SCIENCE WITHOUT LAWS

MODEL SYSTEMS, CASES, EXEMPLARY NARRATIVES

Edited by Angela N. H. Creager, Elizabeth Lunbeck, and M. Norton Wise

Duke University Press Durham and London 2007

The Curious Case of the Prisoner's Dilemma: Model Situation? Exemplary Narrative?

MARY S. MORGAN

DILEMMA I: THE HISTORIANS' DILEMMA: GAME OR WAR?

The Prisoner's Dilemma game is one of the classic games discussed in game theory, the study of strategic decision making in situations of conflict, which stretches between mathematics and the social sciences. Game theory was primarily developed during the late 1940s and into the 1960s at a number of research sites funded by various arms of the U.S. military establishment as part of their Cold War research. As an introduction to this article on game thinking in economics, let me present the historian's dilemma in dealing with this background in its starkest form with two quotations. The first comes from Jacob Bronowski in 1954:

> The scale of the damage of Nagasaki drained the blood from my heart then [in the autumn of 1945], and does so now when I speak of it. For three miles my road lay through a desert which man had made in a second. Now, nine years later, the hydrogen bomb is ready to dwarf this scale, and to turn each mile of destruction into ten miles, and citizens and scientists stare at one another and ask: "How did we blunder into this nightmare?"[1]

Writing in 1966, Anatol Rapoport provides something of an answer to how Bronowski's 1954 nightmare came about. At least part of the blame lay at the door of game theorizing. Rapoport, a highly respected figure in game theory, was keen to point out the extreme limitations on applications of formal game theory to strategic thinking and to warn of the dangers of placing reliance on such theory in Cold War actions. In making his warning, he suggests why game theory became so important to Cold War ways of thinking. I quote his words at length, for his arguments about the role of game theory in Cold War thinking prompt some of the claims I will make later about game theory and economic reasoning:

> If two-person game theory is an extension of rational decision theory to situations in which outcomes are controlled by two decision-makers whose interests are at least partially in conflict, then the range of appli-

cations of two-person game theory ought to be the range of such situations. It is understandable why, in the period following the close of World War II, when so much attention was paid especially in the United States, to the impending power struggle between the Communist and the non-Communist worlds, the appearance of game theory on the scientific horizon was hailed with enthusiasm and with great expectations.

People had witnessed the increasing abstruseness of the sciences geared to military applications. World War I had been called the chemists' war. World War II was called the physicists' war. Towards its final phases, World War II was rapidly becoming a mathematicians' war with cybernetic devices and electronic computers beginning to play a decisive role. It is assumed in many quarters that World War III (which many feel to be a matter-of-fact culmination of existing trends) will be truly a mathematicians' war.

Moreover, mathematics is assumed in those quarters to be not merely an appendage to physical science but also the foundation of strategy. . . .

Wars to come are imagined by the strategists to be either "limited wars" or "nuclear exchanges,"[2] both being envisaged as wars of strategy rather than of attrition. It seems that those strategists who are actively concerned with the conduct of limited war view such wars as "rational" instruments of national policy, in contrast to nuclear war which, because of its awesome destructiveness, falls outside the scope of rational policy. Those strategists, on the other hand, who are concerned with nuclear "exchanges," although noncommittal about the "rationality" of such maneuvers [*sic*], view the *potential* for waging nuclear war as bargaining leverage in international affairs. They view the use of this potential as a basis for rational diplo-military policy. . . .

One finds in the writings of contemporary strategists a deliberate striving to rehabilitate war as a normal event among civilized nations. The reestablishment of high intellectual content in military strategy doubtless serves this purpose. In my opinion, the tremendous interest aroused by game theory is in no small measure due to the climate in which the rehabilitation of war, or at least of the sophisticated power struggle, was undertaken.

