1

Games and Solutions

Game theory and decision theory have a symbiotic relationship. Game theory motivates revisions in principles of rational choice. And decision theory motivates revisions in accounts of solutions to games. My project is to revise game theory in light of reflections on rational choice. I provide a new account of equilibrium in games – equilibrium among strategic reasoners – motivated by equilibrium's connection with solutions and rational choice.

To explain equilibria and solutions, let us examine a game of Hide and Seek with a time limit. The Seeker has to find the Hider before time expires; otherwise the Hider wins. The Hider can conceal herself on either the first or the second floor of a house. The second floor is smaller than the first so that the Seeker can search all of it before time expires, whereas he can search only half the first floor in the time available. The second floor has windows from which, looking down through skylights, parts of the first floor can be seen. As a result, half the first floor can be searched while searching the second floor. The players know about the windows but do not know which parts of the first floor are visible from the second floor.

Shuttling between floors would be time wasting for the Seeker. His only effective strategies are (1) searching the first floor and (2) searching the second floor. The Hider also has to choose between the first and second floors. Suppose that the players know that the chances of the Hider's being found are as follows for each of their strategy combinations: (a) 100% if the Hider is on the second floor and the Seeker searches there, (b) 0% if the Hider is on the second floor and the Seeker searches the first floor, (c) 50% if the Hider is on the first floor and the Seeker searches there, and (d) 50% if the Hider is on the first floor and the Seeker searches the second floor. Figure 1.1 summarizes these probabilistic outcomes of their strategy combinations. What should the players do given that each is psychologically astute and able to anticipate the other's strategy?

Plainly, each player should decide in light of the other's choice. This rules out some strategy combinations. If the Hider goes to the second floor, then, being an accurate predictor, the Seeker looks there. The Hider has an incentive to switch floors; her strategy is not self-supporting. Only if the Hider goes to the first floor and the Seeker searches the second floor are their strategies jointly self-supporting. Such an outcome is an equilibrium since

		Hider	
		Floor 1	*Floor 2*
Seeker	*Floor 1*	50% discovery chance	0% discovery chance
	Floor 2	50% discovery chance	100% discovery chance

Figure 1.1 Hide and Seek.

		Mismatcher	
		Heads	*Tails*
Matcher	*Heads*	Matcher wins	Mismatcher wins
	Tails	Mismatcher wins	Matcher wins

Figure 1.2 Matching Pennies.

it balances the strategic considerations of the players. Joint self-support is a prerequisite of joint rationality, which is the hallmark of a solution. So the game's unique equilibrium – hiding on the first floor and seeking on the second floor – is also its unique solution. Here a solution is understood in a subjective sense that makes it depend on the rationality of strategies rather than their successfulness. The game has no solution in an objective sense according to which a solution gives each player success. Whatever the players do, one wins and one loses.

1.1. MISSING EQUILIBRIA

Now consider another game, Matching Pennies. This game has two players, each of whom has a penny he can display with either Heads or Tails showing. The players display their coins simultaneously. One wins if the coins match; the other wins if they do not match. Figure 1.2 depicts their situation. Suppose that the players cannot randomize their choices, and each will anticipate the other's move. What should they do? This game has no strategy combination in which each strategy is a best reply to the other. If both players display Heads, for example, the mismatcher does better with Tails. What are the consequences of such games for an account of solutions and equilibria?

The outcome of a game is a strategy for each player. A (subjective) solution to a game is an outcome in which the players are jointly rational. Does every game have a solution? I hold that it is always possible to choose rationally. I deny the existence of dilemmas of rationality in which all choices are irrational. I do not argue for this view but take it as a starting point. This view motivates the related but independent view that every ideal game has a solution. I argue that since the players in a game do not face dilemmas of rationality, their joint rationality is possible in ideal games; that is, a solution exists. The restriction to ideal games is essential since idealizations concerning the rationality and knowledge of the agents are needed to remove obstacles to the existence of a solution. The possibility of individual rationality for each agent is not sufficient by itself to ground the possibility of joint rationality for all agents. But, I contend, individual rationality (Chapter 4's topic) and appropriate idealizations (Chapter 2's topic) together ensure the existence of a solution.

An equilibrium of a game is an outcome in which the strategies adopted by the players are jointly self-supporting. Since a rational strategy is self-supporting, a solution of a game has strategies that are jointly self-supporting. In other words, a solution is an equilibrium. Since every ideal game has a solution, it also has an equilibrium. This conclusion forces a revision of views about equilibrium.

