Computer Simulations in Game Theory

Paul Weirich
Philosophy Department
University of Missouri
Columbia, MO 65203, USA

A computer simulation runs a model generating a phenomenon under investigation. For the simulation to be explanatory, the model has to be explanatory. The model must be isomorphic to the natural system that realizes the phenomenon. This paper elaborates the method of assessing a simulation's explanatory power. Then it illustrates the method by applying it to two simulations in game theory. The first is Brian Skyrms's (1990) simulation of interactive deliberations. It is intended to explain the emergence of a Nash equilibrium in a noncooperative game. The second is Skyrms's (2004) simulation of the evolution of cooperation. It is intended to explain cooperation in assurance games. The final section suggests ways of enhancing the explanatory power of these simulations.

1. The Explanatory Power of Simulations

A simulation runs a model, in particular, a dynamic model. Although a model may be static or dynamic, a simulation runs a model of a process. Not all simulations have explanatory goals. A machine selling train tickets may simulate speech. A machine issuing game-board instructions may simulate playing chess. These simulations succeed if they produce appropriate behavior.

Some simulations may aim not only to produce an outcome but also to produce it as nature does. Simulations that represent an organism's behavior may replicate natural processes because evolution has optimized production of the behavior. A machine simulating speech may replicate human cognitive processes because they efficiently generate speech. The simulation's goal may be to produce speech as humans do. Then it may combine the goals of producing and explaining speech.

In some cases simulations have a predictive rather than an explanatory goal. A simulation of an economic market may aim to predict the market's outcome without trying to replicate the process that produces it. A predictive simulation may depart from nature to implement a reliable shortcut. However, the goals of prediction and explanation generally go hand in hand. Explanatory mechanisms afford the most accurate predictive devices.[1]

One evaluates a simulation with respect to a goal, so an evaluation begins by identifying the goal. First, what phenomenon does the simulation study? Simulations of the emergence a signaling system under various conditions may, for example, examine the prevalence of signaling systems rather than the process of their emergence. Second, does the simulation aim only to reproduce a phenomenon, or also to reproduce the process behind the phenomenon's natural occurrence? If it has the latter goal, which aspects of the process does it seek to replicate? Does it, for example, try to replicate the speed of the process? Third, does the simulation seek to explain the phenomenon? If so, then it must replicate components of the natural process that explains the phenomenon.

---

[1] Simulations in the social sciences are diverse and raise a variety of interesting philosophical issues. For a good introduction, see Rainer Hegselmann, Ulrich Mueller, and Klaus Troitzsch (1996).

This paper examines simulations that aim to explain a natural phenomenon's occurrence. They aim to replicate the process of the phenomenon's occurrence. The simulations treat interactive human behavior. They target Nash equilibrium, coordination, and cooperation in simultaneous-move games. Although many simulations have these goals, the paper focuses on Brian Skryms's computer simulations of interaction deliberations and of the evolution of social norms. It presents a standard of explanatory success and by way of illustration applies it to Skryms's simulations.

May simulations using cellular automata and agent-based models explain behavior in games? The models and simulations use assumptions that do not hold for people. They may seem unrealistic and unable to explain human behavior. May simulations such as Skryms's explain human deliberation and cooperation despite their unrealistic assumptions?

Paul Humphreys (2004: 132) point out a problem with agent-based modeling that uses only simple local influences on individuals to obtain complex macro phenomena. He observes, "Because the goal of many agent-based procedures is to find a set of conditions that is sufficient to reproduce the behavior, rather than to isolate conditions which are necessary to achieve that result, a misplaced sense of understanding is always a danger." As he warns, "It is often possible to recapture observed structural patterns by using simple models that have nothing to do with the underlying reality" (p. 134).

An argument that a simulation has explanatory power typically shows that the simulation robustly generates the target phenomenon. It shows that the simulation generates the phenomenon even if details of the simulation change. Skyrms, for example, supports the explanatory power of his simulations of human interaction by demonstrating robustness with respect to granularity of options, neighborhoods of interaction, learning rules, and so on. Robustness is not enough for explanation, however. For example, a nonexplanatory chess program, using brute force instead of pattern recognition, may be robust with respect to various methods of evaluating prospects. Robustness is no more than a part of an adequate standard of explanation. Also, only certain forms of robustness are desirable. A simulation should not robustly yield the same results as it becomes more and more realistic. As it replicates additional explanatory factors, results should change. An adequate explanatory standard specifies the types of robustness a simulation should achieve.[2]

The next section presents a standard for a simulation's explanatory success. The standard is, roughly, imitation of a natural process. That requires running a model that incorporates only explanatory factors. Later sections argue that Skryms's models and simulations incorporate nonexplanatory factors. That weakens their power to explain human behavior. The last section shows that adding strategic reasoning to his models and simulations increases their explanatory power.

2. Partial Explanations

[2] Robert Axlerod (1984) argues that his computer simulations of the repeated prisoner's dilemma explain the evolution of cooperation. They illustrate the virtues of the strategy of tit-for-tat. John Kagel and Alvin Roth (1995: 29) observe that tit-for-tat's success in those simulations is sensitive to the field of alternative strategies. A strong case for tit-for-tat requires showing that it is robustly victorious.

Ron Giere (1988: Chap. 3) offers an account of the evaluation of models. To illustrate his account, he presents Hooke's law together with idealizations for it (pp. 68–70). The law is not exactly true because the idealizations do not hold for natural systems (pp. 76–78). The law belongs to a model, and a model is an idealized system. A model does not have the same structure that a natural system has. It may nonetheless explain a natural system by being similar to the natural system (pp. 80–81). This section refines Giere's standard of evaluation.

There are various types of explanatory success. One type of success is just approximating a natural system, as Giere observes. This section examines another type of explanatory success, namely, partial explanation. A partial explanation requires an exact treatment of some explanatory factors. A model that merely approximates the workings of some explanatory factors does not yield a partial explanation. To yield a partial explanation, it must treat their workings with precision.

A model may yield a partial explanation despite resting on false assumptions. Its assumptions may put aside some explanatory factors to facilitate treatment of other factors. The partial explanation it offers is a step forward. An expansion of the model may yield a more complete explanation. I consider whether a model succeeds as a partial explanation, because enjoying that type of explanatory success facilitates theoretical progress.

Simulations use simplifying assumptions to make their dynamics manageable. Reliance on simplifying assumptions is compatible with achieving a partial explanation. A standard method of building a full theory of a phenomenon starts with a simplified theory. Later stages of the method remove simplifying assumptions to make the theory more general. A simplified theory has explanatory power as long as its assumptions control for explanatory factors. In that case the simplified theory, although not general, exhibits the workings of remaining explanatory factors and yields a partial explanation. One may improve a simplified model by making it cover more features of the phenomenon to be explained in a wider range of cases. My assessment of models condones simplifications and just asks for a partial explanation.

Similarity is not necessary for a partial explanation. Models in game theory and economics do not achieve similarity. Their ideal agents are nothing like real agents. Yet the models yield partial explanations. They yield exact accounts of some explanatory factors after controlling for other explanatory factors. Also, similarity is not sufficient for a partial explanation. A partial explanation treats precisely some factors involved in a phenomenon's production. To obtain a standard for a model's success as a partial explanation, I replace Giere's appeal to similarity with an appeal to a type of isomorphism.

