament concept in three directions: first, elements are permitted not to be comparable (i.e., to tie); second, elements that are comparable have a label on their relationship, which can be interpreted as a qualitative assessment of the strength of the comparison; and third, infinite tournaments are allowed, as long as they satisfy the basic "niceness" property of condition (d). For instance, in the voting context ties are no longer a problem, and the labels might correspond to the relative number of votes for one option when paired against another, incorporating more information into the model. In Section 4 below, a method for deriving a generalized tournament from an arbitrary binary relation is described, but for now we simply assume that we are given one from whatever source and begin the process of comparing and ranking elements.

Definition 3. For $x, y \in T$, by x *covers* y (written $x \succ y$) we understand that xRy and that $\forall z \in T$ $yRz \Rightarrow xRz$. The uncovered set $\mathrm{UC}(T)$ is then $\{x \in T \mid \nexists y \in T : y \succ x\}$.

The idea is that if one element covers another, this implies an unambiguous comparison between the two in favor of the first one: it beats the second and it beats everything the second one beats as well. The uncovered set is simply the set of maximal elements with respect to this order. Since **T** induces a canonical relation on any subset of T, we can define $\mathrm{UC}^2(T) = \mathrm{UC}(\mathrm{UC}(T)), \mathrm{UC}^3(T), \ldots, \mathrm{UC}^\infty(T)$, known as the iterated uncovered sets. However it is possible to refine this idea even further.

Definition 4. The minimal covering set $\mathrm{MC}(T)$ is the smallest subset of T satisfying

$$\forall x \in T \setminus \mathrm{MC}(T), \quad x \notin \mathrm{UC}(\mathrm{MC}(T) \cup \{x\}). \tag{1}$$

When applied to standard tournaments, this is equivalent to the definition given by Dutta [1]. It says that any other element, when added to this set, will be covered by some element of the set. In some sense, these are the core stable elements, at least with respect to the covering relation and, for now, excluding information from the labels. The minimal covering set has many desirable properties for a tournament solution concept, several of which will be needed in the sequel and are therefore stated here.

Proposition 5. *It is shown as follows.*

(a) *Covering is a transitive relation.*

(b) *$MC(T)$ exists and is well defined for all generalized tournaments* **T**.

(c) *$MC(T) \subseteq UC^\infty(T)$.*

(d) *$MC(MC(T)) = MC(T)$, and hence $UC(MC(T)) = MC(T)$.*

(e) *MC is monotonic; that is, if $x \in MC(T)$ and R is changed so as to strengthen x (leaving all else unchanged), it is still the case that $x \in MC(T)$.*

(f) *MC is independent of losers; that is, neither deleting any $x \notin MC(T)$ nor changing R outside of $MC(T)$ has any effect on $MC(T)$.*

(g) *the minimal covering set can be characterized axiomatically.*

The idea is to now extend the minimal covering set solution in two ways. The first of these is to get a full ranking rather than a set of winners (along the lines of looking at $\mathrm{MC}(T \setminus \mathrm{MC}(T))$, and so on). The second is to refine these sets further, in part by using the label information that is available. This is because it is necessary to distinguish and rank elements as discriminately as possible. The approach is to define two further relations on the same set T, without labels, and to use their minimal covering sets as necessary in an iterative procedure that continuously breaks T down into disjoint subsets, with a complete ordering between the subsets. This will be the tournament catalog. For clarity, if $A \subseteq T$ we write $\mathrm{MC}^R(A)$ for the minimal covering set of A according to the relation on A induced by R on T.

Definition 6. The binary relation S_A on $A \subseteq T$ is given by

$$\forall x, y \in A,$$

$$x S_A y \quad \text{iff } x \in \mathrm{MC}^R(A \setminus \{y\}) \text{ but } y \notin \mathrm{MC}^R(A \setminus \{x\}). \tag{2}$$

In essence, this says that x is superior to y, according to this relation, whenever removing x means that y is not a "winner," but not vice versa. That is, x does not depend on y in order to argue that it should be chosen. If x, but not y, is in $\mathrm{MC}(A)$ then xS_Ay holds trivially, but this relation is particularly useful for comparing elements within the minimal covering set itself without introducing any new concepts. It is this refinement that we are interested in. Naturally, we denote minimal covering sets with respect to this relation by MC^S, with the subscript on S implied by the argument to the MC operator.

Definition 7. If $\mathbf{A} = \{\mathbf{A}_i\}_{i=1}^n \subseteq T^n$ such that $A_i \subseteq A_{i+1}$ for all $i = 1, \ldots, n-1$, the binary relation L_A on A_1 is given by for all $x, y \in A_1 xL_Ay$ if and only if there exists i such that either

(a) for all $z \in A_i : xRz$ and yRz, $\lambda(x,z) \geq \lambda(y,z)$ with at least one strict inequality

or

(b) $\{z \in A_i \mid yRz\} \subset \{z \in A_i \mid xRz\}$ and for all $z \in A_i$: yRz $\lambda(x,z) \geq \lambda(y,z)$.

This third relation makes use of the strength assessments to differentiate even further among elements. Given a nested sequence of sets, we compare elements of the innermost set by checking to see if one "weakly dominates" the other in label strength over the set of things beaten by both or if one beats strictly more elements than the other and at least as strongly. To allow the widest room for comparison, the domination can take place at any superset in the given sequence.[11] Because direct comparisons between elements according to the basic relation R are more relevant to choosing a winner, the label relation, denoted by a superscript L, is reserved for third-string efforts to produce a ranking (recall that the goal is to distinguish as finely as possible). In other settings, however, it is possible that labels would be the primary or only means of comparison.

