Does Collective Rationality Entail Efficiency?

Paul Weirich

University of Missouri

Columbia, Missouri 65211, USA

8-09

Abstract

Collective rationality in its ordinary sense is rationality's extension to groups. It does not entail efficiency by definition. Showing that it entails efficiency requires a normative argument. Game theorists treating cooperative games generally assume that collective rationality entails efficiency, but formulating the reasoning that leads individuals to efficiency, and verifying the rationality of its steps, presents challenging philosophical issues. This paper constructs a framework for addressing those issues and reaches some preliminary results about the prospects of rational agents achieving efficiency in coalitional games. It concludes that only under strong idealizations does collective rationality entail efficiency.

Welfare economics formulates idealizations, such as perfect competition, under which markets yield efficient allocations of goods. In a similar spirit, this essay formulates idealizations under which efficiency emerges in coalitional games. These are games in which players may form coalitions and act jointly. Many theorists propose the standard of efficiency for solutions to cooperative games, including coalitional games, and so suggest that collectively rational players achieve an efficient outcome.[1] However, this essay finds that in a coalitional game, the players' collective rationality in its ordinary sense does not ensure efficiency even if the game satisfies standard idealizations. The essay therefore strengthens standard idealizations to make collective rationality generate efficiency. To obtain efficiency, it adds the idealization that players rationally prepare for their game.

The first section explains why, in coalitional games, collectively rational players may not achieve efficiency. It describes collective rationality and efficiency, and shows that although collective rationality aims for efficiency, obstacles often prevent its attainment. The second section examines collective rationality's requirement that players achieve a solution in a coalitional game. A solution assigns strategies to players so that the strategies are jointly rational, and consequently form an equilibrium among players' incentives. So that equilibrium is attainable, this section takes equilibrium as strategic equilibrium, a type of equilibrium that does not require all players to pursue all incentives—that is impossible in some coalitional games. It requires only that players pursue compelling incentives. Players who meet this equilibrium standard may not achieve efficiency, however. So collective rationality's requiring a solution does not in general ensure efficiency. The final section shows that a coalitional game's solution yields an efficient outcome if the players rationally prepare for their game. Prior to their game, fully rational players coordinate their pursuit of incentives so that it ensures

[1] See, for example, John von Neumann and Oskar Morgenstern (1944: Sec. 4) and John Nash (1950).

efficiency in the game. Because their rational preparation is a suitable idealization, collective rationality entails efficiency in thoroughly ideal conditions.

The essay's argument that collective rationality entails efficiency in an ideal coalitional game shows first that in an ideal game collective rationality requires each individual's rationality, and shows next that the rationality of all individuals entails efficiency. It claims that fully rational individuals prepare for a coalitional game by coordinating their pursuit of incentives so that it yields an efficient outcome in the game.

1. Collectively Rational Inefficiency

This essay gives rationality its ordinary sense and so takes rationality to require more than instrumental rationality, or rationality in pursuit of goals. Rationality evaluates beliefs and goals, as well as means of pursuing goals given beliefs. However, the essay relies mainly on principles of rationality governing pursuit of goals, such as the principle to maximize expected utility, which addresses cases where probability and utility functions represent, respectively, beliefs and desires. For an ideal agent in a standard decision problem, an option is rational just in case it maximizes expected utility. Although this is not a general definition of rationality, it covers the decision problems this essay treats. The principle to maximize expected utility, formulated in a general way, governs decisions that players make in games, as Weirich (1998) argues.

This section characterizes efficiency and collective rationality, examines obstacles to efficiency, and reviews idealizations that put aside some obstacles. Collective rationality aims at efficiency, and under suitable idealizations achieves efficiency, although standard idealizations do not ensure its success. Although efficiency may seem to be a requirement of collective irrationality, collectively rational groups may for various reasons fail to be efficient. Efficiency is a goal of collective rationality in the sense that in ideal conditions a group acting rationally achieves efficiency.

By efficiency, I mean Pareto optimality. For a group, one option is strictly Pareto-superior to another option if and only if all the group's members prefer the first option to the second option. A group's act is (weakly) efficient if and only if no alternative is strictly Pareto-superior to it. In a game, the options for the group of players are the possible combinations of their strategies. This essay does not distinguish efficiency and weak efficiency because it treats only games in which weakly efficient outcomes are efficient.

Collective rationality, in the ordinary sense the essay adopts, is rationality extended from individuals to groups. Not being technically defined as efficiency, if collective rationality entails efficiency, the entailment holds because of normative principles. A basic principle of collective rationality is its entailment by individual rationality: if the members of a group act rationally, then the group acts rationally. Weirich (2009) argues for this principle, but this essay takes the principle for granted. Using the principle to show that collective rationality entails efficiency requires showing that rational individuals achieve efficiency.

A group’s rationality is not equivalent to the rationality of all the group’s members because a group can act rationally even if some members act irrationally. For example, an electorate can rationally elect the best candidate although some voters irrationally choose another candidate. This essay does not advance general necessary and sufficient conditions of collective rationality. It characterizes collective rationality only for games in which outcomes are individuated by combinations of individuals’ strategies. In these games, collective rationality amounts to the rationality of all individuals. In games where each individual’s contribution to a solution is essential, collective rationality entails each individual’s rationality.

Because collective rationality differs from efficiency, individual rationality and collective rationality do not conflict in the Prisoner's Dilemma. In the Dilemma each player's defection is rational. So, if each defects, the pair acts rationally. Applying the basic principle of collective rationality, the pair's failure to cooperate is collectively rational. The players' inability to communicate and to reach binding agreements excuses the pair's inefficiency. Efficiency is a goal of collective rationality, but collectively rational players may not achieve that goal in adverse circumstances.