The previous section noted that theorists argue for a simulation's success by demonstrating its robustness. Although robustness is not enough for explanation, robustness bears on a simulation's success in generating a partial explanation. Studies of robustness may show that a simulation responds to an irrelevant feature of a case. Then the simulation does not obtain the target phenomenon from explanatory factors. It fails as a partial explanation. Studies of robustness furnish evidence concerning explanatory power. May robustness be combined with other criteria to obtain a test of partial explanation?

Justin D'Arms, Robert Batterman, and Krzyzstof Górny (1998) present guidelines for evaluating simulations of human behavior in evolutionary game theory. They endorse game theoretic explanations of human behavior that allow us to abstract from misleading particulars to better recognize and appreciate broad patterns in the phenomena of human social life. Their guidelines for assessment of simulations and models of adaptive behavior are being representative, robust, and flexible (pp. 89–92)). A model is representative if its structure is realized frequently in the environment of the evolutionary adaptation. A result is robust if it is achieved across a variety of different starting conditions and/or parameters. A model for an adaptive strategy is flexible if it presents processes that various mechanisms may follow to realize the adaptive strategy.

Robustness together with representativeness and flexibility is still not enough for a partial explanation, however. To offer a partial explanation, a model must be exactly right about the explanatory factors it treats. Robustness, representativeness, and flexibility are all matters of degree and even when attained to a high degree need not yield a partial explanation of a phenomenon. For example, Skyrms's (2004) account of cooperation in the stag hunt meets those criteria but still does not yield a partial explanation of cooperation, as Section 4 argues.

A simulation shows the results of assumptions and so resembles a theorem. I apply to simulations common methods of evaluating the explanatory power of theorems. Suppose that a theorem states a dynamic model's results. Then the model and a simulation running it are explanatory only if the theorem is explanatory. The theorem is explanatory if its assumptions control for some explanatory factors and its results reveal the workings of other explanatory factors. The theorem's proof may show how the theorem's results arise from its assumptions.[3]

A simulation may be so complex that no one understands the process by which it produces results. There may be no theorem deriving its results from its assumptions. Nonetheless, the simulation may be explanatory as long as its model is explanatory. Its model may be explanatory because its assumptions control for explanatory factors. One may establish this despite incomplete information concerning the model's dynamics.

A model may explain the working of some explanatory factors. It may propose laws for them that yield a simulation's outcome, even if the details of the process remain beneath the surface of the simulation. It is not necessary to have analytic results confirming the outcome of the simulation. The model may yield a partial explanation of a phenomenon without offering a transparent process for production of the phenomenon. It may succeed by identifying factors that generate the phenomenon in a controlled environment.

Because the explanatory standards for models and theorems are the same, one may illustration those standards by applying them to a theorem. Take Leonard Savage's (1954) representation theorem showing that, given certain axioms of preference, probability and utility functions may represent preferences. Typically, a person's preferences do not satisfy the axioms. So how can the theorem explain anything about people? To be explanatory, the theorem's assumptions must meet certain conditions. Not every theorem resting on simplifying assumptions has explanatory value. Savage's theorem uses the assumption that preferences are transitive. That is a suitable assumption

[3] As Stephan Hartmann (1996: 84) observes, even when an analytic treatment of a simulation's result is available, the simulation may still be useful. It may display the process producing the result.

because the preferences of a rational ideal agent are transitive. The theorem also assumes an infinite number of preferences. That is an unsuitable assumption because a rational ideal agent may have only a finite preference structure. Hence the theorem only partly meets explanatory standards.

As a theorem's explanatory power depends on its assumptions, the explanatory power of a simulation also depends on its assumptions, that is, the assumptions of the model it runs. Do the assumptions control for explanatory factors? Are operative factors in the model isomorphic to operative factors in the target system? If so, then the model is explanatory. Consider, for example, a model of rational decision making for an agent with full information. According to the model, the agent maximizes utility. The model controls for an explanatory factor, namely, an agent's information. The remaining operative factors, namely, the utilities of options, are explanatory. So, the model succeeds. It is explanatory despite the unrealistic assumption of full information because that assumption controls for a factor in the explanation of a decision's rationality, namely, the agent's information. The assumption sets that factor at full information.

Evaluating a model's assumptions for control of explanatory factors requires an inventory of explanatory factors and methods of controlling them. Individuation and identification of explanatory factors is a theoretical enterprise. Methods of controlling for them are products of investigation. A model's evaluation therefore uses hypotheses about explanatory factors. An evaluation may be mistaken because it misconceives the structure of a general theory. Using a partial theory as a step toward a complete theory is a boot-strapping procedure. One uses an outline of the complete theory to guide construction of the partial theory and also uses the partial theory to guide steps toward the complete theory.

Assumptions that control for explanatory factors are sometimes called abstractions and idealizations. To clarify these assumptions, I compare them with abstractions and idealizations as Paul Humphreys (2004: 141-47) characterizes those procedures.[4] According to Humphreys, abstraction (in one sense) removes some properties without changing remaining properties. As he says, "Abstraction$_1$ is the process of omitting properties or their related predicates from a system or model, the omission process being either cognitive, representational, or physical, leaving the remaining properties or predicates unchanged" (p. 143). As an example of abstraction, he mentions reducing to zero the Coulomb force between electrons in a model of the atom.[5]

A model for a phenomenon may use abstraction. It may remove properties and study the effect of remaining properties on the target phenomenon. Removing an explanatory factor is a way to control for it. Therefore abstraction resembles adopting assumptions that control for explanatory factors.[6] However removal of a property controls for an

---

[4] Weirich (2004) classifies as idealizations all assumptions that control for explanatory factors. For simplicity, I put aside that work's terminology for classifying assumptions.

[5] Humphreys holds that abstraction increases generality because removing some properties means that more systems have the remaining properties (p. 146). Abstraction has an opposite effect on another type of generality. Adding assumptions that implement abstraction decreases a model's generality because then fewer systems satisfy the model's assumptions.

[6] Humphreys uses the term "properties" to cover relations, and treats an individual as a bundle of properties. Thus abstraction covers a wide range of explanatory factors. However, he also imposes restrictions on properties and, for instance, does not assume that if *P* is a property, then not-*P* is a property. His restrictions reduce the range of explanatory factors that abstraction may remove. To facilitate

explanatory factor only if the property is an explanatory factor. Abstraction of a property does not count as control of an explanatory factor if the property abstracted, say, color, plays no role in the target phenomenon's production.[7]

Also, one may control for an explanatory factor not by removing it but by holding it constant. For example, one may hold constant an agent's preferences in a series of decisions. That step facilitates inferences moving from one decision to another. This type of control resembles idealization in Humphreys's sense. He says, "Idealization involves taking a property and transforming that property into one that is related to the original but possesses desirable features which are lacking in the original" (p. 146). For example, an idealization may treat the planets as perfect spheres. It sets shape at a convenient, constant value. However, adding a desirable feature does not always control for an explanatory factor. For example, some decision principles may treat only decision problems with finite utility for the sake of mathematical tractability, or to put aside vexing puzzles about infinite utility. However, in an account of rational decisions, the range of the utility scale is not an explanatory factor. So the assumption of finite utility, although it introduces a desirable feature, does not control for an explanatory factor.