We are now ready to give the formal, recursive algorithm that produces the tournament catalog. Fix a generalized tournament $\mathbf{T}$ and consider subsets A_σ of T, where $\sigma \in [0,1]^{<\infty}$ is a finite list of numbers in the unit interval. Notationally, we take $A_\varnothing = T$. By $\sigma|_n$ we understand the restriction of σ to its first n components, and by σ, α for $\alpha \in [0,1]$ the concatenation of σ and α, therefore, is also a point in $[0,1]^{<\infty}$.

Definition 8. Given $A_\varnothing \subseteq T$, the process $\wp$ is as follows.

(1) If $\mathrm{MC}^R(A_\sigma) = A_\sigma$, go to (2).

Otherwise, pick an index set $A_\sigma \subseteq [0,1]$ containing $\{0,1\}$ and define

$A_{\sigma,\beta} = \mathrm{MC}^R(\cup\{A_{\sigma,\alpha} \mid \alpha \leq \beta, \alpha \in \beta_\sigma\})$, so that in particular
$A_{\sigma,1} = \mathrm{MC}^R(A_\sigma)$. Now apply $\wp$ to $A_{\sigma,\beta}$ for each $\beta \in B_\sigma$.

(2) If $\mathrm{MC}^s(A_\sigma) = A_\sigma$, go to (3).

Otherwise, pick an index set $A_\sigma \subseteq [0,1]$ containing $\{0,1\}$ and define

$A_{\sigma,\beta} = \mathrm{MC}^s(\cup\{A_{\sigma,\alpha} \mid \alpha \leq \beta, \alpha \in \beta_\sigma\})$, so that in particular
$A_{\sigma,1} = \mathrm{MC}^s(A_\sigma)$. Now apply $\wp$ to $A_{\sigma,\beta}$ for each $\beta \in B_\sigma$.

(3) Let $n = |\sigma|$ and set $A_i = A_{\sigma|_{n+1-i}}$ for $i = 1, \ldots, n+1$, so that $\mathbf{A} = \{\mathbf{A}_i\}_{i=1}^{n+1}$ is a nested sequence of sets in T and L_A is thereby defined.

If $\mathrm{MC}^L(A_\sigma) = A_\sigma$, apply $\wp$ to $A_{\sigma,1} = A_\sigma$.

Otherwise, pick an index set $B_\sigma \subseteq [0,1]$ containing $\{0,1\}$ and define

$A_{\sigma,\beta} = \mathrm{MC}^L(\cup\{A_{\sigma,\alpha} \mid \alpha \leq \beta, \alpha \in \beta_\sigma\})$, so that in particular
$A_{\sigma,1} = \mathrm{MC}^L(A_\sigma)$. Now apply $\wp$ to $A_{\sigma,\beta}$ for each $\beta \in B_\sigma$.

This completes the definition. The idea is that, at each step, we first look to see if the standard minimal covering set has any effect. If it does, we use it to divide the original set into disjoint subsets, each of which is the minimal covering set for everything up to and including itself. These sets are naturally ordered, and their indices β retain that order. We then apply $\wp$ iteratively to each subset, attempting to refine further. If MC^R has no effect, we turn to the relation S and perform the same check, proceeding analogously if possible. Finally, the label relation L is employed, and if it fails then the process gives up. In nonpathological cases, very few steps will not be vacuous. If we apply $\wp$ to the whole tournament, that is, to $A_\varnothing = T$, then for each $x \in T$ and $n < \infty$ there is exactly one $\sigma_n(x)$ with $|\sigma| = n$ and $x = A_\sigma$. Furthermore, these are consistent in the sense that if $n < n'$ then $\sigma_n(x) = \sigma_n(x)|_n$, and hence $\sigma(x) \in [0,1]^\infty$ is well defined as the "limit" of $\sigma_n(x)$'s.

Definition 9. For a generalized tournament $\mathbf{T}$, the catalog order is given by

$$\forall x, y \in T x <_c y \Longleftrightarrow \sigma(x) <_{\mathrm{lex}} \sigma(y). \tag{3}$$

The tournament catalog is the pair $(T, <_c)$.

Of course, $<_{\mathrm{lex}}$ refers to the lexicographic order. It is appropriate in this case because the first component of $\sigma(x)$ refers to the first step in the process $\wp$, which is more authoritative in the sense that later steps are refinements of it. So elements are compared first on this basis, next on the second step, and so on. Of course, several elements may induce the same sequence σ; they are indistinguishable by $\wp$. The tournament catalog is a "pseudochain": two elements may not be comparable, but in this case the set of successors (or predecessors) of one of them is identical to that of the other one.[12] In some contexts, there is a distinguished element $x^0 \in T$, such as the status quo in voting or more generally any default or focal point. By definition, this element can be differentiated from all others.

Definition 10. Let $I(\mathbf{T})$ be the closure of $\{x \in T \mid \nexists y \in T : \sigma(x) = \sigma(y)\} \cup \{x^0\}$.

Then the catalog index for $\mathbf{T}$ is the pair $(I(\mathbf{T}), <_c)$.

The index of a tournament thus picks out all elements that can be unambiguously compared with every element of the tournament (not just of the index),[13] and it is maximal with respect to this property. It is in fact completely ordered, as seen below.