Providing opportunities for communication and binding agreements improves the prospects that rational players will achieve efficiency. Changing the circumstances of the Prisoner's Dilemma by adding these opportunities creates a cooperative version of the game. Suppose that in it, the players know their situation and know they are rational. They realize that each is better off if both cooperate than if both defect. Neither thinks that the other will decline a binding contract for mutual cooperation. Given ideal conditions for joint action, one player proposes the contract and the other accepts it. Thus the pair cooperates.

Suppose that to improve the prospects for efficiency in a cooperative game, we adopt the assumption that conditions are ideal for joint action. Besides opportunities for costless communication and binding contracts, agents are informed about their situation and each other. They know enough about their game and each other to have foreknowledge of their joint action. Strengthening the idealizations for cooperative games this way nonetheless fails to ensure efficiency.

Suppose that agents face a classical bargaining problem. Any bargain requires the agreement of all. The agents have opportunities for costless communication and binding contracts, and in ideal conditions for joint action have foreknowledge of joint strategies. The ideal conditions ensure that no individual has an incentive to block efficiency because he does better in an expected inefficient outcome. When two agents bargain over division of a good, none declines an efficient 50–50 split because he thinks that he will be on the good end of an inefficient 60–30 split (which they might receive if they enlist an arbitrator who charges a ten-percent fee). Rational bargainers form the grand coalition and efficiently divide its profits. Bargaining in ideal conditions produces efficiency.

In contrast with bargaining problems, players in a coalitional game may form productive coalitions smaller than the grand coalition. The ideal conditions that ensure efficiency in bargaining problems do not ensure efficiency in coalitional games. In a coalitional game, a player's bargaining leverage may vary according to the coalition to which she belongs. A player with much power in a small coalition may have little power in the grand coalition. A rational player joins a coalition that favors her even if that blocks formation of the grand coalition and efficiency. In ideal conditions players may be able to solve efficiently an isolated bargaining problem without being able to solve efficiently the multiple bargaining problems that a coalitional game creates.

Suppose that in a particular coalitional game, three people *A*, *B*, and *C* may form work teams. Ignoring unproductive 'teams' with just one member, the values of the teams are: $v(AB) = v(BC) = v(AC) = 8$ and $v(ABC) = 12$. A core allocation is a division of gains that gives each team at least its value. In this example, using alphabetical order to list payoffs for players, the outcome (4, 4, 4) is a core allocation. It accords every coalition at least its value.

In another three-person coalitional game, imagine that the values of productive work teams are: $v(AB) = v(BC) = v(AC) = 8$ and $v(ABC) = 9$. Any division of 9 units leaves some pair with an amount less than it can obtain on its own. So the game's core is empty. Suppose that *A* and *B* foresee 3 units apiece if the team {*A, B, C*} forms. They may decide to form the team {*A*, *B*} to generate 8 units and allot each member 4 units. *C* is excluded, and the outcome *A* and *B* produce, namely, (4, 4, 0), is inefficient. Inefficiency results despite ideal conditions for joint action in this game of exclusion.

To simplify, I consider the case for efficiency in elementary coalitional games. An elementary coalitional game, as I define it, has three properties. First is strict superadditivity: adding a member to a group increases the group's value. For example, a three-person coalition has more value than any two-person coalition. Second is transferable utility: a coalition's value is divisible among its members. Third is value-independence: outsider's acts do not affect a coalition's value. For example, if a two-person coalition forms, its value is the same whatever any third party does. Because I treat only elementary coalitional games, for brevity I often call these games coalitional games *tout court*.

In an elementary coalitional game, efficiency requires the grand coalition's formation. However, other coalitions may have incentives to fracture the grand coalition. A rational coalition may block the grand coalition's formation and so prevent efficiency. This may happen even in ideal conditions for joint action because the coalitional game may have an empty core. If the core is empty, then no matter how gains are divided, some coalition receives less than it can obtain on its own. Its members have an incentive to form that coalition rather than accept the division of gains. If a coalition smaller than the grand coalition pursues its incentive to form, the grand coalition fails to form and the outcome is not efficient. Rational players may realize an inefficient outcome.

Can results in economics show how to dodge obstacles to efficiency and establish that collective rationality entails efficiency? Economics contains celebrated results about efficiency's realization by rational bargainers. Coase's Theorem asserts that rational bargainers with symmetric information achieve efficiency in the absence of transaction costs. Coase did not prove this proposition, and others have proved it in special cases only. Although Coase's Theorem establishes rational bargaining's efficiency in some cases, it does not establish rational bargaining's efficiency in coalitional games with empty cores.[2]

---

[2] Ronald Coase (1960) presents the proposition known as Coase's Theorem, and George Stigler (1966: 110–114) argues for it. Drew Fudenberg and Jean Tirole (1991: 245) state the proposition this way: 'In the absence of transaction costs and with symmetric information, bargaining among parties concerned by a decision leads to an efficient decision, i.e., to the realization of gains from trade.' William Samuelson (1985) explains that when individuals' information is asymmetric or incomplete, rationality need not lead to efficiency in bargaining problems. A. Halpin (2007) points out the need for restrictions on a person's options in addition to restrictions concerning information and transaction costs. Furthermore, coalitional games with empty cores present obstacles to efficiency, as argue V. Aivazian and J. Callen (1981) and V. Aivazian, J. Callen, and I. Lipnowski (1987). Coase (1981) shows how to overcome those obstacles by

The First Fundamental Theorem of Welfare Economics states that efficiency emerges in a perfectly competitive market. It assumes a market with a nonempty core. It does not establish that rationality yields efficiency in coalitional games with empty cores. New methods are necessary for showing that rational agents achieve efficiency in those adverse circumstances.[3]