It becomes, therefore, extremely tempting to those actively involved in game theory and also interested in its application potential to reply in the affirmative to the question "Is game theory useful?" Since rationality in conflict enjoys extremely high prestige in our day when "realism" and

> "tough-mindedness" are extolled as evidence of sophistication and maturity, game theory can indeed be sold as a useful science.[3]

At this point Rapoport goes on to outline the enormous difficulties and strong limitations that arise when one tries to apply game theory beyond tic-tac-toe and into the field of international conflict.[4]

In his argument above, Rapoport links the development of game theory with a change of thinking in which war is "rehabilitated as a normal event among civilized nations." The link comes via the authority of mathematics and the ascription of a scientifically based rationality. But while Rapoport suggests that game theory was taken up because of the "civilization" of war, it seems equally part of the process that war became acceptable because it was reinterpreted in game theory terms. Whereas game theory initially provided a mathematically formulated theory of rational action in certain situations that might be *applied* to the Cold War world, the process of using game theory to think about that world turned the relationship around: the Cold War came to be seen as a set of game situations. Whereas the Prisoner's Dilemma begins as a game that might be applied, by analogy, to the nuclear arms race, it gradually comes to the point at which we understand and interpret that race as a Prisoner's Dilemma game. No doubt we are familiar enough with this idea that the first task for the historian is to recreate how strange this is.

William Poundstone's book *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb* covers the ground between Bronowski and Rapoport above. It opens with the following story, apparently a traditional African dilemma tale: "A man was crossing a river with his wife and mother. A giraffe appeared on the opposite bank. The man drew his gun [traditional?] on the beast and the giraffe said, 'If you shoot, your mother will die. If you don't shoot, your wife will die.' What should the man do?"[5] Poundstone, sensing that talking giraffes cut little ice in 1990s America, updates the tale—but only as far as the science fiction regime appropriate perhaps for 1950s or 1960s Cold War America—by replacing the talking giraffe with some mad scientists (surely some game theorists from RAND?):

> You are supposed to imagine that the pronouncements of talking giraffes are always true. You can restate the dilemma in more Western and technological terms: you, your spouse, and your mother are kidnapped by mad scientists and placed in a room with a strange machine. All three of you are bound immobile to chairs. In front of you is a push button within reach. A machine gun looms in front of your spouse and mother,

> and a menacing clock ticks away on the wall. One of the scientists announces that if you push the button the mechanism will aim the gun at your mother and shoot her dead. If you *don't* push it within sixty seconds it will aim and fire at your spouse.[6]

These two tales produce a feeling of incredulity in me: whereas the ethical dilemma is surely serious, the kind of situation outlined is so bizarre that it creates a dissonance between problem and situation. Poundstone began with the "dilemma" aspect of his project because it enabled him to trace the parallels between the development of game theory and the nuclear arms race. In that context, the Prisoner's Dilemma game, understood as a model of the Cold War conflict, represented the dilemma of whether to cooperate to prevent a third world war or to strike first with an H-bomb, accepting the danger that both superpowers would destroy each other. The idea that our safety might have hung on the correct application of a Prisoner's Dilemma game is surely something we need to find incredible: we should feel the profound distance between the seriousness of the real problem and the analytical tool, a tool that used an extremely narrow idea of rational behavior and fitted the world situation into a 2 x 2 matrix.[7] And, of course, Poundstone's opening is designed precisely to make us feel the problem, to induce incredulity into the reader by posing the horrible dilemma within an extremely oddly constructed, almost surreal, situation and narrative.

Poundstone's treatment adds something important to Rapoport's observations. It is not just that war could be written about, analyzed, and discussed in mathematical (high intellectual) terms but that the realm of mathematical theory treated war and international diplomacy as a series of game situations, to be played out according to the kind of game they most resembled. Game theorizing of the day relied not just on mathematics but also on mathematicians and social scientists playing games. In this regard, game theory was associated not so much with the civilization of war as the reduction of war studies to the playroom.[8]

There is not nearly so much of a dilemma for the historian of economics in writing about the role of game theory as for the historian of international affairs. But a certain level of dissonance needs to be present even here. Otherwise we have no power to ask how it is that economists can use the Prisoner's Dilemma (PD) game to reason about quite complicated economic behaviors and actions. What implications did this strange little story and its matrix of numbers have for economics?