How does understanding the workings of uncontrolled factors contribute to understanding the workings of all factors? May the uncontrolled factors behave differently when other factors are present? Explanatory factors carve at the joints. Although they may interact differently in some respects when more factors are present, they always obey the same basic laws. They are the factors for which the basic laws of interaction are constant.[8] Because laws governing explanatory factors are constant, a law governing uncontrolled factors is an instance of the law governing all factors. The laws governing an explanatory model are instances of general laws. For example, a model of decision making that assumes full information asserts that agents maximize utility. A more general model without that assumption asserts that agents maximize expected utility. The principle to maximize utility is an instance of the general principle to maximize expected utility. In many cases, being ignorant of general laws, one cannot verify that a model's law is an instance of a general law. However, an outline of a general law may support a judgment that a model's law is an instance.

A model whose assumptions control for explanatory factors offers a partial explanation because it is isomorphic to a natural system. Suppose that a natural system has three components, and a model controls for one component to investigate the remaining two components. Some aspects of the interaction of the remaining components may be altered by the absence of the one removed. To be explanatory, a type of isomorphism must nonetheless hold between the model and the natural system. Suppose that the law for the two-component model asserts that *Rab* and the law for the three-component system asserts that *Sabc*. The relevant type of isomorphism holds if the first law is an instance of the second law.

---

comparison of abstraction and control of explanatory factors, I assume that explanatory factors are properties.

[7] Because removing a property is vague, it is best to use an assumption to state the manner of the property's removal. For example, to remove the property of friction, one may assume that objects are frictionless. I do not assume that under assumptions removing properties, remaining properties are completely unaffected. Abstracting away cognitive limits changes an agent's inferential power, for example.

[8] Whether such factors exist is an empirical matter, but theoretical science searches for them under the working hypothesis that they exist.

An isomorphism between a model and a natural system may serve various purposes. Van Fraassen (1980: 64) uses such an isomorphism as a test of a theory's empirical adequacy. He says, "To present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those models (the *empirical substructures*) as candidates for the direct representation of observable phenomena. The structures which can be described in experimental and measurement reports we can call *appearances*: the theory is empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model." Van Fraassen means by model an interpretation of a formal theory that makes the theory true. This sense of model differs from but is similar to the sense of model in which a simulation runs a model.

The type of isomorphism between model and natural system that van Fraassen considers resembles the type of isomorphism that generates explanatory power. However an explanatory isomorphism has special features. First, it does not involve individuals because their interactions in nature are not the same as in the model. Instead, it involves explanatory factors and their interaction. The explanatory factors are the same in the model and in the natural system. Second, the isomorphism preserves the laws governing the explanatory factors. The law governing their interaction in the model is an instance of the law governing their interaction in the natural system. These relationships establish an explanatory isomorphism between the model and the natural system.

An isomorphism involving properties and their interaction may not establish explanatory power. A model and a natural system may involve properties, a single law of the model, and a single law of the natural system. If the law of the model and the law of the natural system are both true, then the model is isomorphic to the natural system under the identity function from properties in the model to properties in the natural system. An explanatory isomorphism requires in addition that the properties be explanatory factors and that the law of the model be an instance of a law of the natural system. A partial explanation requires a particular type of isomorphism between a model and a natural system. The model must capture the right structures in the right ways. The elements must be explanatory factors, and their relation in the model must be an instance of their relation in the natural system.

The model earlier taken from decision theory exhibits the required type of isomorphism. The elements of the model are the properties of being an agent with a decision problem, having a utility assignment for options in the problem, and maximizing utility in the problem with respect to the options' utility assignments. The law governing the properties in the model states that an agent maximizes utility. This is an instance of the general law relating the properties in a natural system where agents may lack full information. The general law states that an agent maximizes expected utility and that an option's expected utility equals its utility.[9]

3. Simulation of Interactive Deliberations

A simulation represents a process. The moves in a simultaneous-move game do not constitute a process, but the deliberations of agents do. A simulation may replicate their deliberations. This section evaluates Brian Skyrms's (1990: Chap. 2) simulations of

[9] These two laws also adopt background assumptions about an agent's rationality and cognitive power.

deliberations in simultaneous-move, noncooperative games. Those simulations aim to explain the emergence of a Nash equilibrium. Skryms derives a Nash equilibrium's realization from principles for agents with bounded rationality. This section argues that because the derivation excessively restricts the agents' cognitive abilities, it does not explain why people realize a Nash equilibrium in a game of strategy.

As Skyrms envisages agents, they deliberate in stages instead of all at once. They use the results of one stage of deliberation as additional information for the next stage. Because agents in games of strategy try to take account of each other's deliberations, the results of an agent's deliberation in one stage often provide evidence about other agents' strategies. The agent can then use this evidence in his deliberation's next stage.

An agent begins deliberation with an initial assignment of probabilities to his strategies and those of other agents. Using the probabilities for others' strategies, he finds a strategy for himself that maximizes expected utility. Typically, he then increases that strategy's probability. He does not increase its probability all the way to one because he is aware that his probability assignments are tentative. They do not accommodate all relevant considerations, so he expects revisions as his deliberations proceed. Next, the agent revises his probability assignment for other agents' strategies in light of his new probability assignment for his own strategies. Then using the revised probabilities, he recalculates the expected utilities of his own strategies and readjusts their probabilities. The process of revision uses each stage of deliberation as input for a rule of bounded rationality that brings the agent to the next stage. This process continues until the probabilities of the agent's strategies do not lead to any further revision in his probability assignment for other agents' strategies. When all the agents reach this stopping point, they achieve a joint deliberational equilibrium. In suitable conditions this joint deliberational equilibrium is a Nash equilibrium.

Figure 1 illustrates a possible course of deliberations in the game of Matching Pennies. In that game two agents simultaneously display pennies, one seeking a match in the sides displayed and the other a mismatch. In Figure 1, Agent 1 seeks a match and Agent 2 seeks a mismatch. The agents know each other's initial probability assignments and updating rules so that they know each other's probability assignments at every stage of deliberations. Deliberations start with each agent assigning a probability of slightly more than 0 to playing Heads. Agent 2 then starts to increase the probability of Heads, seeking a mismatch. After the probability of Agent 2's playing Heads exceeds 0.5, Agent 1 begins to increase the probability of Heads also, seeking to obtain a match. When the probability of his playing Heads exceeds 0.5, Agent 2 begins to decrease the probability of Heads. Mutual adjustment of probability assignments leads to the joint deliberational equilibrium in which each agent settles on the mixed strategy of playing Heads and Tails with equal probability. Their mixed strategies form a Nash equilibrium.