Proposition 11. *It is shown as follows.*

(a) *The tournament catalog and its index are well defined.*

(b) *The catalog index is a chain.*

This concludes the presentation of the catalog (and its index) as a solution concept for generalized tournaments. In the following section, this concept is utilized to help develop a notion of renegotiation-proofness (and some examples of catalogs are given). However, it has many other applications, including the simple voting context used throughout the course of the discussion. One other possible application is to the equilibrium selection problem in normal-form games. In this case we might say that xRy if a majority (or two-thirds or all) of the players prefer outcome x to outcome y, and the labels are based on individual payoff differences between the equilibria. The motivation for such a model is that rational players, if faced with a selection problem, should try to pick out the "best" (i.e., winning in the terminology of tournaments) equilibrium and base their actions choices on that decision, assuming that others are doing likewise.[14]

4. Renegotiation

This section develops the central model of the paper. Five axioms are discussed, which together lead to the definition of renegotiation perfect equilibria. The axioms attempt to capture some of what is meant by being renegotiation-proof or immune to the possibility of unanimous deviations to Pareto superior alternatives. Examples are given, and formal relationships with previous work are outlined. The latter conclusions, that renegotiation perfection generalizes previous concepts, allow direct referral to some known characterization results.

We consider arbitrary n-player extensive-form games with perfect recall and at least one Nash equilibrium;[15] the set of all such games is called $\mathbf{G}$. The strategy space of a game $G \in \mathbf{G}$ is denoted by $\Sigma(G) = \times_{i=1}^{n} \Sigma_i(G)$, and payoffs are mappings $u_i : \Sigma(G) \to \mathfrak{R}$. We let $\mathrm{NE}(G)$ be the set of Nash equilibria of G and let $u(\mathrm{NE}(G)) \subseteq \mathfrak{R}^n$ be the associated set of Nash payoffs. If the action history h leads to a singleton information set (such as between stages in a stage game with perfectly observed actions), this begins a well-defined subgame $G' = G|_h$, and then a strategy profile $\sigma \in \Sigma(G)$ induces continuation strategies $\sigma' = \sigma|_h \in \Sigma(G')$ and continuation payoffs $u(\sigma') \in \mathfrak{R}^n$. The axioms are on set-valued correspondences $\mathbf{R} : \mathbf{G} \Rightarrow \mathfrak{R}^n$. That is, for all $G \in \mathbf{G}$, $\mathbf{R} : G \mapsto \mathbf{R}(g) \subseteq \mathfrak{R}^n$. Consider

A.1 (Rationality) $\forall G \in \mathbf{G}, \mathbf{R}(G) \subseteq u(\mathrm{NE}(G))$.

The first axiom simply states that all payoffs in the $\mathbf{R}$-set should be supportable by Nash equilibria. Since, ultimately, the goal is to define a new equilibrium concept addressing renegotiation concerns but within the existing framework; this requirement is straightforward. Consider

A.2 (Consistency) Any $u \in \mathbf{R}(G)$ equals $u(\sigma)$ for some $\sigma \in \Sigma(G)$ such that if $\sigma' = \sigma|_h$ for some history h, then $u(\sigma') \in \mathbf{R}(G')$, where $G' = G|_h$ is the induced subgame.

The second axiom is likely equally uncontroversial, asking for a basic dynamic consistency in the correspondence $\mathbf{R}$. The motivation is that if an equilibrium requires the use of continuation payoffs which are not themselves feasible (according to $\mathbf{R}$), then they will be unable to effectively enforce the original equilibrium, rendering it infeasible as well. This is analogous to the credibility arguments supporting the subgame perfection concept, and indeed it is clear that the first two axioms together imply that all payoffs in the $\mathbf{R}$-set are supportable with subgame perfect equilibria.

Definition 12. If $A, B \subseteq \mathfrak{R}^n$ then A *confounds* B, written $A \rhd B$, if $A \cap [B + \mathfrak{R}^n_{++}] \neq \varnothing$.

This is equivalent to saying that there is some element of A that strictly Pareto dominates some element of B. This is a key notion in the context of renegotiation since the underlying assumption is that there is always the possibility of jumping to Pareto superior points.[16] That is, if players are expecting payoffs from a set B and they may have an incentive to switch to the set A, calling into question the believability of ever expecting B. Consider

A.3 (Internal Stability) $\forall G \in \mathbf{G}, \mathbf{R}(G) \not\rhd \mathbf{R}(G)$.

Axiom three demands that there be no strict Pareto dominance within the $\mathbf{R}$-set itself.[17] The rationale for the axiom then follows from the definition and the fact that the central notion under consideration is credibility: anything in this set is by assumption credible and is therefore a viable alternative at any time. Hence anything dominated by such a point will be renegotiated away from and cannot itself be renegotiation-proof, which is precisely the statement of the axiom. Put slightly differently, the set $\mathbf{R}(G)$ forms a Pareto antichain.

Definition 13. If $\mathbf{R}$ satisfies A.1–A.3, then $\mathbf{R}(G)$ is a weakly renegotiation-proof set in G.

These WRP sets form the building blocks for much of the rest of the discussion. In the context of infinitely repeated stage games, it is easy to see that this definition coincides with that of Farrell and Maskin [3], from whom the terminology is taken. It is thus also equivalent, in that context, to the notion of internally consistent sets as introduced by Bernheim and Ray [2]. For a given game G, we let $W(G)$ be the set of all WRP sets in G. There is typically a multitude of WRP sets in any particular game, some of which may be preferred to others by all players in all circumstances. This lack of external comparison (along with the lack of uniqueness) makes weak

renegotiation-proofness an insufficient concept. Farrell and Maskin characterize all equilibria that can be supported with WRP payoffs and show that generically some of these are efficient.