The history of game theory in economics has a number of features in com-

mon with the discussions above. The authority of mathematics, the ascription of rationality, and the slippage between applying game thinking to economic situations and seeing those situations as games all constitute features of the spread of game thinking in economics. At first sight, the introduction of game theorizing into economics seems pretty innocuous, for the assumptions of formal game theory are closely aligned with those of neoclassical economics. But as it turns out, game theory is associated with two deep-seated changes in economics over the past half century.

One change associated with the PD is rather well understood, if not well documented. The PD game is a classic game just because it presents a dilemma economists could not resolve within the terms of their existing theory. This little two-person dilemma game turned out to have the power to undermine some cherished beliefs of neoclassical economists, and it has surreptitiously eaten away at economists' belief in the benevolence of the invisible hand and the rationality of the individual. The other change is less remarked and understood, for game thinking has introduced a form of reasoning into economics that adds case-based reasoning onto more generalizing kinds of theorizing. In this respect also economists have come to live with the PD game, and they use it very often, but not perhaps in a way they can easily characterize.

Please expect no general history of game theory in economics here.[9] Rather, I use the case of the Prisoner's Dilemma game to discuss these two changes in ways that I hope will highlight the curiosities of what has happened.

DILEMMA 2: THE PRISONER'S DILEMMA: COLLABORATE OR DEFECT?

The Prisoner's Dilemma game is so named because it involves a dilemma for each of the individuals playing the game. It is one of a number of simple "games" studied in the social sciences embodying not just situations of conflict in which choices must be made but situations in which the players face an element of dilemma in choosing what actions to take. Unlike many games in game theory, the PD game actually began life as a game. It was possibly first played at RAND during January 1950 following a design by Merrill Flood and Melvin Dresher.[10] In the game, each of two players (John Williams from RAND and Armen Alchian from the University of California, Los Angeles) had to choose—simultaneously and in the absence of knowledge of the other's choice—one of two actions, for which they received payoffs known to both players in advance. Howard Raiffa, working at the University of Michigan on contracts for the Office of Naval Research, was at the same time investigating a similar game and

carrying out game-playing experiments with it.[11] At this stage the game had no name.

The rewards from the choices made by the two players in such a game are typically given in a so-called matrix of payoffs. By convention these are treated as utilities, and under certain usual assumptions, these can be treated as monetary payoffs. Here I provide a version of the matrix of numbers given by Duncan Luce and Raiffa from their respected early text that remains a reference source in the field:[12]

	Player B, Collaborate		Player B, Defect	
Player A, Collaborate	5	5	−4	6
Player A, Defect	6	−4	−3	−3

Here B, the column player, has outcome payoffs on the right, while A, the row player, has outcome payoffs on the left, from whatever joint choices they make. For the convenience of later discussion, I have called row 1–column 1 choices the CC ("collaborate with each other") choice and row 2–column 2 choices the DD ("defect against the other") choice.

Here I have presented the game in the form in which it most usually appears in economics writings, namely, as a particular matrix of payoffs. Many variations of the numbers may be used, and while the matrix is usually symmetric, it need not be. (For example, in that first recorded game at RAND, the matrix was not symmetric, to make it less clear what the situation was.) But this variation in numbers is misleading for the numbers used cannot be just any numbers. Even slight changes in the numbers may change the matrix to represent a different game, maybe even another dilemma game.[13]

The numbers in a PD matrix are particularly chosen, but they also follow some general rule, namely, that relations between the numbers must conform to a set of inequalities. These are important in defining a PD game matrix, for they provide a more general description of the payoffs, though, surprisingly, they are rarely given.[14] Where they are given, they appear in the following form:[15]

a) $T > R > P > S$, and
b) $2R > (T + S) > 2P$

Of course, these inequalities are defined in terms that make no sense yet, and they make no sense because so far I have not reported the text that goes with the game. This in itself offers a good pointer to the importance of the text in defining the structure of the game, and whereas the game, strangely, very often

appears in economics books without the inequalities, it (almost) never appears without the text.