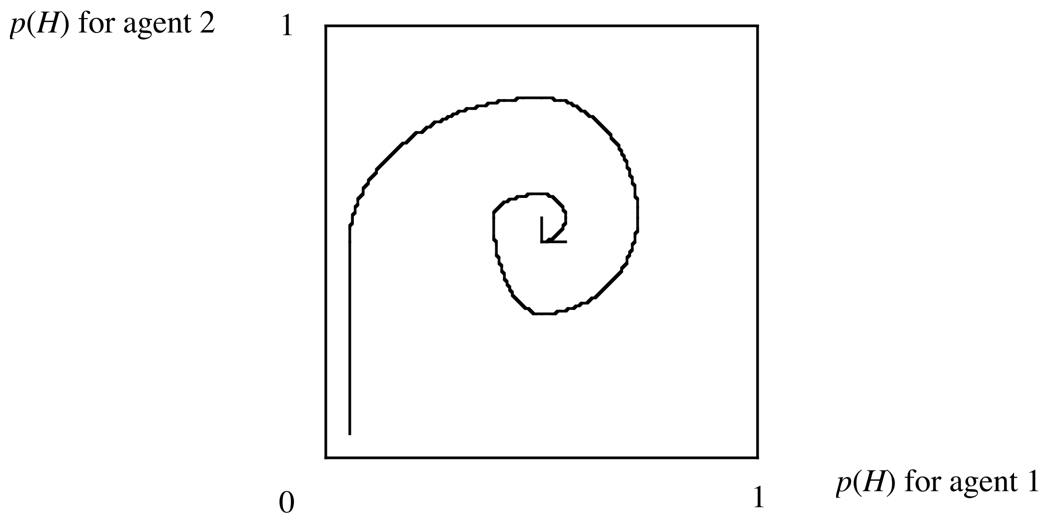

Figure 1

In games with multiple Nash equilibria, the equilibrium reached depends on the agents' initial probability assignments (p. 34). Although the dynamics do not explain the particular equilibrium reached, Skyrms holds that they explain why deliberations culminate in a Nash equilibrium. He shows a theorem with the following form. Under certain assumptions, “A joint deliberational equilibrium on the part of all the players corresponds to a Nash equilibrium point of the game” (p. 29). This result has intrinsic interest because it shows how a Nash equilibrium in a game may arise from the agents’ deliberations. However, the assumptions on which it rests are too restrictive to yield a partial explanation of a Nash equilibrium's realization.

Skyrms’s theorem shows that given certain assumptions, rational agents reach a joint deliberational equilibrium if and only if they reach a Nash equilibrium. But the theorem’s assumptions provide no guarantee that the agents reach a joint deliberational equilibrium. Under the assumptions, whether the agents reach a joint deliberational equilibrium depends on their initial probabilities and their rules for revising probabilities. For instance, agents who revise probabilities using a noncontinuous rule may not reach a joint deliberational equilibrium even though one exists. They may skip over it (p. 32). Furthermore, in the game of Matching Pennies “Aristotelian” deliberators, who revise probability assignments according the continuous rule $dp(H)/dt = EU(H) - EU(T)$, may fail to reach a joint deliberational equilibrium (pp. 80–81). Given almost every initial probability assignment, they circle the 50–50 mixed-strategy Nash equilibrium as in Figure 2. Because the theorem’s assumptions do not guarantee a joint deliberational equilibrium, the theorem provides at best only a partial explanation of realization of a Nash equilibrium. It supports Nash equilibrium only in cases where the agents reach a joint deliberational equilibrium.

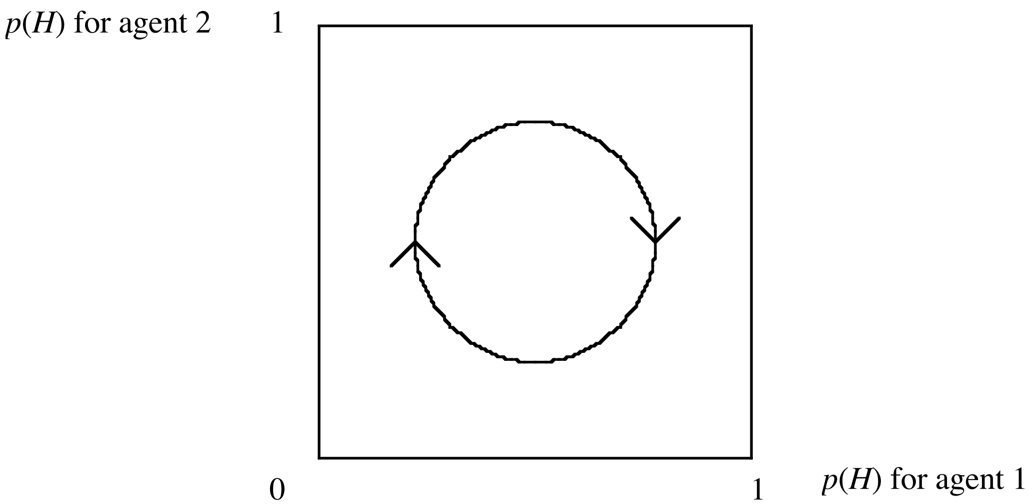

Figure 2

In Skyrms's model of deliberation agents reach a Nash equilibrium by adjusting their probability assignments in light of their knowledge of other agents' probability assignments (p. 33). This is a type of strategic reasoning, but even short-sighted agents may exercise it. In characterizing the deliberators in his examples he says, "These players follow their noses in the direction of the current apparent good, with no real memory of where they have been, no capability of recognizing patterns, and no sense of where they are going" (p. 152). His model does not require that the deliberators be as simple-minded as this, but the explanation of Nash equilibrium he gives succeeds for such deliberators. As he observes (pp. 58, 152), his model of deliberation resembles the evolutionary models of animal behavior proposed by John Maynard Smith (1982). It does not require looking ahead to future stages. It has a dynamics suited to mindless organisms as well as humans.

Skyrms's model of deliberation for agents with bounded rationality is obviously useful in applications of game theory to humans. Also, it is clear that his theorem exposes an interesting connection between joint deliberational equilibrium and Nash equilibrium. The significance of this connection is moreover supported by his investigations of its robustness (pp. 62–86). However, the theorem's assumptions have several drawbacks with respect to the project of obtaining a partial explanation of realization of a Nash equilibrium. For theoretical appeal, the assumptions should control for factors that explain rational behavior. This section argues that the theorem's assumptions fall short with respect to that standard.

The theorem's assumptions appear in this summary: "In a game played by Bayesian deliberators with a common prior, an adaptive rule that seeks the good, and a feedback process that updates by emulation, with common knowledge of all the foregoing, each player is at a deliberational equilibrium at a state of the system if and only if the assignment of the default mixed acts to each player constitutes a Nash equilibrium of the game" (p. 33). A common prior is a common initial assignment of probabilities to all the strategies of all the players. An adaptive rule that seeks the good revises the probabilities of an agent's strategies, after calculations of their expected utilities, so that only strategies with a higher expected utility than the agent's tentative mixed strategy receive a higher probability and so that the sum of the probabilities of such strategies increases. A feedback process that updates by emulation revises the probabilities assigned to other agents' strategies by replicating their revision procedures. Common knowledge of the process of deliberation entails that all know it, know that all know it, and so on. The default mixed act of a player is the randomized strategy that follows the agent's current assignment of probabilities to his pure strategies.