In order to make comparisons within the class of WRP sets, we shall use the tournament index concept developed in the previous section. This requires a relation on the set $W(G)$ to make it into a generalized tournament. The basic comparison available is whether or not one set confounds another, directly or indirectly (i.e., through a sequence of sets, each confounding the next).[18] Formally, fixing G and for $x, y \in W(G)$ define $k(x, y) = \min\langle n \mid \exists x = x_0, x_1, \ldots, x_n = y \text{ s.t. } x_i \in W(G) \forall i \text{ and } x_i \rhd x_{i+1} \forall i = 0, 1, \ldots, n-1\rangle$, with $k(x, y) = \infty$ if no such n exists. We also make use of the set of predecessors of x in $W(G)$, namely, $\rhd^+(x) = \{y \in W(G) \mid y \rhd x\}$.

Definition 14. Fix G and let $x, y \in W(G)$ be WRP sets in G. Then xRy if

(a) $k(x, y) < k(y, x)$;

(b) $k(x, y) = k(y, x)$ but $x \supset y$; or

(c) $k(x, y) = k(y, x), x \not\subseteq y$ but $\rhd^+(x) \subset \rhd^+(y)$.

Thus to compare x and y, we first check to see if one confounds the other more directly. Recall the definitions and motivations: this means that it is easier to renegotiate (by way of Pareto jumps) from y to x than it is from x to y. For instance, it may be that x indirectly confounds y in a minimal chain of three steps but y cannot indirectly confound x at all. Then x is in a stronger, or more dominant, position and so xRy. If on the other hand this comparison cannot be made, then a strictly larger set is preferable because (other aspects being equal) we wish to be as open-minded as possible when considering whether or not certain payoffs are renegotiation-proof.[19,20] Finally, since once again the goal is to use as much information as possible to distinguish elements (in this case WRP sets), a simple comparison between predecessors can be made. One set is in a stronger position than another if everything that confounds it also confounds the other but not vice versa.[21]

Definition 15. The renegotiation tournament $\mathbf{T}(G)$ is $(W(G), R, \lambda)$ where $\lambda(x, y) =$

(a) $(-k(x, y))/k(y, x)$ if $k(y, x) < \infty$ and $1/k(x, y)$ otherwise,

(b, c) 0 anytime that $k(x, y) = k(y, x)$.

Of course this also yields a catalog order $<_c$ on $W(G)$ and a completely ordered index $I(\mathbf{T}(G))$ for the renegotiation tournament. We define the distinguished element x^0 to be the Pareto frontier of the convex hull of payoffs achievable by myopically optimal subgame perfect equilibria.[22] These, even if nothing else can be agreed upon, are certainly feasible and can be differentiated. As stated previously, it would also be possible to define x^0 as the payoffs arising from the current (i.e., status quo) equilibrium, as long as they formed a WRP set. With this firmly in hand, we are now ready to state the final two axioms of renegotiation perfection. Consider

A.4 (External Stability) $\forall G \in \mathbf{G}, \mathbf{R}(G) \in I(\mathbf{T}(G))$.

The fourth axiom basically requires that the $\mathbf{R}$-set always be comparable in the catalog order with all other WRP sets in G. This is the essence of being externally stable: if it were not distinguishable then by definition it would be at best an arbitrary choice. The true danger here lies in the fact that an equally valid argument could then be made in favor of any WRP set not comparable with it, meaning it could be renegotiated away from if the possibility arose and thus violating the whole motivation.

A.5 (Optimality) $\forall G \in \mathbf{G}$ and $x \in I(\mathbf{T}(G)), x \not>_c \mathbf{R}(G)$.

The final axiom states that no distinguishable (i.e., externally stable) WRP set should be greater in the catalog order than the $\mathbf{R}$-set. If there was such a "better" set, it would not be rational for optimizing players to accept only the $\mathbf{R}$-set, again violating the motivation that we are attempting to list certain properties in agreement with our sense of what renegotiation-proof sets ought to behave like. Separating the external requirements into these two distinct conditions turns out to be, perhaps somewhat surprisingly, one of the most clarifying aspects of this approach. We list the axioms once more in their entirety for the sake of completeness:

A.1 (Rationality) $\forall G \in \mathbf{G}, \mathbf{R}(G) \subseteq u(\mathrm{NE}(G))$

A.2 (Consistency) Any $u \in \mathbf{R}(G)$ equals $u(\sigma)$ for some $\sigma \in \Sigma(G)$ such that if $\sigma' = \sigma|_h$ for some history h, then $u(\sigma') \in \mathbf{R}(G')$, where $G' = G|_h$ is the induced subgame.

A.3 (Internal Stability) $\forall G \in \mathbf{G}, \mathbf{R}(G) \not\rhd \mathbf{R}(G)$.

A.4 (External Stability) $\forall G \in \mathbf{G}, \mathbf{R}(G) \in I(\mathbf{T}(G))$.

A.5 (Optimality) $\forall G \in \mathbf{G}$ and $x \in I(\mathbf{T}(G)), x \not>_c \mathbf{R}(G)$.

Definition 16. A correspondence $\mathbf{R} : \mathbf{G} \Rightarrow \Re^n$ is renegotiation perfect if it satisfies A.1–A.5. In this case, $\mathbf{R}(G)$ is called the renegotiation perfect set (of payoffs) for G, denoted by $\mathrm{RP}(G)$, and a strategy profile $\sigma \in \Sigma(g)$ is a renegotiation perfect equilibrium of G if, for all histories h (inducing strategies σ' on the subgame G'), $u(\sigma') \in \mathbf{R}(G')$.