The text attached to the matrix is the story of two prisoners and their dilemmas. Though there is some debate about who first worked on the game matrix, all parties agree that the name and story were attached to the game by Albert Tucker, a Princeton mathematician, when he wanted to use the game in a popular lecture to psychologists.[16] The PD text, which I again quote from Luce and Raiffa's first account of the game—used by the *Oxford English Dictionary* (OED) to indicate the first written usage of the term *prisoner's dilemma*— goes as follows:[17]

> The following interpretation, known as the prisoner's dilemma, is popular: Two suspects are taken into custody and separated. The district attorney is certain that they are guilty of a specific crime, but he does not have adequate evidence to convict them at a trial. He points out to each prisoner that each has two alternatives: to confess to the crime the police are sure they have done, or not to confess. If they both do not confess, then the district attorney states he will book them on some very minor trumped-up charge such as petty larceny and illegal possession of a weapon, and they will both receive minor punishment; if they both confess they will be prosecuted, but he will recommend less than the most severe sentence; but if one confesses and the other does not, then the confessor will receive lenient treatment for turning state's evidence whereas the latter will get "the book" slapped at him.[18]

Each prisoner faces a strategic choice, but this choice poses a dilemma: should the prisoner choose to cooperate with his fellow prisoner and not confess to the police, he may end up with the rewards of a small prison term, but if his fellow prisoner does the opposite, it makes his own situation the worst it could be; or should he follow his own self-interest and confess hoping his fellow will not, thus enforcing the best outcome for himself and the worst outcome for his fellow, but with the danger that his fellow will also confess, thus leaving them both worse off. That is, the dilemma for both prisoners concerns the choice of whether to trust a fellow prisoner or not, and so whether to chance the outcome of being a Sucker (not confessing when his fellow does) or giving into the Temptation of telling tales on the fellow. Of course, each one may hope that they can both reap the Reward of cooperating with each other, but there is always the possibility that both will pay the Penalty of defection. The dilemma of whether to "collaborate" and reap the joint rewards or "defect" to one's own advantage and to the loss of one's opponent was experienced by the two players

involved in that first recorded playing of the game, as we can learn from the transcripts of their ongoing personal commentary as the game was played over and over again one hundred times.[19]

The economists' analysis of this dilemma is clear cut. The economist automatically assumes that each prisoner is a rational economic actor and will seek to maximize his or her individual utility (their payoff). That is, for A (row player), it is better to play row D regardless of what B (column player) does [$6 > 5$ and $-3 > -4$]; and for B (column player) it is better to play column D [$6 > 5$ and $-3 > -4$] as well. Both prisoners will act rationally by defecting (confess to the police) and both end up with penalties of -3. Although both prisoners might do better if they could agree not to confess, the text embodies the rules of the game, which forbid any discussion between them. A more modern treatment of this text would argue that even if they could agree to collaborate, the agreement is not credible or enforceable, for the payoff structure in this matrix still provides an individual temptation to defect. Thus the outcome of CC is not an "equilibrium" solution, it is not a stable outcome, whereas the DD outcome is.

This analysis reveals the meaning of the constraints on the matrix numbers, and we can now see where the inequalities terminology comes from: T is the reward from giving into temptation and defecting when his fellow does not, R is the reward to each from cooperation with each other, P is the penalty to each when both defect, and S is the loss from being a "sucker," the player who cooperates when his fellow player defects. Our previous matrix can be reinterpreted with the inequality symbols:

	Player B, Collaborate		Player B, Defect	
Player A, Collaborate	R (5)	R (5)	S (−4)	T (6)
Player A, Defect	T (6)	S (−4)	P (−3)	P (−3)

The economists' solution to this dilemma is, at one level, straightforward: rational economic man, the model man who inhabits economic theories and makes them what they are, follows an individual utility-maximizing process which, in this case, translates into making choices that end up with both players paying a penalty rather than reaping rewards. This makes economists uncomfortable, as we can see from Luce and Raiffa's comments upon the solution to the PD game: "No, there appears to be no way around this dilemma. We do not believe that there is anything irrational or perverse about the choice of α_2 and β_2 [of D, D], and we must admit that if we were actually in this position we would make

these choices."[20] There was nothing wrong with the individual choices in the context of the PD game, but clearly the outcome is "wrong" in some sense—this marks the dilemma economists face.

DILEMMA 3: THE ECONOMISTS' DILEMMA: INDIVIDUAL RATIONALITY OR INVISIBLE HAND?