John von Neumann and Oskar Morgenstern (1944: Sec. 4) hold that a solution to a coalitional game is a set of imputations. An imputation is an allocation that is efficient and gives each individual at least as much as she can obtain on her own. Those properties are not enough to characterize a solution to a coalitional game, as they ignore some coalitions' incentives. A set of imputations with a certain stability property is a solution, von Neumann and Morgenstern claim. Their characterization of a solution is inadequate, however, because according to it some coalitional games lack solutions, whereas collective rationality's attainability ensures that every coalitional game has a solution. Also, von Neumann and Morgenstern assume without argument that the allocations that constitute solutions are efficient. Although efficiency is a plausible goal of collective rationality and therefore a plausible requirement of collective rationality in ideal coalitional games, they do not show that rational individuals realize an efficient allocation in an ideal coalitional game with an empty core.

Because of difficulty identifying idealizations that ensure efficiency, one may doubt that efficiency is a goal of collective rationality, that is, a property of an outcome that collectively rational agents achieve in ideal conditions. Perhaps even in conditions perfect for attaining efficiency, collectively rational individuals may fail to achieve it. If

---

restricting the bargaining protocol. P. Gangopadhyay (2000) shows how to overcome them by restricting coalitional games. Later, without sacrificing generality, Section 3, using strengthened idealizations, establishes that efficiency emerges from an ideal elementary coalitional game even if the game's core is empty.

[3] According to the First Fundamental Theorem of Welfare Economics, as R. Pindyck and D. Rubinfeld (1989: 570) explain, if each person trades in the marketplace to maximize her satisfaction, and all mutually beneficial trades are completed, the resulting allocation is efficient in the sense that no alternative allocation yields gains for some without losses for others.

that were so, then efficiency would not be a goal of collective rationality, contrary to this essay's assumption.

Some theorists, such John Broome (1991) and Philip Pettit (1993), raise objections to taking efficiency as a goal of collective rationality. They suggest that collective rationality does not require it even when conditions are perfect for attaining it. One objection to efficiency arises from collective rationality's responsiveness to symmetry. Consider bargaining over an indivisible good such as a painting. Imagine two people with equal claims to the painting. Because collective rationality responds to symmetry, it treats the claimants equally. However, equally dividing and so destroying the painting is inefficient, assuming that each claimant opposes its destruction and would rather that the other claimant have the painting entire than have the painting be destroyed.[4]

Using symmetry to oppose efficiency is unconvincing. In the example, the claimants may settle on an equal opportunity to possess the painting. They may flip a coin to allocate the painting, or they may arrange for each to have the painting in alternate years. Also, one claimant may have superior bargaining skill and may convince the other to relinquish his claim to the painting, perhaps in return for a side payment. The symmetry of their claims does not entail that the bargainers are symmetrical in all matters bearing on their bargaining. Finally, even if nondestructive symmetric settlements are impossible, and no asymmetry settles the bargaining problem, collective rationality uses symmetry only as a requirement for a bargaining problem's set of solutions. The set of solutions to the bargaining problem concerning the painting may have a solution according to which the first bargainer receives the painting and another solution according to which the second bargainer receives the painting. Then the set of solutions honors symmetry even if its elements do not. Symmetry receives its due without inefficiency, assuming that a symmetric set of solutions yields only efficient outcomes.

---

[4] Pettit (1993: 293–295) raises a similar objection.

Another objection to efficiency rests on the possibility that the players in a cooperative game have uninformed preferences. Collective rationality then may not promote efficiency, which is defined in terms of their preferences. For example, suppose that two politicians consider joining forces to back a war. Politician *A* supports war only because he believes it will stimulate the economy. He knows that it will not promote democracy. Politician *B* supports war only because she believes that it will promote democracy. She knows that it will not stimulate the economy. Then for *A* and *B*, war is efficient because each prefers it to peace. According to the objection, efficiency is not a goal of collective rationality in this case. War's efficiency rests on *A*'s and *B*'s uninformed preferences. In ideal conditions, when *A* and *B* are informed, they do not support war given their rationality. Both withdraw support of war. So collective rationality does not promote war in ideal conditions.[5]

This objection shows that efficiency with respect to preferences in nonideal conditions is not a goal of collective rationality. However, collective rationality's goal is efficiency with respect to preferences in ideal conditions. In ideal coalitional games, players have rational preferences. Also, players, and so their preferences, are equally informed about relevant facts. It is as if the players' information had been pooled prior to their game. Their preferences do not rest on incompatible beliefs. Inefficiency in ideal games counts against efficiency's being a goal of collective rationality. However, if the games raising doubts are not ideal in all respects bearing on collective rationality, then their case against the goal of efficiency dissolves. Idealizations about information enlist collective rationality's promotion of efficiency.

The objections discussed notwithstanding, efficiency is a goal of collective rationality. Collective rationality requires attaining that goal when conditions are ideal for its attainment. Although collective rationality does not entail efficiency in all cases, it

---

[5] Broome (1991: Chap. 7) advances an objection of this type.

entails efficiency under appropriate idealizations. This essay specifies idealizations that make collective rationality entail efficiency in coalitional games. Of course, not any conditions entailing efficiency count as idealizations. Conditions count as idealizations just in case they control for factors that explain behavior's collective rationality. The essay's idealizations satisfy this constraint.[6]

2. Strategic Equilibrium and Solutions

Whether collective rationality entails efficiency depends on collective rationality's demands. If collective rationality requires a core allocation, then it requires efficiency because a core allocation is efficient. However, collective rationality, being attainable, does not require a core allocation in coalitional games with empty cores. Collective rationality requires a less demanding, more general type of equilibrium than realization of a core allocation. This section introduces that type of equilibrium, strategic equilibrium, and shows that its realization does not entail efficiency. The appendix offers a technical summary.