Proposition 17. $\forall G \in \mathbf{G}$, *$RP(G)$ is nonempty and well defined (i.e., unique).*

Thus a renegotiation perfect equilibrium is a strategy profile such that all of its continuation payoffs are in the appropriate renegotiation perfect set, and these equilibria exist for all extensive-form games (at least those with a Nash equilibrium). The axioms require that renegotiation perfect sets be self-enforcing and resistant to counter proposals from both within the set and outside it. It is also possible, without invoking an entire cooperative framework, to interpret the axioms in a societal manner; for instance, A.1 is individual rationality, while A.5 is social rationality.[23]

We turn next to the particular setting of finitely repeated stage games, for which there exists already an accepted definition of renegotiation-proofness. This is the Pareto perfection

concept developed by Benoit and Krishna [4].[24] To briefly recap their definition, let G be a one-shot game[25] with a Nash equilibrium and let $P^1 = \text{WEff NE}(G)$, the set of weakly efficient Nash equilibria of G. Next let $Q^2 \subseteq \Sigma(G^2)$ be the set of all subgame perfect equilibria of G^2 that use only continuation equilibria in P^1; with two periods remaining nothing else is credible, given that otherwise renegotiation to some element of P^1 would occur in the final period. Since optimality must hold at this stage as well, they define $P^2 = \text{WEff}\, Q^2$. By iterating this recursive efficiency procedure, one reaches P^m, the set of Pareto perfect equilibria of the m-repetition G^m.

Proposition 18. *For any one-shot game G, $RP(G^m) = u(P^m)$ $\forall m = 1, 2, \ldots$.*

This result and its proof imply that, in finitely repeated stage games, all Pareto perfect equilibria are renegotiation perfect, and vice versa. In this sense, renegotiation perfection is an extension (there may be several) of Pareto perfection to infinite games. Since the latter concept is relatively uncontroversial, this provides a good plausibility check on the former concept. The intuition behind the proof of the proposition, which uses induction to show that $u(P^m)$ satisfies A.1–A.5, is straightforward. Axioms one and three are trivial (minimal WRP sets are all singletons here), axiom two follows from the inductive hypothesis, and axioms four and five follow from the fact that $u(P^m)$ is in fact a Condorcet winner in the renegotiation tournament. To look at it in the other direction, axioms one through three imply that only subgame perfect equilibria with renegotiation perfect continuation payoffs can be considered, that is, Q^m, and then axioms four and five narrow it down to the Pareto frontier of this set, that is, P^m. Everything tends to be simpler in the finite case, but the underlying processes are identical to the general case.

We next briefly consider the particular case of infinitely repeated games, which has occupied much of the attention in the renegotiation literature. One partial solution in this context was introduced in the original paper by Farrell and Maskin. In the present terminology, it can be stated as follows.

Definition 19. A set $x \in W(G)$ is strongly renegotiation-proof (SRP) in G if there exists no $y \in W(g)$ such that $y \rhd x$.

That is, no element of the WRP set x should be Pareto dominated by any element of any other WRP set. In terms of logical consistency of the definition, this requirement is definitely too strict (as the authors explicitly acknowledge in their choice of a name), since it rules out sets simply because they are confounded by other sets that may or may not be SRP themselves. This avoids many of the circularity problems but also avoids the underlying motivation that only jumps to continuation payoffs that are themselves feasible should constitute valid objections or deviations. However, these equilibria are easier to define and, precisely because the constraint is overly strict, any payoff that is supported by such an equilibrium is indisputably renegotiation-proof.[26] The problem, of course, is that there may exist no such equilibria at all, which again runs counter to the idea behind models of renegotiation-proofness that they are a search for what outcomes might reasonably be agreed upon by rational players in this setting.

Proposition 20. *If G is an infinitely repeated game and x is SRP in G, then $x \subseteq RP(G)$.*

This result says that SRP payoffs, if they exist, are also renegotiation perfect. Since SRP equilibria (supported entirely by SRP payoffs) do exist in some games of economic interest, and since Farrell and Maskin [3] are able to show some properties of SRP equilibria (including some results about when they exist), this is a useful proposition in terms of better understanding renegotiation perfect equilibria. Characterizing its formal relationship with other proposed concepts for infinite games would be a logical next step in this regard.

5. Conclusion

This paper has presented a new definition, renegotiation perfect equilibrium, for extensive-form games. The problem it attempts to address is of what happens if players are given the opportunity to "renegotiate" to Pareto superior payoffs between actions in the game. Given that this is the case, punishments may no longer be credible. Rational players will be able to foresee this problem and will avoid equilibria which cannot be supported, but what, in the end, is the set of feasible equilibria? This is an important question since it merges the off-path credibility requirements epitomized by subgame perfection with the social choice arguments behind Pareto optimality. At least in the self-referential world of infinitely repeated games, it has also proven to be a difficult question to address. While renegotiation perfection does not have all the answers, it is the first proposed solution that applies in all contexts and exists for all games.

The innovation in the present axiomatic approach is twofold: first, the application of tournament theory to the renegotiation problem (including the development of the catalog, a new solution tool for generalized tournaments that is independently useful); and second, considering the questions of external stability (i.e., distinguishability) and optimality (i.e., maximality) separately. The main results are that renegotiation perfect payoffs exist for all extensive-form games, finite and infinite, and that they capture key previous concepts from the literature.