Economic theory has long assumed a "rational economic man" to lie at the heart of economic reasoning. Adam Smith's economic man was a complex mixture of propensities, preferences, talents, and motives (including self-interest). A more narrowly defined *Homo economicus* only became established for economists by John Stuart Mill in 1836.[21] Mill argued that only by adopting a thin psychological profile of individuals could economics make any progress as a science. This thinly profiled or abstracted account of economic behavior, picturing man as governed overwhelmingly by wealth-seeking self-interest, was adopted within the later classical school (including Marx), and changed little in the following neoclassical and Austrian traditions. The "rationality" tag was a twentieth-century addition, where rationality was defined as utility-maximizing behavior and being consistent in choices. From Mill onward, the abstraction of rational economic man was thought to capture the essential elements of economic behavior, though it was known to serve inadequately as a description of actual behavior, which was subject to many other impulses, economic and otherwise.[22]

During the period from the 1940s through the 1960s, the period of early development for game theory, the role of rational economic man gained greater significance for he constituted a necessary component of the nexus of general equilibrium analysis and the perfectly competitive economy that formed the focus of mathematical theorizing during those years. Economists' emphasis on such narrowly defined, hard-edged economic rationality seemed during this period to turn their favorite creature into a caricature, nowhere more evident than in game theory.

This degree of rational behavior was seen as a particular problem in the PD game. The primary expression of the prisoner's dilemma is that in theoretical analysis, the outcome of individually rational and strategic behavior in the PD game leads to an outcome that is jointly irrational. That is, by following the economists' injunction to maximize their individual gain, both prisoners end up with a worse outcome than if they had both collaborated.

But perhaps, economists argued, the result might not hold if two people played a succession of such games. It is tempting, after all, to suppose that a succession of games would ensure the cooperative outcome. Even here, the

theoretical analysis led to the same result: the usual method of analyzing the outcome of a PD game repeated a finite number of times is one of so-called backward induction.[23] It calls for first working out what will happen on the last game: because it is the last game, there is no reason for the individual to cooperate and every reason to defect, hoping that the opponent will not and thus reaping the best outcome, as in the single-game case. Since both individuals follow the same rationality, the last game will result in a DD outcome. Moving back one game to the penultimate game produces the same analysis. The sequence continues back to the beginning, so the theoretical analysis results in a series of bad outcomes.[24] Rational economic man has no learning power, no power to trust; he can only choose strategically the best option in any given situation.

The theoretical results were incontrovertible, leaving the economist extremely uneasy—either the rationality assumption was wrong, or the outcome equilibrium was wrong. But neither could easily be given up: too much was at stake.

On the one hand, the rational economic man was deeply embedded in the high theory of the day. Because of the position of this rational character as a central building block of modern theory, softening or broadening the rational-man portrait would undermine more than just game theory or the PD result.

On the other hand, economists rely equally on another important result in economic theory, namely, the invisible-hand outcome that individuals following their own self-interest will end up doing naturally what is best for each other. According to the argument of the invisible hand, the outcome of a game in which the players follow their own self-interest should be a good one. This invisible-hand argument forms the basis of the idea that free markets and free individuals will deliver a more efficient and better outcome than that arising from government planning.

Thus, the *theory* of what happens in the PD game leaves economists in a particular nasty double dilemma. Constraining the rationality of individual play, means accepting the irrational outcome that both players are worse off than if they collaborated. But accepting that irrational outcome, means denying the benevolence of the invisible hand (and thus the efficiency of the market), while weakening the rationality of individual self-interest, means potentially undermining the mathematical work on general equilibrium analysis and perfect competition.