The common view takes equilibrium to be Nash equilibrium. This view interprets self-support in ideal games, in which players are aware of what others do, in terms of best replies. A *Nash equilibrium* of an ideal game is an outcome in which each player adopts a best reply to the others. There are two ways to interpret the standard of Nash equilibrium. The *objective* interpretation states the standard in terms of payoff increases: An outcome is a Nash equilibrium if and only if no strategy change by a single agent produces a payoff increase for the agent. The *subjective* interpretation states the standard in terms of preferences of the agents: An outcome is a Nash equilibrium if and only if given the outcome, no agent prefers an outcome that he can reach by changing strategy unilaterally. A subjective Nash equilibrium is an outcome that is *incentive-proof*, or free of (subjective) incentives to switch strategy. The objective interpretation is the canonical one. But the standard is usually advanced for ideal games, where agents are rational and informed about the game. The idealizations make payoff increases and incentives coincide so that the objective and subjective interpretations agree. In ideal games the objective Nash equilibria are exactly the subjective Nash equilibria. I generally adopt the subjective interpretation of the standard, since it makes more explicit the connection between Nash equilibrium and rationality requirements for agents. But I also generally limit myself to ideal games, in which the two interpretations of the standard agree. Consequently, the distinction between the two interpretations of the standard is not critical for my purposes.

Some ideal games, such as the version of Matching Pennies discussed, lack Nash equilibria. The subjective interpretation of the standard of Nash equilibrium insists on the absence of agents' incentives to switch strategy. But the standard cannot be met in all games. This creates a problem given the foregoing views about solutions and equilibria. Here is the problem in a nutshell. Every ideal game has a solution. Moreover, only equilibria are solutions since self-support is necessary for rationality. Consequently, every ideal game has an equilibrium. But, uncontroversially, there are games without Nash equilibria and, I argue, also ideal games without Nash equilibria. Therefore the common view of equilibrium must be revised to provide for the existence in ideal games of equilibria taken as combinations of strategies that are jointly self-supporting. We need an account of equilibrium weaker than Nash equilibrium that allows for the existence of equilibria in ideal games. The full argument for revising contemporary views about equilibrium takes us to the end of Chapter 3. Chapters 4–8 work out the revision.

The addition of randomized or *mixed* strategies guarantees the existence of a Nash equilibrium in a *finite* game (a game with a finite number of agents and of unrandomized or *pure* strategies for each agent). At least it does so under the usual assumption that the value of a mixed strategy is the probability-weighted average of the values of component pure strategies. Nash's famous theorem (1950) establishes this. The theorem stands even if an agent's mixed strategies are taken, as in Harsanyi (1973), as probability mixtures only with respect to other agents' probability assignments. Nonetheless, the problem of missing Nash equilibria remains for games without mixed strategies (because of the unavailability of randomization or the predictive powers of agents) and for games with an infinite number of pure strategies for some agent. A general theory of equilibrium must address such games.

Some theorists quickly dismiss the problem. According to the common view that a Nash equilibrium is a solution, a game with no Nash equilibrium is simply a game with no solution. However, this response runs contrary to the intuition that every ideal game has a solution (putting aside special cases with an infinite number of agents) and wreaks havoc with the theory of rationality, as we shall see. We need to explore alternative, less devastating responses to the problem. We need to reexamine rationality requirements for strategies, and equilibrium standards for solutions.

Equilibria as generally conceived are absent in ideal games because of the stringent type of self-support used to define equilibria. Commonly a strategy is said to be self-supporting if and only if it maximizes expected utility on the assumption that it is adopted, in other words, if and only if it is incentive-proof, or ratifiable. This standard of self-support is too high. It cannot be met in all decision problems. Chapter 4 provides examples. Our approach to the problem of missing equilibria is to lower the standard of self-support. We obtain a weaker, more easily attained equilibrium standard by using a less demanding type of self-support to fill out the definition of equilibrium.

I modify the definition of an equilibrium, a profile of jointly self-supporting strategies, by taking *self-support* as the absence of a sufficient reason to switch strategy, but not necessarily the absence of an incentive to switch strategy. As Chapter 4 argues, not every incentive to switch strategy is a sufficient reason to switch. An incentive may be insufficient, for instance, if it is undermined by the supposition that it is pursued, in particular, if given its pursuit the other agents' response undermines the incentive. Our new type of equilibrium, motivated by decision principles concerning self-support, is attainable in ideal games.