Naturally, the analysis of renegotiation perfection is incomplete. It would be beneficial to have true formal models of the underlying communication and bargaining processes (including perhaps adding the possibility of coalition formation). A fuller characterization of renegotiation perfect payoffs, both in specific games and in general, would lead to a deeper understanding of the concept. But quite possibly the most fruitful investigations, however, are in the field of stochastic games. It is only in this context that actual renegotiation can be expected to take place in equilibrium, so that the goal is to eventually have not just a theory

of renegotiation-proofness but rather an actual working theory of renegotiation. Such an extension would provide an excellent opportunity to test the implications and the relative merits of any competing theories of renegotiation-proofness. This paper attempts to provide one step in the progression toward these future endeavors.

Appendix

Proof of Proposition 5. (a) Suppose that $x \succ y$ and $y \succ z$. Then yRz so xRz since x covers y. Furthermore if zRw then yRw by $y \succ z$, and hence xRw by $x \succ y$. Thus $x \succ z$.

(b) A covering set in $\mathbf{T}$ is any subset of T satisfying the property in the definition of minimal covering set. It is clear that these exist (e.g., T itself is one) and it can be shown (e.g., following the proof in [12]) that the arbitrary intersection of covering sets is again a covering set. Thus there is a unique minimal covering set. Note that the empty set is never a covering set, so for all T, $\mathrm{MC}(T) \neq \varnothing$.

(c) If $x \notin \mathrm{UC}^{\infty}(T)$ then there exist y, k such that y covers x in $\mathrm{UC}^{k}(T)$ and hence in any subset of $\mathrm{UC}^{k}(T)$. If $y \notin \mathrm{UC}^{\infty}(T)$ then there exist z, $k' > k$ such that $z \succ y$ in $\mathrm{UC}^{k'}(T)$. Since $\mathrm{UC}^{k'}(T) \subseteq \mathrm{UC}^{k}(T)$, $z \succ x$ also in $\mathrm{UC}^{k}(T)$. Continuing, we find some $w \in \mathrm{UC}^{\infty}(T)$ with $w \succ x$ in some superset of $\mathrm{UC}^{\infty}(T)$, and so $w \succ x$ in $\mathrm{UC}^{\infty}(T) \cup \{x\}$. This proves that $\mathrm{UC}^{\infty}(T)$ is a covering set, and the result follows.

(d) The first statement is clear from the definition (in particular, from minimality) and the second statement is then implied by part (c).

(e) Let S be the new relation in which x has been strengthened (but nothing else has been changed). We know that $x \in \mathrm{MC}^{R}(T)$ and, letting $A = \mathrm{MC}^{S}(T)$, assume toward a contradiction that $x \notin A$. But then R and S agree on A, so $\mathrm{UC}^{R}(A) = A$ (since certainly $\mathrm{UC}^{S}(A) = A$). Now if $y \notin A$ and $y \neq x$ then R and S agree on $A \cup \{y\}$ so, by definition of A, $y \notin \mathrm{UC}^{R}(A \cup \{y\})$. On the other hand, $x \notin \mathrm{UC}^{S}(A \cup \{s\})$ because A is a covering set under S, and x is weaker under R, so $x \notin \mathrm{UC}^{R}(A \cup \{s\})$ either. Together this implies that A is a covering set under R, but since A does not contain x and $x \in \mathrm{MC}^{R}$, this is a contradiction.

(f) The only relations that arise in the definition involve members of $\mathrm{MC}(T)$ in at least one of the two positions, from which the result follows directly.

(g) Analogous to Laslier [12, pages 117–120], mutatis mutandis. □

Proof of Proposition 11. (a) It suffices to show that the catalog order $<_c$ is well-defined. Clearly the index sets B_σ in the process $\wp$ may be chosen in more than one way, but we show that the induced decomposition is unique. Indeed, suppose that $\{A_{\alpha,\beta}\}_{\beta \in B_\varnothing}$ and $\{A_{\alpha,\beta'}\}_{\beta' \in B'_\varnothing}$ both satisfy the conditions of the definition for $\wp$ (in any one of the three stages). Using the notation $A_{\sigma,\beta}(x)$ for the unique member of $\{A_{\alpha,\beta}\}_{\beta \in B_\varnothing}$ containing x and likewise for $A_{\sigma,\beta'}(x)$, consider $\overline{A} = \{x \mid A_{\sigma,\beta}(x) \neq A_{\sigma,\beta'}(x)\}$. As $\{x \mid \alpha \in B_\sigma, A_{\sigma,\alpha} \cap \overline{A} \neq \varnothing \Rightarrow \alpha < \beta\} = \{x \mid \alpha' \in B'_\sigma, A_{\sigma,\alpha'} \cap \overline{A} \neq \varnothing \Rightarrow \alpha' < \beta'\}$ by construction of $\overline{A}$, we know that $\mathrm{MC}(\overline{A}) = \mathrm{MC}(\cup\{A_{\sigma,\beta} \mid \beta \in B_\sigma, \beta < \overline{\beta}\}) = \mathrm{MC}(\cup\{A_{\sigma,\beta'} \mid \beta' \in B'_\sigma, \beta' < \overline{\beta'}\})$ by Proposition 5 part (f), where $\overline{\beta} = \mathrm{lub}\langle \beta \in B_\sigma, A_{\sigma,\alpha} \cap \overline{A} \neq \varnothing \Rightarrow \alpha < \beta \rangle$ and $\overline{\beta'}$ is defined similarly. But this means that if the latter two sets exist then they are the same, are members of $\{A_{\alpha,\beta}\}_{\beta \in B_\varnothing}$ and $\{A_{\alpha,\beta'}\}_{\beta' \in B'_\varnothing}$, respectively, and have nonempty intersection with $\overline{A}$, which contradicts the definition of $\overline{A}$. Hence they do not exist, that is, $\overline{A} = \varnothing$, therefore the decomposition is unique. There are many possible index sets, but they give rise to the same disjoint subsets and the same ordinal ranking over them, meaning $\sigma(x) <_{\mathrm{lex}} \sigma(y)$ for one choice of indices if and only if $\sigma(x) <_{\mathrm{lex}} \sigma(y)$ for any other, as desired.