Economists have made various attempts to get around these dilemmas. We can classify these ways as attempts to broaden individual rationality in order to get both prisoners to the good outcome, or as ways to broaden the invisible-hand argument to accept bad outcomes. These two ways of seeking solutions

to their dilemmas with the PD game are signified by economists' placement of the apostrophe in the PD. If it is the prisoner's dilemma, the question focuses on individual rationality. If it is the prisoners' dilemma, the focus is on coordination or "social" outcomes. Most economists use the term Prisoner's Dilemma and focus on the individual problem (a rare exception being David Kreps, in both his microeconomics text and his charming book on modeling and game theory, both published in 1990[25]). But when economists write about the PD as a problem of invisible-hand outcomes and issues of coordination, they more often use the form Prisoners' Dilemma.[26]

Theoretical ways around the prisoner's dilemma depend on arguments suggesting that each individual even without an opportunity to collaborate will nevertheless rationally play the cooperative move in the first place. Perhaps people have a "disposition" to trust each other? Perhaps people follow a moral code that makes them trust each other? Perhaps people begin by trusting until that is proved wrong? On the whole, economists have not found these attempts to broaden their rationality principle very convincing,[27] leaving these kinds of speculations to those more philosophically inclined.[28]

However, *experimental* work on how the PD game is played by real people for money reveals a different picture from economists' *theoretical* accounts of how rational economic man plays the game. For example, the original experiments at RAND showed that in repeated rounds of the game some collaboration would occur. Since then, there have been many experiments with the game, in both economics and psychology.[29] The most famous of these is probably the extensive set of experiments conducted by Rapoport and Albert Chammah showing how players tended to converge to either CC or DD outcomes over a series of experiments,[30] but they also report considerable variations in outcomes. Robert Axelrod's series of tournaments reported in a 1984 book investigated various strategies or sequences of moves in playing a repeated PD game.[31] These strategies were programmed to play against each other in a computer tournament. Those that did best turned out to be rather collaborative and rather forgiving—"nice" strategies if you like. The results of experimental studies of the game hit at the heart of the theoretical results obtained for the PD. These experiments, along with other research directions, have pushed economists into exploring economic behavior more broadly, maintaining rational economic man only as a benchmark and reducing economists' concern about the individual prisoner's rationality.

If the question is concerned with the prisoners' dilemma, the solution lies not in the PD itself, but in the nature of the invisible-hand argument. Economists and philosophers have a preference for a benevolent version of the in-

visible hand, often dated to Adam Smith's use of the term and epitomized in this oft-quoted passage from his argument: "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages."[32] But individual self-interest does not necessarily produce benevolent outcomes, as the PD case demonstrates so easily. Here economists would do better to turn to an earlier manifestation of the invisible-hand argument available within the tradition of political economy, namely, the one proffered by Bernard Mandeville's *Private Vices, Publick Benefits* of 1705/1724.[33] In this fable, private vices lead via an invisible-hand process to public benefits, but the participants in the economy interpret the outcome as a bad one: the invisible hand is a malevolent one after all.

It is not possible here to trace the winding route and many paths by which neoclassical economists came to modify their belief in the pervasiveness of a benevolent invisible hand. Nevertheless, I believe I am on safe ground in arguing that game theory, and particularly the classic example of the PD, functions as an exemplary narrative in making this turn. In this respect I can do no better than quote from one of my London School of Economics colleagues in teaching students the lesson drawn from using the PD case during a middle-level course in economic theory: "The first law of economics is that individuals left to follow their own self-interest will reach a mutually beneficial outcome; the second law of economics is that this won't necessarily happen."[34]

DILEMMA 4: THE COMMENTATOR'S DILEMMA: MODEL SITUATION? EXEMPLARY NARRATIVE?

When we look at the way the PD game is used in economics, as opposed to the theoretical, empirical, and experimental study of the game itself, we encounter the problem of how to characterize its usage and role. On the one hand, we find that the use of mathematical notation is inextricably bound up with narratives. On the other, we find that game theory spawns models that are useful in particular cases rather than for general use. Game thinking, reasoning using game theory, does not fit easily with modern economists' picture of their science as a mathematical discipline, producing general theories derived deductively from general principles of rational behavior.

I focus on three elements to provide a characterization and thus resolve my own dilemma about how best to describe the way the PD game is used in economics. These elements I pose as questions: What are the roles of narrative?

How do economists reason about situations? Why does game theory spawn taxonomies?

What Are the Roles of Narratives?