Taking Skyrms's theorem as an explanation of realization of a Nash equilibrium in a game of strategy, the role of an assumption is either to facilitate the attainment of joint deliberational equilibrium or to bring joint deliberational equilibrium into alignment with Nash equilibrium. My evaluation of the theorem's assumptions examines the way in which they fulfill these roles.

The first drawback is that some of the assumptions do not control for factors that explain rational behavior. To show that an assumption fails this test without advancing an inventory of factors that explain rational behavior, I show that modifying the assumption in a way that has no significance according to a general theory of rationality (insofar as its contours are known) either blocks joint deliberational equilibrium or else upsets the alignment of joint deliberational equilibrium and Nash equilibrium.

Some assumptions meet the standard of evaluation. Take the common knowledge assumption. It controls an explanatory factor, namely, each agent's knowledge about others. Clearly, a general theory of rational behavior uses this factor to help explain rational behavior in strategic situations. Likewise, the assumption of a common prior, in the context of the common knowledge assumption, controls each agent's knowledge about others.

The other assumptions, however, lack this sort of theoretical warrant. Take the assumption that all agents have the same adaptive rule for revision of probabilities. Because, as Skyrms allows, many adaptive rules are rational, this assumption does not control for an explanatory factor. Also, consider the assumption that the adaptive rule seeks the good. This assumption excludes adaptive rules that permit raising the probabilities of strategies that are just as good as one's tentative mixed strategy. Because probability increases of this sort are not irrational, excluding adaptive rules that permit them does not control for an explanatory factor.

Moreover, the exclusion of the permissive adaptive rules is not merely a matter of technical convenience. It is necessary for the correspondence between deliberational and Nash equilibrium. To see this, consider a game with a unique Nash equilibrium that is not strict. Given a permissive adaptive rule, the agents may be at a Nash equilibrium but not achieve joint deliberational equilibrium. They may oscillate away from and back to the Nash equilibrium until lack of time forces them to halt deliberations, perhaps away from the Nash equilibrium.

The assumption that the agents update by emulation is also nonexplanatory. According to it, an agent's stage n deliberations about others replicate their stage n deliberations. It has the consequence that an agent learns about others' stage n deliberations during his stage n deliberations. But this feature of an agent's learning lacks theoretical motivation. Why should an agent learn about others' stage n deliberations during his stage n deliberations as opposed to his stage $n - 1$ or $n + 1$ deliberations? The symmetry of the agents' situations does not require their learning to be in step with deliberations rather than a step behind or ahead. It requires only that each agent's learning follow the same pattern. Rationality, although it requires gathering information about others, does not require that agents in stage *n* deliberations gather information about others by replicating their stage *n* deliberations as opposed to their stage $n - 1$ or $n + 1$ deliberations. Because predictions about others are subject to revision in later stages of deliberation, they need not stem from replications of any particular stage.

However, the timing of information about others is critical for the attainment of a joint deliberational equilibrium. Suppose that in a two-person game, one agent's learning about the other agent falls a step behind the other agent's deliberations. It is as if the agents take turns going through the stages of deliberation. Then a joint deliberational equilibrium may move out of reach. In the game of Matching Pennies, for instance, each agent may end up oscillating perpetually between two mixed acts. When Agent 1, who wants a match, learns of an increase in the probability of Heads for his opponent, he raises the probability of his playing Heads. Then, when Agent 2 learns of this change, she lowers the probability of her playing Heads. This triggers a decrease in the probability of Agent 1's playing Heads, and so on. As a result, the agents do not reach the mixed acts that constitute deliberational and Nash equilibrium. They trace a box around the Nash equilibrium as in Figure 3 until time for deliberation ends.

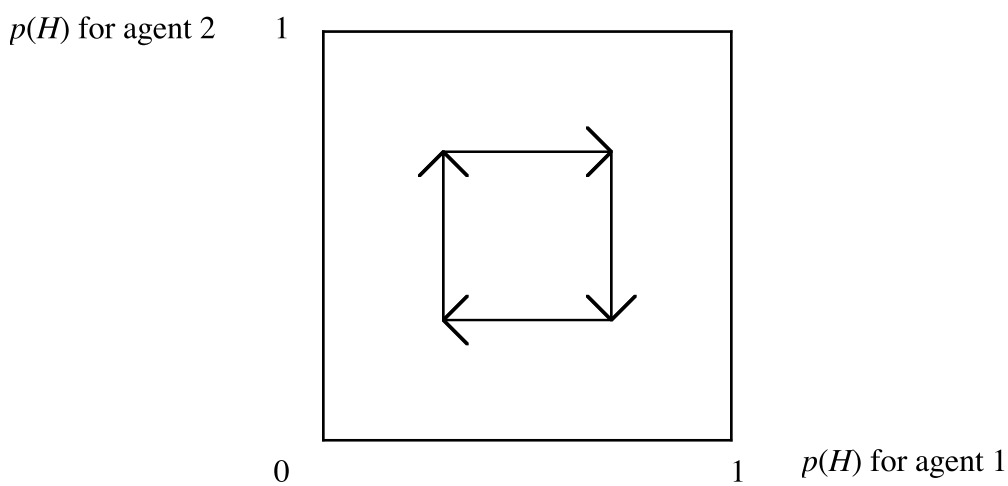

Figure 3

Do Skyrms's simulations of joint deliberations partially explain realization of a Nash equilibrium? As shown, not all assumptions control for explanatory factors. Although assumptions about timing are critical, they do not control for an explanatory factor. The outcomes people achieve in simultaneous-move, noncooperative games do not depend on the timing of their deliberations. The simulations should be robust with respect to assumptions about the timing of agents' deliberations but are not robust. Thus, Skyrms's simulations of interactive deliberations do not generate partial explanations of a Nash equilibrium's realization. Section 5 proposes modifying the dynamics of Skyrms's simulations to mitigate the effect of timing on attainment of an equilibrium.

4. Simulation of Behavioral Evolution

Skyrms (2004) offers computer simulations to explain the emergence of cooperation. The simulations run evolutionary dynamics that yield cooperative behavior. To the basic replicator dynamics, he adds various ways for organisms to interact with their neighbors in a population. His simulations use the stag hunt as the paradigm problem of cooperation and show how a population may reach an efficient equilibrium despite obstacles such as risk. They also treat coordination in the game divide-the-dollar. His simulations apply to many types of organism, but this section examines their application to people only. It asks whether the simulations explain human behavior.

Cultural evolution and genetic evolution both yield evolution of behavior. Models of cultural evolution use the replicator dynamics applied to behavior rather than to genes. Skryms takes this approach and so simulates cultural evolution (p. 10). His target is cultural evolution of cooperation and also the speed of cooperation's emergence.

Part I of Skyrms's book presents the simulations I treat. Other parts modify those simulations to facilitate cooperation. I do not review the modifications. My objective is just to illustrate evaluation of a simulation, and the simulations in Part I furnish good examples. Also, the book's modifications of those simulations do not rectify the explanatory shortcomings that this section identifies.