(b) Let $I_o(\mathbf{T}) = \{x \in T \mid \nexists y \in T : \sigma(x) = \sigma(y)\}$ be the members of the catalog that are strictly comparable to all others, so $I(\mathbf{T}) = \overline{I_o(\mathbf{T})} \cup \{x^0\}$. It is obvious that $I_o(\mathbf{T})$ forms a chain under $<_c$, and that it will remain a chain as long as no two elements from the same σ-equivalence class (i.e., with identical sequences σ) are added, which is true for the closure. Thus it remains to show that if x is a limit point for $I_o(\mathbf{T})$ and $\sigma(x) = \sigma(x^0)$ then $x = x^0$. If $x = \lim_{n\to\infty} x_n$ and $\sigma(x) = \sigma(x^0)$, then $x^0 = \lim_{n\to\infty} x_n^0$ with $\sigma(x_n) = \sigma(x_n^0)$ for arbitrarily large n (else x and x^0 could be differentiated), using the continuity condition (d) for tournaments. Recall that $x_n \in I_0(\mathbf{T})$ $\forall n$ and therefore by definition it must be that $x_n = x_n^0$ for large n. Hence, since they are the respective limits, $x = x^0$ as desired. □

Proof of Proposition 17. Since $x^0 \in I(\mathbf{T}(G))$, the latter is nonempty. It is by construction closed and by part (b) of Proposition 11 totally ordered, so it has a unique maximal element. This proves that $\mathrm{RP}(G)$ is well defined. Note that the empty set satisfies A.1–A.3 and is therefore a WRP set for all games and an element of $\mathbf{T}(G)$. However, by part (b) of the definition for the relation in this tournament, $xR\varnothing$ $\forall x \in W(G), x \neq \varnothing$ and so the empty set is a Condorcet loser. This implies that $\mathrm{RP}(G)$ is always nonempty (in particular, although $\varnothing \in I(\mathbf{T}(G))$, $x^0 >_c \varnothing$). □

Proof of Proposition 18. We prove by induction on m that $u(P^m)$ satisfies A.1–A.5. If $m = 1$, then A.1 and A.3 follow directly from the definition of P^1 (the efficient frontier admits no strict Pareto comparisons), and A.2 is vacuous in this case as there are no nontrivial histories. Now, if $x \in W(G^1)$ and $x \neq u(P^1)$, then either $u(P^1) \rhd x$ and $x \not\rhd u(P^1)$, or $x \subset u(P^1)$. Both possibilities imply that $[u(P^1)]Rx$ in $\mathbf{T}(G)$, and hence that $u(P^1)$ is a Condorcet winner in the tournament, beating all other elements. In particular, $\mathrm{MC}(\mathbf{T}(G^1)) = \{u(P^1)\}$, a singleton, and so $\sigma(u(P^1)) = (1, 1, 1, \ldots)$ and no other set has that property. It is now clear that $u(P^1)$ satisfies A.4-A.5 as well.

Given that it holds up to $m - 1$, consider $u(P^m)$. Once again, A.1 and A.3 are trivial, while A.2 follows from the inductive hypothesis. Since, by A.2, all $x \in W(G^m)$ must only use continuation payoffs in $u(P^{m-1})$, $u(P^m)$ once again constitutes the Pareto frontier of all payoffs from any set in $W(G^m)$, and so the argument for A.4 and A.5 goes through exactly as above. □

The reason for this simplicity is that, in the finite case, continuation payoffs are only defined in strict subgames and so the set of feasible continuation payoffs is known when deciding which equilibria of the current game are renegotiation-proof. In fact, due precisely to this lack of circularity, all minimal WRP sets (with respect to size) are singletons and so can be very easily compared and joined if necessary.

Proof of Proposition 20. Let $S(G) = \cup\{x \in W(G) \mid x \text{ is SRP}\}$. Then $S(G)$ is itself SRP; in fact it is the (maximal) strongly renegotiation-proof set in G. It always exists, though it is not always nonempty. We show $S(G) \subseteq \mathrm{RP}(G)$. It is clear by part (c) of the definition for the relation in $\mathbf{T}(G)$ that $S(G)$ must be comparable with all other elements of $W(G)$, and in fact that $S(G) \in I(\mathbf{T}(G))$. Therefore either $S(G) = \mathrm{RP}(G)$, in which case we are done, or $S(G) <_c \mathrm{RP}(G)$. But since by the nature of $S(G)$ the latter relationship cannot occur via either parts (a) or (c) of the definition of R, it must be that $S(G) \subset \mathrm{RP}(G)$. □

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this article.