In a coalitional game, a coalition may form or not form. If it forms, it may adopt various joint strategies involving participation of its members. Some joint strategies are productive and yield gains divisible among the coalition's members. A strategy profile for the coalitions in a game generates an assignment of individuals to coalitions that form and divisions of those coalitions' gains. Given a strategy profile, some coalition may have an incentive to change its strategy. Given realization of a core allocation, no coalition can on its own improve its allocation. However, because some coalitional games lack core allocations, collective rationality does not require realization of a core allocation. Rationality does not demand the impossible. In a coalitional game with an

[6] Weirich (2004: Chap. 3) supports and expands this account of idealizations.

empty core, not every coalition can pursue all incentives. Rationality does not require an individual or a coalition to pursue every incentive when doing that is endless. Not every incentive is a sufficient reason to act. Rationality requires only responding to sufficient reasons to act.[7]

To generalize the equilibrium requirement for coalitional games, I propose a type of equilibrium that does not require pursuit of every incentive but only every sufficient incentive. Speaking informally, an incentive is sufficient if and only if pursuing it generates an outcome. Whether pursuing an incentive generates an outcome depends on how others respond to pursuit of the incentive. Its pursuit may lead to their pursuit of other incentives so that pursuit of incentives is endless. Then its pursuit does not generate an outcome, and the incentive is not sufficient. If a profile of strategies is such that no agent has a sufficient incentive to unilaterally deviate, I call the profile a *strategic equilibrium*. Realization of a core allocation is a strategic equilibrium because no coalition gains from unilateral deviation. In a game with an empty core, where incentives are endless, a stopping place for pursuit of incentives is also a strategic equilibrium. The

[7] Unattainability is not the only common objection to requiring realization of a core allocation. Robert Aumann (1987: 470–471) and Howard Raiffa (2002: 444–446) object that core allocations are too demanding in some coalitional games with nonempty cores. Raiffa examines the game represented by this value function: $v(AB) = v(AC) = 10$, $v(BC) = 0$, and $v(ABC) = 10$. The only core allocation is (10, 0, 0). Agent *A* has tremendous bargaining power because no team produces anything without enlisting him. He plays *B* and *C* against each other. To win his collaboration, they lower their demands. He may induce them to lower their demands to zero. Raiffa claims that *B* and *C* should collude to end their bidding war. Together they may extract concessions from *A* because to gain he needs at least one's assistance.

The game Raiffa presents does not qualify as an elementary coalitional game because the coalition {*B, C*} is not more productive than is *B* alone or *C* alone. This essay addresses elementary coalitional games only. Also, Raiffa's objection does not target core allocations so much as the adequacy of a coalitional game's representation. If *B* and *C* may collude, then the value function stated does not adequately represent their resources. If they may enter an effective agreement not to engage in cutthroat competition to be *A*'s partner, then the value of their coalition is greater than 0. Granting that the value function adequately represents the coalitional game, so that *B* and *C* may not collude, the objection to the core allocation dissolves. If *B* and *C* cannot collude to restrain their competition, then they cannot stop a slide to the outcome (10, 0, 0). As long as a value function adequately represents agents' abilities to collaborate, core allocations survive the objection that they are excessively competitive.

following paragraphs elaborate this account of sufficient incentives and strategic equilibrium.[8]

An agent pursues an incentive to switch from one strategy to another if and only if, were the incentive to arise, she would take steps to realize the second strategy. An agent's pursuing an incentive is therefore a disposition to adopt a strategy in certain conditions. Agents' pursuit of incentives creates a dynamic structure. The dynamics for coalitions depends on the dynamics for individuals. For example, the coalition {*A*, *B*} has an incentive to move from (3, 3, 3) to (4, 4, 0) because *A* has an incentive to propose to *B* forming {*A*, *B*} and realizing (4, 4, 0), and *B* has an incentive to accept. The coalition pursues its incentives if its members pursue their incentives.

Paths of pursued incentives indicate bargaining's movement from one proposal to another. A hypothetical conditional expresses a step in a path. The conditional says that if one proposal is made, another proposal follows. In Section 1's game of exclusion, a conditional expressing a path's step may say that if the coalition {*A, B*} proposes (4, 4, 0), then the coalition {*B, C*} proposes (0, 6, 2). Such hypothetical conditionals describe the dynamics of bargaining. The dynamics they describe resemble the dynamics of a ball rolling in a basin. A strategic equilibrium corresponds to a low spot of the basin where the ball may settle.

The dynamics of the agents' pursuit of incentives settles the strategic equilibria of their game. For example, in Section 1's game of exclusion, an efficient outcome such as (3, 3, 3) may be the outcome of a strategic equilibrium. Whether it is depends on the

---

[8] Weirich (1998, 2009) presents a thorough account of strategic equilibrium for all games, both cooperative and noncooperative. In noncooperative games, strategic equilibrium is a generalization of Nash equilibrium. It exists more widely than does Nash equilibrium. For example, it exists in a game of Matching Pennies that lacks mixed strategies.

In a coalitional game, a strategy profile from which no coalition has an incentive to depart is a *strong Nash equilibrium* according to terminology that Robert Aumann (1959) introduces. Assuming that a coalition has an incentive to switch from one strategy profile to another if each member of the coalition has that incentive, only a strategy profile producing a core allocation is a strong Nash equilibrium.

incentives that rational agents pursue. If none pursues an incentive away from (3, 3, 3), then the strategies producing that outcome form a strategic equilibrium.