The new account of self-support considers paths of incentives to switch options, say an incentive to switch from A to B, from B to C, and so on. A path of pursued incentives terminates if it has a last member and no extension. To be sufficient in the games treated, an incentive must go somewhere; it must start a terminating path of pursued incentives. As Chapter 4 argues, in these games a self-supporting strategy is one that does not start a terminating path of pursued incentives. Chapter 5 applies this view to equilibria – they must be composed of strategies that are jointly self-supporting – and shows that all games of a certain type have an equilibrium. Chapter 6 shows that in some of those games a profile is an equilibrium if and only if no agent pursues an incentive away from it and uses this feature of equilibria to identify them. Chapter 7 explains that an ideal version of Matching Pennies has an equilibrium dependent on the players' pattern of pursuit of incentives. If, for instance, the mismatcher abandons endless pursuit of incentives at the point where both players show Heads, then that point is an equilibrium.

My main concern is the existence of equilibria. Showing that equilibria exist is of course different from showing that they are realized. Showing that equilibria exist in a game requires showing the possibility of joint self-support. It is another matter to show that the possibility is realized, and that the outcome of the game is an equilibrium. I put aside the realization of equilibria. I do not consider whether given appropriate idealizations rational agents realize an equilibrium. Nor do I consider whether in games with multiple equilibria, rational agents do their parts in a particular equilibrium. I offer no principles of equilibrium selection. The problem of equilibrium selection is challenging, and is exacerbated by our new equilibrium standard since it provides equilibria in addition to Nash equilibria from which to choose. Still, I consider only whether given appropriate idealizations equilibria exist. Rather than use decision principles to derive the realization of an equilibrium, I use them to establish the existence of an equilibrium. Instead of deriving the realization of an equilibrium from compliance with decision principles, I reduce being an equilibrium to compliance with certain decision principles. Then I show the possibility of attaining the equilibrium standard by demonstrating the possibility of compliance with the decision principles. Explaining the realization of equilibria is an important project, but a project for another occasion.

One may wonder whether the issues of existence and realization of equilibrium can be separated as I propose. Doesn't support for a definition of equilibrium require showing not only that equilibria exist in ideal games but also that an equilibrium is realized? If certain profiles are jointly self-supporting in ideal games, then one of them must be realized. The realization of other profiles is incompatible with the agents' rationality and knowledge. To verify the definition of equilibrium, shouldn't we explain how principles of individual rationality lead each agent to his part in one of the equilibria? This verification is indeed desirable. But it requires a much broader account of rationality than an account of self-support, and a much broader account of agents' knowledge in ideal games than we provide later. To make our project manageable, we have to forgo this type of verification. The case we make for the revised definition of equilibrium is nonetheless strong.

1.2. NORMAL-FORM GAMES

The equilibrium standard for solutions applies to all games since it stems from the general rationality requirement that an agent's choice be self-supporting. However, our investigation of equilibrium is limited to a certain class of games – normal-form, noncooperative, single-stage games of strategy with a finite number of agents. The terminology used to define this class of games is from standard works in game theory; it is reviewed to inform those new to the field and to make precise the interpretation adopted here. These games clearly bring out a problem with interpreting self-support as incentive-proofness and hence equilibrium as Nash equilibrium.

A *game* is a decision situation involving two or more agents where each agent's outcome depends on the actions of other agents as well as his own action. This definition covers familiar games such as chess, but also competition between firms, coordination between drivers in traffic, bargaining over a division of resources, and many other types of personal interaction. The definition may be extended to include degenerate cases where there is just one agent, or the outcome for an agent depends only on his own action. Then any decision situation for any number of agents counts as a game. But the central cases, and the ones that interest us, involve multiple agents with the outcome for each agent dependent on the actions of all.

We call the options available to an agent in a game *strategies*. The strategies may be simple actions or complex conditional actions replete with contingency plans and the like. We use the term *strategies* because the reasons for an agent's action in a game are often strategic in a broad sense,

Cambridge University Press
978-0-521-03802-7 - Equilibrium and Rationality: Game Theory Revised by Decision Rules
Paul Weirich
Excerpt
More information

looking ahead to the responses of other agents. We call an assignment of strategies to agents, one strategy for each agent, a strategy *profile*. A strategy profile determines the outcomes or *payoffs* of the game for the agents. Since a strategy profile determines the outcome, it may substitute for a possible outcome.