In the preface Skyrms says of his book, "Each part isolates a simple empirically motivated modification of the most thoroughly studied evolutionary model and shows how the modification makes a dramatic difference in the evolutionary dynamics of the stag hunt and related interactions" (p. xiii). The responsiveness of models to simple modifications highlights a question about explanatory power. Does a model explain cooperation if simple modifications make a dramatic difference? Skyrms believes that the simulations explain the emergence of cooperation (p. 123). He supports his claim by studying robustness of results. However, changes in some nonexplanatory factors influence results. Changes in those factors do not affect results among people but do affect results among agents in the model. That lack of robustness shows that the factors controlling results in the model are not the factors controlling results in the world. The model's assumptions do not control for explanatory factors, and so the model does not yield a partial explanation of cooperation.

Skyrms controls for some explanatory factors governing the emergence of cooperation, such as incentives to exploit cooperators. He controls for these incentives by addressing the stag hunt rather than the prisoner's dilemma and by adjusting the payoff matrix of the stag hunt.

Because the stag hunt is the game that Skyrms treats most fully, I start with it. This two-person noncooperative game has two Nash equilibria in pure strategies. One is efficient, and the other is not efficient. The stag hunt has a rewarding but risky equilibrium, namely, stag hunting. It also has a risk-free equilibrium, namely, hare hunting. Various payoff matrices realize the stag hunt. One matrix is this.

	Cooperate	Defect
Cooperate	4, 4	0, 3
Defect	3, 0	1, 1

With the standard replicator dynamics, the basin of attraction for cooperation drives a population with at least 75% cooperators to cooperation (p. 11). Chance may bounce a population from one equilibrium to another. Without chance, a population of hare hunters is stuck with hare hunting despite its inefficiency.

In Chapter 3, Skyrms treats the stag hunt in a population organized into neighborhoods of interaction. In that analysis, the stag hunt has this payoff matrix.

	Cooperate	Defect
Cooperate	3, 3	0, 2
Defect	2, 0	2, 2

Defecting, or hunting hare, is a risk-dominant equilibrium in the sense that given equi-probability of strategies, defecting maximizes expected utility. Models for this game differ in their dynamics. Glenn Ellison's (1993) dynamics use a best-response updating rule, and Jason Alexander's (2000) dynamics use an imitate-the-best updating rule. Also, models differ in type of interaction with neighbors. Interaction may occur in a one- or two-dimensional neighborhood, for example. Interaction with neighbors speeds attainment of an equilibrium, but whether the efficient equilibrium evolves depends on details of the interaction. Imitating the most successful neighbor in a two-dimensional neighborhood tends to yield the efficient equilibrium, whereas adopting a best response to an anticipated act of a neighbor in a one-dimensional neighborhood tends to yield the inefficient equilibrium (pp. 31–32). Skyrms summarizes results this way: "Local interaction opens up possibilities of cooperation that do not exist in a more traditional setting, and imitation dynamics is often more conducive to cooperation than best-response dynamics" (p. 41).

The equilibrium's sensitivity to the details of interaction raises questions about the explanatory power of the simulations. Such details do not seem to affect the emergence of cooperation among people. People engaged in the stag hunt do not reveal a similar sensitivity to types of neighborhood. The model's sensitivity to those types shows that its results rest on assumptions about nonexplanatory factors. The model therefore does not yield a partial explanation of cooperation.

In Chapter 2, Skyrms studies the evolution of the norm of equity in a two-agent division of a windfall. The game studied is called divide-the-dollar. It is not a bargaining problem because the agents do not communicate. The game is a simultaneous-move, noncooperative game, in particular, a coordination problem. Each agent simultaneously submits a bid and then receives a portion of the dollar available only if the sum of their bids is less than or equal to a dollar. Each obtains his demand in that case. To make the game tractable, Skyrms limits an agent's options to demand 1/3, demand 1/2, and demand 2/3. Then he studies robustness of results with respect to granularity among options.

Suppose that the members of a population interact in random fashion. Then demand 1/2 is the only evolutionarily stable strategy. If the whole population follows it, mutant strategies cannot take over. However, half the population's demanding 1/3 and the other half's demanding 2/3 makes a stable polymorphism (p. 20). Next, suppose that strategies' probabilities increase in proportion to their success and vary a bit by chance. Then fair-division is the uniquely stochastically stable equilibrium (pp. 22–23). It is the highest probability evolutionarily stable strategy, and a population spends most of its time with it despite the possibility of a stable polymorphism.

Skyrms's account of the emergence of fair division in divide-the-dollar summarizes results of Alexander and Skryms (1999) and Alexander (2000). Those articles descend from Skyrms (1996) and respond to criticisms from D'Arms, Batterman, and Górny (1998).

Skyrms's (1996) model for the emergence of fair division in divide-the-dollar introduces correlation of strategies of agents playing the game. D'Arms et al. object to assuming correlation without specifying a mechanism to generate it. An evolutionary account of behavior may fail, they observe, because it abstracts away crucial details of the particular ecologies and ontogenies of the creatures whose behavior it seeks to explain (p. 82). They ask for a particularist account of the evolution of behavior that identifies the mechanism that produces the behavior to be explained. "Selective pressures deriving from the fitness consequences of various social relations such as cooperation, reciprocity, coalition building, and competition for social status, have forged similarly specific adaptive psychological mechanisms which mediate cognition and motivation in these domains" (p. 84). They are dissatisfied with Skryms's account of the emergence of fair division because it is a generalist account. It just identifies gains from correlation of strategies without explaining how organisms achieve that correlation. An explanation of some behavior should identify a mechanism that produces the behavior and not just appeal to the advantages of the behavior. They ask for the mechanism that generates correlation of strategies in divide-the-dollar. They find no evidence of a genetic mechanism that yields the correlation. Moreover, assuming a genetic mechanism sacrifices flexibility. It precludes a cultural mechanism that uses teaching or competition for the allegiance of rational deliberators.

Skyrms shows that fair division emerges in a variety of ways in a variety of circumstances. But D'Arms et al. hold that robustness is not enough to explain fair division. To flesh out an evolutionary model of behavior, they assume rational deliberators, substitute a best-response dynamics for Skyrms's imitation dynamics, and allow agents to seek out partners for interaction. Their model does not support correlation of strategies. In fact, it yields anti-correlation of strategies because those who demand 2/3 seek out those who demand 1/3. With anti-correlation, the demand 1/3 and demand 2/3 polymorphism reemerges. They conclude that Skyrms's model is not sufficiently flexible and robust to explain fair division because adding some realistic features brings back the inequitable polymorphism (p. 98).

Alexander (2000) defends Skyrms's (1996) appeal to correlation of strategies. To Skrym's model, he adds a mechanism of correlation, namely, interaction in a fixed neighborhood. The addition of neighborhoods not only correlates strategies but also makes fair division relatively independent of a population's initial mix of strategies. Almost every initial mix is in the basin of attraction for fair division. Justice is contagious given neighborhoods.

Skryms (2004: Chap. 2) adopts the spatial model of interaction to improve his original model for fair division. In the new model, correlation of strategies arises from imitation of successful neighbors rather than from genetic influence. For instance, suppose that agents adopt the most successful strategy in their neighborhood. Then fair division almost always emerges and emerges quickly (p. 24). An agent's bargaining with random members of a population yields fair division as an equilibrium after, on average, 46 generations. However, an agent's bargaining with neighbors yields the same equilibrium after, on average, only 16 generations. Rapid fixation of fair division occurs because of neighborhoods and not because of the type of imitation. If instead of imitating the best strategy, an agent imitates strategies with probabilities in proportion to their success rates, fair division still emerges rapidly (p. 26). A dynamics with neighborhoods yields the target phenomenon, namely, fair division plus the speed of its emergence, better than a dynamics with random interaction does.