Whether a strategy profile is a strategic equilibrium depends on the structure of incentives that agents pursue. Their incentive structure consists of paths of pursued incentives. A graph with directed edges between strategy profiles may represent it. A strategic equilibrium is a profile at the end of a path, because an incentive is sufficient just in case it starts a path of pursued incentives. All paths of pursued incentives end because agents finish their game and do not pursue incentives endlessly. A strategic equilibrium exists in every coalitional game because a profile fails to be a strategic equilibrium only if it starts a path of pursued incentives. The end of that path is a strategic equilibrium. A profile's failure to be a strategic equilibrium requires the existence of a strategic equilibrium.

Strategic equilibrium generalizes the type of equilibrium that realization of a core allocation provides. Every core allocation is the outcome of a strategic equilibrium, but not every strategic equilibrium yields a core allocation. Although (3, 3, 3) in Section 1's game of exclusion may be the outcome of a strategic equilibrium, it is not a core allocation.

The literature offers other generalizations of equilibrium in coalitional games, but strategic equilibrium has superior decision-theoretic foundations. As Weirich (2009) shows, it fits into a unified theory of rationality that uses general principles, such as responding to a sufficient reason, to derive special principles for individuals and for groups. The unified theory contains, for example, a generalization of the principle of expected-utility maximization, namely, the principle of self-support, which requires that an option's adoption be self-supporting in light of information that the option's adoption provides about the option's outcome. It also contains an equilibrium principle for cooperative and for noncooperative games that requires realization of a strategic equilibrium. If a profile of strategies gives no agent a sufficient incentive to switch strategy, then the agents' strategies are jointly self-supporting. So the principle requires a profile of strategies that are jointly self-supporting.[9]

Collectively rational agents realize a solution to their game. Being an equilibrium is necessary for being a solution to a game, but a solution may demand more than does being an equilibrium. To show that collective rationality entails efficiency, one may try to show that a solution to a game yields efficiency. Features of a solution, besides being a strategic equilibrium, may ensure efficiency. For example, to be a solution, a strategy profile must be the culmination of rational pursuit of incentives. Does this constraint ensure that a solution yields efficiency? The remainder of this section, after characterizing a solution to a game, shows that it does not.[10]

A solution to a coalitional game is a profile of strategies, one for each agent, such that the strategies are jointly rational. A strategy profile has jointly rational strategies just in case each agent's strategy is rational given the strategy profile. If agents have foreknowledge of the strategy profile realized, as in an ideal coalitional game, the strategies in a profile are jointly rational just in case each agent's strategy is rational given knowledge of the profile's realization.

To be rational given a profile, a strategy must be self-supporting given the profile. So a profile of jointly rational strategies forms a strategic equilibrium. Because a coalitional game's incentive structure affects its equilibria, a representation of a coalitional game adequate for identifying its solutions includes the game's incentive structure along with

[9] Other generalizations of the set of core allocations are the bargaining set, the kernel, and the nucleolus. James Friedman (1990: 243, 256, 273, 275) and Joseph Greenberg (1990: 77) note some problems with these proposals. The proposals' analyses of objections to a strategy profile assume that a game's coalition structure is static and that pursuit of incentives ends after two steps.

[10] Ken Binmore (1987) shows that in some cases Nash's bargaining solution, and so efficiency, follows from the rationality of individuals. He demonstrates that in a two-person bargaining problem, if the bargainers take turns making offers, and there is a constant chance that bargaining breaks down after each offer, then the bargain struck converges to Nash's solution as the probability of a breakdown approaches 0. Binmore's demonstration rests on Ariel Rubinstein's (1982) proof that every bargaining problem the demonstration treats has a unique subgame perfect equilibrium. Binmore (1985) investigates extending Rubinstein's methods to some three-person bargaining problems but not to problems with an arbitrary number of people. In the last section I argue for efficiency in every elementary coalition game, suitably idealized, no matter how many people the game involves.

its value function. An adequate representation covers all the features that settle the game's solutions. A coalitional game's value function is an abstract representation of a game. A concrete realization of a coalitional game includes the agents’ psychological types. I treat concrete coalitional games, and not just their abstract representations. An important component of a concrete coalitional game is the agents' pursuit of incentives.

Because a solution to a coalitional game originates in the agents’ rationality, a strategic equilibrium among coalitions' incentives is not a solution to a coalitional game unless the equilibrium results from rational pursuit of incentives. Agents cannot pursue all incentives in a coalitional game with an empty core, so they should judiciously choose incentives to pursue. A rule for pursuit of incentives concerns an agent's selection of an incentive to pursue when the agent has multiple incentives. The rule of optimization requires pursuit of an optimal incentive. Another rule for pursuit of incentives requires stopping pursuit of incentives only if continuing pursuit of incentives is endless. An exhaustive set of such rules for rational pursuit of incentives does not settle agents' pursuit of incentives; it leaves agents a degree of latitude. The agents' psychological types, as well as their rationality, settle the dynamics of their pursuit of incentives.

Because the agents in an ideal coalitional game are rational, their pattern of pursuit of incentives complies with all applicable rules of rationality. Do the rules, despite not settling the dynamics of pursuit of incentives, ensure that solutions are efficient, or at least that some solution is efficient? If no solution is efficient, then the standards of efficiency and joint rationality conflict. In that case, joint rationality blocks rather than generates efficiency.