We focus on games in which rational decision relies on principles of strategic reasoning. We call these *games of strategy*. We put aside other games, for example, repeated games, in which rational decision requires learning about opponents' future choices from their past choices.

As in nondegenerate games of any sort, in a game of strategy the outcome of one agent's strategy is dependent on other agents' strategies. A situation in which two agents simultaneously choose from a dinner menu, for example, does not constitute a game of strategy. Also, and more important, in a game of strategy an agent's choice affects his preferences among strategies, or his incentives to adopt strategies. Given one choice an agent may have an incentive to adopt a certain strategy, but given another choice that incentive may disappear. An agent's incentives to adopt strategies are not constant given his possible choices, in contrast with most decision situations. In games of strategy rational agents use strategic reasoning to take account of the anticipated effect of a choice on incentives to adopt strategies. To explain these games, let us consider the characteristics of strategic reasoning.

In many games of strategy incentives are inconstant with respect to strategies because the agents can anticipate, at least to some degree, the responses of other agents to their strategies. An agent's adoption of a strategy provides evidence about other agents' strategies. Rational agents use strategic reasoning to make their choices because it takes account of anticipated responses.

Strategic reasoning involving anticipation of responses is most clearly exemplified in games where an agent makes a move observed by other agents, and then the other agents make moves observed by the agent, and so on. For instance, in Tic-Tac-Toe, if the agent who plays first, X, takes the middle, then the agent who plays second, O, takes a corner, reasoning that if she takes any other spot, X will respond in a way that brings him victory as in Figure 1.3, where X's first move is X1, his second move is X2, and so on, and similarly for O. O uses her strategically inferred information about X's responses to her strategies to guide her selection of a move.

Some games do not have a sequence of publicly observed moves. In *single-stage games* all agents choose during a single period using initial information about the game, and their choices are not made public until the period ends. Games where choices are made simultaneously are single-stage. But simultaneous choice is not essential. The characteristic feature of single-stage games is that no agent can observe the choices of other agents before choosing. Information about their choices must come, directly or

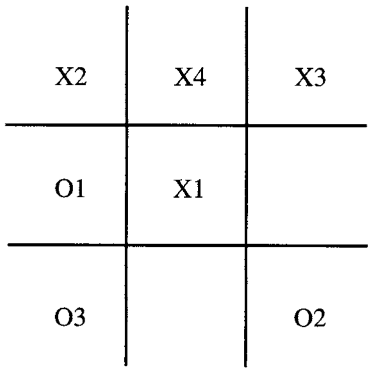

Figure 1.3 Tic-Tac-Toe.

by inference, from initial information about the game. I treat single-stage games of strategy.

In single-stage games of strategy, strategic reasoning may involve an agent's anticipation of other agents' choices, and his anticipation of their choices may depend on his information about their abilities to anticipate his own choice. An agent may know that other agents anticipate a certain choice by him and that they make their choices according to what they anticipate. For instance, in Scissors-Paper-Stone, an agent may know that his opponent anticipates Paper from him and that she will play Scissors as a result. The agent may then surprise his opponent by playing Stone.

Strategic reasoning in single-stage games of strategy is analogous to strategic reasoning in multistage games, except that instead of considering the response of other agents to a move, an agent considers the evidence a strategy provides for other agents' strategies. Strategic reasoning plays a role via foresight of the incentives that obtain if a choice is made. An agent has foresight of those incentives either because of anticipation of the other agents' response to his choice or because of anticipation of changes in the circumstances for incentives given his choice. An agent's incentives are relative to his choice, and strategic reasoning uses knowledge of this relativity to make a choice. Since the analogy between strategic reasoning in single-stage and multistage games is strong, we use the terminology of strategic reasoning in multistage games to describe strategic reasoning in single-stage games. We speak of the "response" of other agents to an agent's strategy in a single-stage game if his strategy is conclusive evidence for certain strategies by other agents, even though, strictly speaking, because his strategy has no causal influence on their strategies, they do not respond to his strategy.