Despite improvements, Skyrms's model still does not offer a partial explanation of fair division. The model is not robust with respect to nonexplanatory factors. The model's results depend on its rule for updating behavior. In particular, the results depend on the updating's being synchronic, as Alexander (2000: 507) notes. This sensitivity to timing is nonexplanatory because among people the timing of interaction does not make a difference. Moreover, results about the speed of fair division's emergence depend on the size of neighborhoods, as Alexander (pp. 507–10) shows. The size of neighborhoods of interaction is not an explanatory factor, and variations concerning it should not matter to cooperation's emergence. People playing divide-the-dollar adopt a 50-50 split immediately without any apparent sensitivity to the size of neighborhoods of interaction.

To succeed as a partial explanation, a model's assumptions must control for explanatory factors only. Some features of Skyrms's models of coordination and cooperation are not explanatory. Despite other successes, those models do not yield partial explanations of the emergence of fair division and stag hunting. The next section considers how to remedy this problem.

5. Incorporating Strategic Reasoning

Skyrms's simulations reach solutions to games using limited resources. In particular, they do not use strategic reasoning. Although the simulations' power is remarkable, realism about human behavior requires adding strategic reasoning. Explanations of human behavior should include strategic reasoning because people use it.

A model may adjust to the type of population it treats. For people, it may incorporate reasoning that governs their interactions. An agent's response to others need not be simple imitation but may involve strategic reasoning. A model's dynamics may allow for that reasoning. This section suggests adding strategic reasoning to the models Skyrms's simulations run, at least, when they target human behavior in games of strategy.

Chapter 7 of Skryms's book adds some strategic reasoning to the models earlier chapters present. He says, "Let us equip our agents with a little bit of strategic thinking. Now, when an agent revises her strategy, she chooses a best response to the strategy she has last encountered. … Best-response to the last strategy encountered is related to *best-response to some more or less sophisticated inductive estimate of what strategy you are likely to encounter*" (p. 107). The type of best-response dynamics Skyrms introduces is a very simplified form of strategic reasoning. Why not add more sophisticated reasoning? Perhaps reluctance arises because people have only bounded rationality. Building in sophisticated reasoning may not yield an explanation of behavior. Also, decision-theoretic support for solutions to games is a research project. Standard principles of rationality such as utility maximization do not explain realization of a Nash equilibrium.

The replicator dynamics generates a Nash equilibrium more straightforwardly than strategic reasoning does. Models may resolve these problems concerning bounded rationality and the theory of rationality by introducing realistic forms of strategic reasoning and by investigating forms of strategic reasoning that lead to appropriate types of equilibrium.

A dynamics for strategic reasoning should include an agent's estimation of other agents' strategies, and that estimation should acknowledge their strategies' dependence on their estimation of his strategy. The interactive nature of strategic reasoning generates equilibria, and some equilibria are more efficient than others are. The dynamics should add the global influence of strategic reasoning to the local influence of agents on each other. Alexander (2000: 502–03) describes a sort of global influence that may help explain the emergence of fair division. He observes, "If demand half is a stochastically robust strategy and *A* knows this, *A* knows that the strategy of fair division will shortly dominate the population. This provides a strong incentive for *A* to demand half, since, given *A*'s expectation that fair division will shortly dominate the population, demand half is the strategy which will maximize *A*'s amount of cake when fair division comes to dominate." While respecting bounded rationality, a model may incorporate such global influences on behavior.

How should models and simulations incorporate strategic reasoning? If one builds a model in which agents above all else want to cooperate, then cooperation emerges easily. For a deeper-going explanation, one may add features that explain why people want to cooperate. The target phenomenon may include more than cooperation. It may include the desire to cooperate. The general goal is to explain actions by individuals that yield cooperation in a group. Additions to the dynamics are not ad hoc as long as they concern individuals rather than the whole group.

Realism requires giving agents no more reasoning power than they have, but a model's explanatory goals may not require realism in all matters. A model with simplifying assumptions may yield a partial explanation if it replicates some factors explaining human behavior. A model incorporating reasoning may use abstraction and idealization to simplify its treatment of reasoning. More realistic models may later rescind those abstractions and idealizations.

Incorporating strategic reasoning may alter the target phenomenon. Models that incorporate reasoning may seek explanations of rational behavior. Such explanations amount to justifications of behavior. Idealized forms of strategic reasoning fit this project. A model may explain how rational behavior emerges from strategic reasoning. A model with a normative objective may be used to improve human behavior. If the model is appropriate for people with bounded rationality, then it may also explain behavior because people are rational by and large.

Moving to an explanation of rational behavior has some advantages. It enhances the prospect of obtaining a partial explanation. Because it is hard to treat empirical systems with precision, it is hard to obtain a partial explanation of human behavior. A treatment of rational behavior, being normative, may attain the type of precision that logic attains. One may evaluate a model of rational behavior as one evaluates a theorem. The model may treat reasons with precision. A priori judgments may confirm that the model's assumptions control for explanatory factors only and that the model accurately handles remaining explanatory factors. The shift from descriptive to normative models thus facilitates confirmation of explanatory success.

This section proposes using strategic reasoning to promote explanation of rational behavior. One explanatory objective is an individualistic account of the rationality of participation in a Nash equilibrium. Skyrms uses individualistic factors to obtain a Nash equilibrium, but his assumptions are not explanatory, as Section 3 argued. This section considers how to modify those assumptions so that they better serve explanatory purposes. Another explanatory objective is an individualistic account of the rationality of coordination and cooperation in games such as divide-the-dollar and the stag hunt. Skryms's simulations use individualistic factors to obtain fair division and to select an efficient equilibrium. His assumptions are not explanatory, however, as Section 4 argued.

First, consider adding strategic reasoning to Skyrms's deliberational dynamics. Those dynamics attempt only to explain realization of a Nash equilibrium and do not attempt to explain realization of a particular Nash equilibrium. In games with multiple equilibria, the equilibrium realized depends on the agents' initial probability assignments to the strategies of other agents. Models that incorporate strategic reasoning may explain the realization of a particular Nash equilibrium, for example, an efficient Nash equilibrium. Adding strategic reasoning gives agents foresight. They can look ahead to see where their initial probability assignments will take them. Then they can reason their way to probability assignments that will take them to an efficient Nash equilibrium. Adding strategic reasoning eliminates an equilibrium's dependence on the agents' initial position in the deliberational dynamics. They need not fall by chance into the equilibrium's basin of attraction. A simulation incorporating strategic reasoning is robust with respect to the nonexplanatory factor of initial position.

Strategic reasoning also adds robustness with respect to the timing of the agents' deliberations. Adding strategic reasoning swamps out the affects of timing. Agents assess the direction of their current deliberations and make adjustments so that their deliberations end well. The agents may independently make adjustments at various stages of their deliberations. They may identify the equilibria of their game, and an efficient equilibrium may be a focal point of their reasoning.