Every coalitional game has both efficient profiles and solutions (taken as profiles of jointly rational strategies), so the standards of efficiency and of joint rationality conflict if the set of efficient profiles and the set of solutions do not overlap. Some realizations of Section 1's game of exclusion prevent an overlap and create a conflict. In any realization every efficient profile requires formation of the three-person coalition and its division of

9 units of utility. The solutions depend on the agents' pursuit of incentives. The agents may rationally pursue incentives so that each of the three-person coalition's joint strategies starts a path of pursued incentives that terminates in a two-person coalition's joint strategy. Then no efficient profile is a solution, that is, a profile of strategies that are jointly rational for coalitions. In any concrete game in which paths of pursued incentives lead away from every efficient profile, no efficient profile is an equilibrium and so none is a solution.

To illustrate, consider a case in which *B* prevents formation of the grand coalition. She does better in {*A*, *B*} than in {*A*, *B*, *C*} because she bargains better with *A* alone than with both *A* and *C*. She foresees that if {*A*, *B*} forms, the outcome is (2, 6, 0), whereas if {*A*, *B*, *C*} forms to pursue the incentive to achieve (2.4, 6.4, 0.2), or any other efficient allocation, the final outcome is (3, 3, 3). Anticipating more from the two-person coalition than from the three-person coalition, she blocks the coalition {*A*, *B*}'s expanding to include *C*.

Because the coalition {*A*, *B*, *C*} does not pursue its incentive to move from (2, 6, 0) to (2.4, 6.4, 0.2), the allocation (3, 3, 3) is not the final outcome. In fact, the coalition {*A*, *B*} pursues its incentive to switch from realization of (3, 3, 3) to realization of (4, 4, 0). The coalition {*A*, *C*} responds with (6, 0, 2). Then the coalition {*B*, *C*} switches to (0, 4, 4). Finally, the coalition {*A*, *B*} responds with (2, 6, 0). Coalitions stop pursuit of incentives at (2, 6, 0). Thus, coalitions pursue incentives along the path (3, 3, 3) → (4, 4, 0) → (6, 0, 2) → (0, 4, 4) → (2, 6, 0) and stop at the end. Their pursuit of incentives complies with rules of rationality. Realizing (2, 6, 0) is jointly rational but is not efficient.

The standard of joint rationality for coalitions survives conflict with the principle of efficiency. The principle of efficiency fails in noncooperative games, such as the Prisoner's Dilemma, because it ignores individuals' incentives. In coalitional games, the principle of efficiency ignores incentives of coalitions smaller than the grand coalition. It does not acknowledge a subcoalition's incentive to deviate from an efficient profile. Efficiency is too narrow-minded to be a general standard for coalitional games.

Suppose that rules of rational pursuit of incentives give priority to the grand coalition's incentives and require pursuit of optimal incentives. Suppose also that they make the grand coalition the last to halt pursuit of incentives and allow it to halt only at a joint strategy. Then the rules ensure that paths of pursued incentives contain efficient profiles and stop only at efficient profiles. However, the rules for rational pursuit of incentives have no reason to favor the grand coalition's incentives to adopt joint strategies. The rules rest on individuals' reasons for pursuit of incentives so that the rules may support collective standards for solutions. Efficiency is a collective reason for pursuit of incentives and so not good grounding for the rules. Plausible rules for pursuit of incentives do not eliminate the possibility of inefficient profiles of jointly rational strategies.

In the example, the coalition $\{A, B, C\}$ has an incentive to switch from a profile realizing (2, 6, 0) to a profile realizing (2.4, 6.4, 0.2). Achieving the second outcome is better for each member of the coalition. The incentive to switch is insufficient, however. It starts a path of incentives leading back to (2, 6, 0). Consequently, $\{A, B, C\}$ may justifiably fail to pursue its incentive to achieve (2.4, 6.4, 0.2). Pursuit starts a cycle of incentives, and a coalition may rationally forgo nonproductive relentless pursuit of incentives. Rationality allows the grand coalition to halt pursuit of its incentives. It may rationally abandon pursuit of an incentive to adopt a joint strategy. Not only proper subcoalitions may rationally forgo incentives.

Rules of rationality for pursuit of incentives do not ensure the efficiency of a coalitional game's solution. It is possible that a game's strategic equilibria and solutions are all inefficient, and *a fortiori*, that not all its solutions are efficient. Idealizations stronger than the players' rationality in the game are necessary to guarantee that solutions are efficient.

3. Full Rationality

To complete the essay’s project, this final section presents an assumption that ensures efficiency and then establishes that the assumption is an appropriate idealization. The idealization regulates agents’ preparations for a coalitional game. The argument from the idealization to efficiency rests on straightforward utility maximization. Adding the idealization to standard idealizations makes collective rationality entail efficiency in ideal coalitional games.

Section 1 briefly describes rationality. It notes that rationality requires more than maximization of expected utility. It imposes constraints on probability and utility assignments, for instance. An individual's full rationality in a standard decision problem requires not only maximizing expected utility, but also having rational probability and utility assignments to ground options' expected utilities. To be fully rational in a decision problem, and not just rational granting certain circumstances, an individual must be rational in all matters bearing on her decision. A decision that maximizes expected utility with respect to irrational probability and utility assignments is rational given those assignments, but is not rational if the irrational assignments adversely affect the decision in a significant way.

In a coalitional game, a fully rational player passes a comprehensive evaluation and not just a limited evaluation taking some circumstances for granted. Full rationality requires rationality in all matters relevant to the game and not just rationality in strategy selection, taking for granted the player’s circumstances. In addition to requiring players to pursue incentives rationally, full rationality requires them to have rationally coordinated their pursuit of incentives. The players control the dynamics of their pursuit of incentives, within limits that rationality imposes. When the game starts, the dynamics are settled, but prior to the game's start, the players have had opportunities to adjust the dynamics. Although they cannot coordinate pursuit of incentives during their coalitional game, they may coordinate prior to their game, so that when their game begins, they have a fully rational pattern of pursuit of incentives. Preparation for a game changes not the game's value function but its concrete realization.