In recent literature Nash equilibrium is generally taken as the appropriate conception of equilibrium only for single-stage games, such as Matching Pennies, as opposed to multistage games, such as chess. Multistage games raise issues of strategy ignored by Nash equilibrium, as Selten (1965) observes. Although I believe that our new type of equilibrium applies to multistage games, I do not work out the details. Its main advantage over Nash equilibrium in multistage games is that it draws on more features of a game than does Nash equilibrium. Nash equilibrium takes into account only the features represented by a game's *payoff matrix*, that is, a table giving the payoff for each agent of each strategy profile. The new type of equilibrium takes account of other strategic considerations ignored by Nash equilibrium.

I focus on a particular type of single-stage game of strategy – *noncooperative* games. Noncooperative games are sometimes characterized as games without opportunities for binding agreements. It is better to take them as games where the agents are not coalitions or are not treated as coalitions. Binding agreements are just a means of forming coalitions that act as agents. The inability of agents to form coalitions and act jointly may arise because they are unable to communicate, because their interests conflict, or because their agreements cannot be made binding.

It is also better to apply the distinction between the cooperative and the noncooperative to representations of games rather than to games themselves. A representation of a game standardly identifies, among other things, the agents and their strategies. In a noncooperative representation of a game, the agents are individuals, or groups treated as individuals. In a cooperative representation of a game, the agents are coalitions, including single-membered coalitions, of individuals. The coalitions are treated as coalitions, not as individuals; such treatment entails acknowledging that they may not form, and so assigns them the strategies of formation and nonformation. A single game may receive both a cooperative and a noncooperative representation. A game with a cooperative representation may be represented as a noncooperative game by reducing the strategies of coalitions to combinations of strategies of individuals. Generally, the reduction requires treating a game with a single-stage, cooperative representation as a multistage, noncooperative game in which individuals make offers of collaboration, counteroffers, and the like, before they realize a coalition structure and joint strategies for its coalitions. The strategies assigned to the members of a coalition by a profile in the game's noncooperative representation constitute a strategy for the coalition in the game's cooperative representation.

We treat noncooperative games because Nash equilibrium is generally proposed as the appropriate conception of equilibrium for them and because noncooperative games form the fundamental type of game. Cooperative games are in principle reducible to noncooperative games. Although

Cambridge University Press
978-0-521-03802-7 - Equilibrium and Rationality: Game Theory Revised by Decision Rules
Paul Weirich
Excerpt
More information

I investigate equilibrium in noncooperative games only, my definition of equilibrium is intended for cooperative games as well. Its extension to cooperative games faces two issues, however. First, one must fully work out a method of treating a coalition as an agent so that self-support is precisely defined for a coalition's strategies. Second, one must specify the idealizations that are appropriate for cooperative games in order to verify that equilibria exist in all ideal cooperative games. I leave these two issues open but assume they can be settled favorably.[1]

In noncooperative games of strategy an agent's assumption of a choice affects his incentives. That is, incentives are relative to choices. This is the distinguishing feature of games of strategy. But furthermore, in noncooperative games a choice affects incentives through evidence about other agents' responses to the choice, not through changes in the circumstances of choice, as may happen in cooperative games. Agents can anticipate to some degree responses to their choices. This anticipation of responses is not direct. It is not, for example, a product of communication between agents as in cooperative games. It is indirect and involves reasoning. The reasoning may be strategic and go from an agent's choice to the response of other agents, or it may be nonstrategic and rest, for example, on psychological insight into the choices of other agents.

We focus more narrowly still on the noncooperative games of strategy known as *normal-form* games. Normal-form noncooperative games are generally distinguished from *extensive-form* noncooperative games. Again, the distinction fits representations of games better than games themselves. A normal-form representation of a game lists agents, strategies, and payoffs. An extensive-form representation of a game also indicates the order of moves. The representation is nondegenerate only if the game is multistage and so has a nontrivial order of moves. Strategies of a multistage game's normal-form representation are more complex than strategies of its extensive-form representation. In its normal-form representation strategies are contingency plans specifying the move to be made given various moves by other agents. Because the contingency plans, or strategies in a narrow sense of the word, are not represented by the moves composing them, normal-form representations are sometimes called "strategic-form" representations.

The normal-form representation of a game is generally proposed for single-stage games in which strategies are causally independent, or for multistage games where solutions are independent of the order of moves. The extensive-form representation of a game is generally proposed for multistage

[1] For a treatment of coalitions as agents in ideal cooperative games and for an exposition of the view that solutions to such games reveal collective preferences and collective utility, see Weirich (1990, 1991a, 1991b).