Next, consider cultural evolution in the game divide-the-dollar and in the stag hunt. The key to fair division and stag hunting is correlation of strategies. Skyrms's mechanism for achieving it is a neighborhood of interaction. His dynamics use imitation of successful strategies in the neighborhood. Adding strategic reasoning alters the dynamics. Justification of cooperation requires a best-response dynamics rather than an imitation dynamics. Imitation of success sounds rational, but those succeeding may be cooperators in a prisoner's dilemma. Then imitation of their success yields an irrational strategy. Adding strategic reasoning also introduces interaction of strategic considerations. Strategic reasoners may use focal points to coordinate in divide-the-dollar and to reach the efficient equilibrium in the stag hunt. As a result, cooperation becomes robust with respect to types of neighborhoods. Given that agents see the advantages of stag hunting, it does not matter whether they are interacting in one-dimensional or two-dimensional neighborhoods. It is as if the neighborhoods of learning and interaction were distinguished, and as if the neighborhood of learning were the whole population.

In Skyrms's model of cooperation, the equilibrium a simulation achieves in a game with multiple equilibria depends on the population's starting point and on the chance events that make a population jump from one basin of attraction to another basin of attraction. Strategic reasoning replaces initial conditions and chance as a means of settling the equilibrium reached. Strategic reasoners are not dependent on the population's starting point or chance perturbations.

A simulation needs features handling initial conditions, type of interaction, and dynamics. Given strategic reasoning, results are robust with respect to changes in those features. Because agents look for an efficient equilibrium and take steps to attain it, the details of their starting position and local interaction are not important. A simulation gains explanatory power if its results are robust with respect to changes in these nonexplanatory factors.

Furthermore, in Skyrms's model, because basins of attraction govern a population's path toward an equilibrium, and because chance moves a population from one basin to another, the emergence of a desirable equilibrium may be a relatively slow process. It may take a large number of mutants to shift a population from one basin to another, and, normally, it takes a long time for a large number of mutants to occur together. Strategic reasoning works faster than chance does. Adding strategic reasoning to simulations of human behavior increases the speed of cooperation's emergence. A dynamics with strategic reasoning accelerates emergence of cooperation in both divide-the-dollar and the stag hunt. A simulation with strategic reasoning yields fair division rapidly in divide-the-dollar because fair division is a focal point. Adding strategic reasoning increases the speed of reaching the efficient equilibrium in the stag hunt because stag hunting is also a focal point. Increasing the speed of cooperation's emergence has explanatory value because the target phenomenon includes the pace of cooperation's emergence, which is rapid among people.

My account of the role of strategic reasoning appeals to focal points. A simulation may aim to explain focal points rather than lean directly on them. Michael Bacharach (2006) explains focal points using the psychological framing of decision problems. Simulations may incorporate his methods. Then the relevant period for cooperation's emergence includes the period for emergence of the pertinent focal point. Strategic reasoning may accelerate emergence of a focal point, too. Fair division is a sustainable practice, and stag hunting is an efficient equilibrium. These considerations draw attention to those equilibria and speed their emergence as focal points.

Finally, I consider a game where the prospects for cooperation are especially bleak, namely, the prisoner's dilemma. Skryms (2004) treats this game, too. He notes that some spatial models with correlation of strategies, such as the haystack model, may support cooperation in the prisoner's dilemma (pp. 6–8). Nonetheless, only defection is rational, so a best-response dynamics yields defection (p. 41). Doesn't adding strategic reasoning to a model diminish the prospects for cooperation in the prisoner's dilemma?

One way to improve prospects for cooperation is to move from the one-shot prisoner's dilemma to repeated prisoner's dilemmas. Repetition, of course, changes the game because choices in one game may affect outcomes in later games. Repeated prisoner's dilemmas create a stag hunt, as Skryms observes (pp. 4–5), and cooperation in the stag hunt is achievable.

Adding strategic reasoning also creates a new, related mechanism for escaping the prisoner's dilemma. Strategic reasoners may change the form of their interaction. That is, they may change the game they are playing. They may alter their social institutions.

The prisoner's dilemma involves agents who cannot communicate. They may escape from the dilemma if they can communicate. As John Harsanyi and Reinhard Selten (1988: 4) observe, if agents can communicate, then one agent may offer to cooperate on condition that the other agent does. The other agent may then cooperate. The conditional offer and positive response constitute a roll back equilibrium of a sequential game underlying the prisoner's dilemma. Participating in the rollback equilibrium is rational for each player. People may foresee the results of changing a prisoner's dilemma to allow for conditional offers and responses. Liking the results, they may make the change.

Strategic reasoning not only applies within a game but also guides selection of games to play. Strategic reasoners may anticipate the results of the game they are playing and may alter the game to achieve better results. Humans may change the context of their interactions by changing their social institutions. They may replace a social institution with a new institution in which people pursuing incentives achieve an equilibrium better than one they achieve by pursuing incentives in the old institution. Human control of social structure is a means of reaching cooperation faster than the replicator dynamics reaches it, even when interaction with neighbors enriches that dynamics. To explain human cooperation, especially the speed of its emergence, a simulation should incorporate the reasoning of agents and its influence on social structure.

Models and simulations of cooperative human behavior gain explanatory power by incorporating strategic reasoning. Strategic reasoning introduces a type of global dynamics that improves robustness. Although humans have only bounded rationality, their strategic reasoning moves them rapidly toward efficient equilibria and toward beneficial changes in social structure. Skyrms's models and simulations are not completely successful in providing partial explanations of human cooperation, but adding strategic reasoning remedies that problem.

References

Alexander, Jason. 2000. "Evolutionary Explanations of Distributive Justice." *Philosophy of Science* 67: 490–516.

Alexander, Jason, and Brian Skyrms. 1999. "Bargaining with Neighbors: Is Justice Contagious?" *Journal of Philosophy* 96: 588–98.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Bacharach, Michael. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton: Princeton University Press.

D'Arms, Justin, Robert Batterman, and Krzyzstof Górny. 1998. "Game Theoretic Explanations and the Evolution of Justice." *Philosophy of Science* 65: 76–102.

Ellison, Glenn. 1993. "Learning, Local Interaction and Coordination." *Econometrica* 61: 1047–71.

Giere, Ronald. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Hartmann, Stephan. 1996. "The World as a Process: Simulations in the Natural and Social Sciences." In Rainer Hegselmann, Ulrich Mueller, and Klaus Troitzsch, eds., *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, pp. 77–100. Dordrecht: Kluwer.

Harsanyi, John, and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Hegselmann, Rainer, Ulrich Mueller, and Klaus Troitzsch. 1996. *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Dordrecht: Kluwer.

Humphreys, Paul. 2004. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press.

Kagel, John and Alvin Roth. 1995. "Introduction to Experimental Economics." In John Kagel and Alvin Roth, eds., *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Savage, Leonard. 1954. *The Foundations of Statistics*. New York: Wiley.

Skyrms, Brian. 1990. *The Dynamics of Rational Deliberation*. Cambridge, MA: Harvard University Press.

Skyrms, Brian. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Clarendon Press.

Weirich, Paul. 2004. *Realistic Decision Theory*. New York: Oxford University Press.