Players coordinate their pursuit of incentives if each player settles her pursuit of incentives jointly with the other players. This coordination may be the result of a pre-game agreement. One player may agree not to pursue a certain incentive if another player agrees not to pursue another incentive. The players may agree to adopt a certain pattern for pursuit of incentives. Suppose that in Section 1's game of exclusion *A*, *B*, and *C* are initially inclined to stop pursuit of incentives at (2, 6, 0). They may agree to stop at (2.4, 6.4, 0.2) instead. Rational coordination ensures efficiency because if a pattern of pursuit of incentives yields an inefficient outcome, all players have a reason to adopt an alternative pattern that yields an efficient outcome.

Each player gains from all players' stopping pursuit of incentives at an efficient outcome rather than at a strictly inferior outcome. Consequently, fully rational players, who not only play their game rationally but also rationally prepare for their game, create a pattern for their pursuit of incentives that ensures the efficiency of the game's equilibrium outcomes. They pursue incentives in a way coordinated to achieve efficiency.

In ideal conditions for a coalitional game, agents have had an opportunity to coordinate pursuit of incentives so that it leads to an efficient outcome. They may do this without violating rationality's constraints on their pursuit of incentives. In an ideal version of the game of exclusion, fully rational agents have coordinated their pursuit of incentives so that it does not stop at (2, 6, 0) but instead continues to an efficient outcome such as (2.4, 6.4, 0.2). Their pursuit of incentives ensures that the coalitions forming do not exclude a productive agent. Because the agents have the opportunity to coordinate pursuit of incentives and because coordination brings gains for each, they coordinate to achieve efficiency.[11]

Preparation for a coalitional game is a coordination problem. Players realize a pattern for pursuit of incentives by each player's settling her pursuit of incentives. According to some but not all patterns, strategic equilibria are efficient. In the example, according to one pattern, the strategic equilibrium realized yields (2, 6, 0), but according to another pattern the strategic equilibrium realized yields (2.4, 6.4, 0.2). Players have a common interest in creating a pattern that generates efficiency.

The players' problem of coordinating pursuit of incentives duplicates some but not all features of the players' coalitional game. Before the game of exclusion starts, players *A* and *B* can exclude player *C* from coordination of pursuit of incentives and thereby exclude *C* from the game's benefits. However, in preparations for the game, *C* can eliminate every reason for *A* and *B* to exclude him from coordination of pursuit of incentives. He can become more accommodating. Preparation offers players additional options. Instead of leaving a game's dynamics to chance, they can eliminate psychological traits that cause exclusion.

In ideal coalitional games, players know how others pursue incentives. Coordination need not occur by agreement. It may occur through recognition of players' psychological types and opportunities to change type. However, players may coordinate by agreement given the opportunities for communication and binding contracts that coalitional games provide. If the players' current pattern of pursuit of incentives generates inefficiency, some player proposes an alternative generating efficiency, and others accept it.

The standard of efficiency for coalitional games assumes ideal conditions. Taking the conditions to include agents' full rationality, realization of a strategic equilibrium and a solution entails realization of an efficient outcome. Efficiency emerges in an ideal coalitional game, even if the core is empty.

---

[11] Weirich (2007) explains how coordination generates efficiency.

An idealization is an assumption that controls for an explanatory factor. The reasons that make players' rational selection of strategies a good idealization also make their full rationality a good idealization. Full rationality controls for factors in the explanation of an outcome's collective rationality, factors such as players' steps toward the game's outcome. The assumption of full rationality ensures that their steps, including their use of opportunities to coordinate pursuit of incentives, do not have a significant adverse effect on the game's outcome. Hence, the assumption is a suitable idealization for supporting efficiency. It removes an obstacle to this goal of collective rationality.

To deepen the argument for efficiency in an ideal coalitional game, one may use principles of rationality such as expected-utility maximization to show in more detail that agents coordinate to realize a pattern for pursuit of incentives that makes the game's equilibria efficient. To broaden the argument, one may add principles of rationality that select a particular efficient equilibrium when several exist, for example, the principle to select one that is subgame perfect with respect to the bargaining protocol. Such principles explain how an agent's bargaining power leads to an efficient equilibrium that favors her. Also, one may extend the argument to cooperative games besides the elementary coalitional games this essay treats. The essay's purpose is just to present the problem of demonstrating that rational agents achieve efficiency in coalitional games with empty cores and to sketch a method of resolving the problem in ideal coalitional games, assuming that the idealizations include the agents' full rationality.

This essay shows that if the individuals in an ideal coalitional game are fully rational, so that they have taken advantage of opportunities to coordinate their pursuit of incentives, then they realize a strategic equilibrium that generates an efficient outcome. If the individuals do not realize a strategic equilibrium, then some individual fails to pursue a sufficient incentive. That individual is not rational. If their strategic equilibrium does not generate an efficient outcome, then some individual has missed an opportunity to coordinate pursuit of incentives to realize a strategy profile she prefers.[12]

Appendix

This appendix formally presents coalitional games, their concrete realizations, paths of incentives, strategic equilibrium, and some basic points about strategic equilibrium. It treats ideal coalitional games in which agents rationally pursue incentives.

A coalitional game with *n* individuals has $2^n - 1$ coalitions. A value function *v* specifies the value of each coalition. A coalition may fail to form, or form and divide its value among its members. A strategy profile *s* specifies which coalitions form and their divisions of their values. $U(s)$ is the utility profile resulting from *s*. It is an *n*-tuple $(u_1, u_2, \ldots, u_i, \ldots, u_n)$ listing the utility each individual receives given *s*'s realization. $U_i(s) = u_i$. $U(s)$ is subject to the constraint that if according to *s* coalition *c* forms, then for $i \in c$, $\sum_i u_i \leq v(c)$.