Acknowledgments

The author wishes to thank Abhijit Banerjee, Jeff Banks, Peter Diamond, Glenn Ellison, Paolo Ghirardato, Jim Mirrlees, David Pearce, Debraj Ray, Lones Smith, Jean Tirole, and seminar audiences at Caltech, MIT, Texas, UBC, UCLA, USC, UCSB, and Northwestern. All remaining mistakes are errors. This paper is the result of the author's independent research and does not necessarily represent the views of the Consumer Financial Protection Bureau or the United States.

Endnotes

1. See, for example, Benoit and Krishna [13] and Fudenberg and Maskin [14].
2. But see the discussion in Section 2 concerning communication.
3. It will be clear that we cannot choose more than one set. This is essentially because we have either that one of the sets Pareto dominates another, violating axiom 3, or the union of them is also WRP and therefore already under consideration on its own merits. See also the discussion of stationarity in Section 2.
4. Note that there is no hope of the infinite case being the limit of the finite case. This can be clearly seen in the example of the Prisoner's Dilemma, where (as with subgame perfection) the possibility of coordination is discontinuously restored at infinity.
5. Blume [5] also uses this term, though with the nominative form renegotiation-perfectness. The two concepts are not related, but there is an unfortunate scarcity of applicable names.
6. For instance, Jamison [15] offers a cheap-talk model of equilibrium selection by rational players.
7. Of course, the actions may still be history-dependent.
8. The current equilibrium can be considered as a status quo and given some preference in the model of this paper as a default choice (see Section 4), if desired. In some contexts this will be appropriate. However, it is not built in to the concept and the model as in other work, which often imposes less stringent conditions on the current equilibrium than it does on proposed deviations.
9. In a slight abuse of terminology, this more general concept is also called weakly renegotiation-proof.
10. In this scenario, we assume an odd number of voters in order to avoid ties. Such an assumption is not required for generalized tournaments, as defined below.
11. The appendix gives a proof that this is indeed well-defined.
12. So there is a natural equivalence relation, and the set of equivalency classes forms a chain.
13. With the single possible exception of the distinguished element.
14. As in Section 4, there will be a unique maximal element in the catalog index.
15. Of course, we could have instead used correlated equilibrium as our foundational concept. The results are qualitatively similar so, both in order to better focus solely on the issues raised by renegotiation and to facilitate comparison with the existing literature, we maintain the Nash paradigm.
16. This point was well recognized in even the earliest papers in the literature.
17. It is primarily this axiom that is not imposed in the work of Abreu et al. [6, 9].
18. Bernheim and Ray [2] formulate some definitions in a similar vein.
19. Note that if we have two WRP sets, neither of which confounds the other, then their union is also WRP.
20. This will certainly be a WRP set since myopic optimality requires no punishment threats, and hence there are no comparisons between continuation payoffs.
21. The observant reader will remark that this is somewhat symmetric to the covering relation of Section 3.
22. In a repeated stage game, myopic optimality translates into Nash equilibria of the one-shot game.
23. This line of reasoning was suggested by Lones Smith.
24. In fact it is clear how to extend their definition to finite stage games, and the analysis carries over.
25. All that is really necessary is that there is no chance to renegotiate in the course of playing G.
26. That is, if one accepts the stationarity assumption and hence WRP as minimally necessary.

References

[1] B. Dutta, "Covering sets and a new condorcet choice correspondence," *Journal of Economic Theory*, vol. 44, no. 1, pp. 63–80, 1988.

[2] B. D. Bernheim and D. Ray, "Collective dynamic consistency in repeated games," *Games and Economic Behavior*, vol. 1, no. 4, pp. 295–326, 1989.

[3] J. Farrell and E. Maskin, "Renegotiation in repeated games," *Games and Economic Behavior*, vol. 1, no. 4, pp. 327–360, 1989.

[4] J. P. Benoit and V. Krishna, "Renegotiation in finitely repeated games," ,*Econometrica*, vol. 61, no. 2, pp. 303–323, 1993.

[5] A. Blume, "Intraplay communication in repeated games," *Games and Economic Behavior*, vol. 6, no. 2, pp. 181–211, 1994.

[6] D. Abreu, D. Pearce, and E. Stacchetti, "Renegotiation and symmetry in repeated games," *Journal of Economic Theory*, vol. 60, no. 2, pp. 217–240, 1993.

[7] B. D. Bernheim, B. Peleg, and M. D. Whinston, "Coalition-proof nash equilibria I: Concepts," *Journal of Economic Theory*, vol. 42, no. 1, pp. 1–12, 1987.

[8] Q. Wen, "On renegotiation-proof equilibria in finitely repeated games," *Games and Economic Behavior*, vol. 13, no. 2, pp. 286–300, 1996.

[9] D. Abreu and D. Pearce, "A perspective on renegotiation in repeated games," in *Game Equilibrium Models II*, R. Selten, Ed., Springer, New York, NY, USA, 1991.

[10] D. Pearce, "Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation," Cowles Foundation Discussion Paper 855, 1991.

[11] J. Bergin and W. B. MacLeod, "Efficiency and renegotiation in repeated games," *Journal of Economic Theory*, vol. 61, no. 1, pp. 42–73, 1993.

[12] J. F. Laslier, *Tournament Solutions and Majority Voting*, Springer, New York, NY, USA, 1997.

[13] J. P. Benoit and V. Krishna, "Finitely repeated games," *Econometrica*, vol. 53, no. 4, pp. 890–904, 1985.

[14] D. Fudenberg and E. Maskin, "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, vol. 54, no. 3, pp. 533–554, 1986.

[15] J. C. Jamison, *Valuable Cheap-Talk and Equilibrium Selection [M.S. thesis]*, 1998.