A concrete coalitional game has, besides a value function, a directed graph with strategy profiles as vertices and between them directed edges for coalitions. Ordered pairs of vertices may represent the edges. Construction of the game's directed graph occurs in three stages. The first stage creates a directed edge for a coalition *c* between any two profiles *s* and *s´* such that for all $i \in c$, $U_i(s) < U_i(s')$ and such that *c* can realize *s´* by changing its strategy in *s*. The second stage removes edges for *c* away from a strategy profile *s* to a profile *s´* if and only if there is another edge for *c* away from *s* to a profile *s´´* such that for all $i \in c$, $U_i(s') < U_i(s'')$, and then removes all but one of the remaining edges for *c* away from *s*. A path of incentives is an ordered set of edges such that for every pair of adjacent edges, the second member of the first pair is the same as the first member of the second pair. The third stage, for each endless path of incentives,

[12] I thank the referees and editors for helpful comments and also an audience at the University of Tilburg, where I presented an early version of this essay.

selects an edge and removes the subsequent edge. The result is the directed graph for the concrete coalitional game.

To obtain from a path starting with an edge for a coalition *c* a path of incentives for *c*, delete from the path every edge between two strategy profiles according to which *c* does not form. In a concrete coalitional game, a coalition *c* has a sufficient incentive to switch from *s* to *s´* if and only if it has a path from *s* to *s´*. A strategic equilibrium of a concrete coalitional game is a strategy profile *s* such that no coalition has a sufficient incentive to switch to another profile, that is, has a path away from *s*.

The following theorem and propositions follow immediately from the foregoing definitions.

Theorem. Every concrete coalitional game has a strategic equilibrium.

Proposition 1. In some concrete coalitional game all strategic equilibria are inefficient.

Proposition 2. If in a concrete coalitional game all paths terminate in efficient strategy profiles, then the game's strategic equilibria are efficient.

References

Aivazian, V. and J. Callen. 1981. 'The Coase Theorem and the Empty Core.' *Journal of Law and Economics* 24: 175–181.

Aivazian, V., J. Callen, and I. Lipnowski. 1987. 'The Coase Theorem and Coalitional Stability.' *Economica* 54: 517–520.

Aumann, R. 1959. 'Acceptable Points in General Cooperative *n*-Person Games.' In R. D. Luce and A. W. Tucker, eds., *Contributions to the Theory of Games* IV, *Annals of Mathematical Studies* 40, pp. 287–324. Princeton: Princeton University Press.

Aumann, R. 1987. 'Game Theory.' In J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave: A Dictionary of Economics*, Vol. 2, pp. 460–482. London: MacMillan.

Binmore, K. 1985. 'Bargaining and Coalitions.' In A. Roth, ed., *Game-Theoretic Models of Bargaining*, pp. 269–304. Cambridge: Cambridge University Press.

Binmore, K. 1987. 'Nash Bargaining Theory II.' In K. Binmore and P. Dasgupta, eds., *The Economics of Bargaining*, pp. 61–76. Oxford: Blackwell.

Broome, J. 1991. *Weighing Goods: Equality, Uncertainty, and Time*. Oxford: Blackwell.

Coase, R. 1960. 'The Problem of Social Cost,' *Journal of Law and Economics* 3: 1–44.

Coase, R. 1981. 'The Coase Theorem and the Empty Core: A Comment.' *Journal of Law and Economics* 24: 183–187.

Friedman, J. 1990. *Game Theory with Applications to Economics*. Second edition. New York: Oxford University Press.

Fudenberg, D. and J. Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.

Gangopadhyay, P. 2000. 'On the Coase Theorem and Coalitional Stability: The Principle of Equal Relative Concession.' *Theory and Decision* 48: 179–191.

Greenberg, J. 1990. *The Theory of Social Situations: An Alternative Game-Theoretic*

Approach. Cambridge: Cambridge University Press.

Halpin, A. 2007. 'Disproving the Coase Theorem.' *Economics and Philosophy* 23: 321–341.

Nash, J. 1950. 'The Bargaining Problem,' *Econometrica* 18: 155–162.

Pettit, P. 1993. *The Common Mind: An Essay on Psychology, Society, and Politics*. New York: Oxford University Press.

Pindyck, R. and D. Rubinfeld. 1989. *Microeconomics.* New York: Macmillan.

Raiffa, H. 2002. *Negotiation Analysis: The Science and Art of Collaborative Decision Making*, written with J. Richardson and D. Metcalf. Cambridge, MA: Harvard University Press.

Rubinstein, A. 1982. 'Perfect Equilibrium in a Bargaining Model.' *Econometrica* 50: 97–109.

Samuelson, W. 1985. 'A Comment on the Coase Theorem.' In A. Roth, ed. *Game-Theoretic Models of Bargaining*, pp. 321–339. Cambridge: Cambridge University Press.

Stigler, G. 1966. *The Theory of Price*. Third edition. New York: Macmillan.

von Neumann, J. and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Weirich, P. 1998. *Equilibrium and Rationality: Game Theory Revised by Decision Rules.* Cambridge: Cambridge University Press.

Weirich, P. 2004. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. New York: Oxford University Press.

Weirich, P. 2007. "Initiating Coordination." *Philosophy of Science* 74: 790–801.

Weirich, P. 2009. *Collective Rationality: Equilibrium in Cooperative Games*. New York: Oxford University Press.