When I first started attending to the role of narratives in economics, I took the trouble to listen carefully to how and where exactly economists used narratives. In seminars without game theory content, narratives were employed in the way economic models were used to answer questions.[35] But in economics seminars employing game theory, the role of narrative seemed truncated—narratives filled in the middle space between a set of individually rational actors with a matrix of numbers and an equilibrium solution. The narratives gave accounts of the situations, but as stories, they remained curiously unsatisfactory, for their endings were already presupposed, and the whole problem was how to get there. Although stories that are all middle do exist, it remains odd to label them as stories.[36] Yet I think now my sense that only middles were involved was, in fact, misleading.

When we look carefully at the account of the PD game given by Luce and Raiffa and the discussion of it given earlier, we can see that it is the interpretative text, not the matrix, that *implicitly* contains economists' traditional assumptions about individual rationality (i.e., which characterize the players) and about the necessity for equilibrium outcomes or solutions (a requirement of "good" economic theories) and *explicitly* contains the rules of the game (i.e., noncollaboration, simultaneous moves, etc.). All these, taken together, characterize the game situation. Thus the interpretative text is as necessary as the matrix of payoffs to the narratives that enable economists to use the game to reason in economic terms about another case at hand.[37] We might say that the narratives embody, or take for granted and depend on, the economic assumptions in the interpretive text, which is perhaps why I initially had the sense that they served only to fill in the middle of the story. In embodying the economic assumptions, they explain why the outcome is as it is, or wherein it is problematic, and how the situation might even be resolved. The narratives translate the prisoners' situation into the economic situation—they link particulars to particulars —and "explain" how it is, for example, that two large firms can end up doing damage to each other just as the prisoners end up with the double defect outcome.

We can also see how narrative elements help in thinking about how to resolve the prisoner's dilemma in Luce and Raiffa's treatment of the PD. Their broader narrative, for example, emphasizes the economists' interpretation of the rationality assumption, namely, that "neither suspect has moral qualms about or fear

of squealing."[38] They explore the possibility that the prisoners might reach the collaborative outcome if cooperation were allowed, but they immediately reject this by pointing out that this scenario goes against both the rationality assumption of each player maximizing individual returns and the equilibrium outcome that follows from that assumption taken in conjunction with the payoff matrix. (That is, the joint cooperation point cannot be an equilibrium solution because both have an incentive to defect.) But this formal issue seems less strong for Luce and Raiffa than their narrative musings about whether a binding agreement to cooperate would be broken by double-crossing, and these musings are generated by the text situation and considerations of whether the game can adequately represent that situation:

> Within the criminal context, such a "double cross" may engender serious reprisals and so it might be argued that it would not be worth while. This seems, however, to deny the utility interpretation of the given numbers [in the matrix]. If we have ignored such considerations in abstracting a game from reality, we had better include the breaking of a binding agreement as an integral aspect of an enlarged game purporting to summarize the conflict of interest. Alternatively, we may suppose that the effect of breaking a binding agreement is so disastrous that it is not considered.[39]

In other words, these ways around the outcome of the PD game lead to the respecification of the rules of the game (cooperation is possible) and/or the revision of the matrix of payoffs to reflect the changes in utility. There is no way to redefine rationality to fit the case and generate the collaborative outcome: "The hopelessness that one feels in such a game as this cannot be overcome by a play on the words 'rational' and 'irrational'; it is inherent in the situation. 'There should be a law against such games!'"[40]

Why should the government step in and "pass a law against such games"? Can the government legislate against the prisoner's dilemma game? Clearly not. But in expressing such sentiments, economists are no longer arguing about the PD game, but rather about the analogous situations to which it is applied. Governments have in the past, and still do, legislate in these kinds of situation, which serves almost as a litmus test of my argument about the way that the PD game is used, namely, that economists reason about economic situations by using games as models for them.

The identification of the model case with the real-world case is not so much a matter of formal analogy but of narrative elements that allow the economist to slip easily between the two cases. For example, Shaun Hargreaves Heap and

Yanis Varoufakis suggest that "the prisoners' dilemma arises as a problem of *trust* in every elemental economic exchange because it is rare for the delivery of a good to be perfectly synchronised with the payment for it."[41] That is, the PD game can be relevant whenever we have trade at a distance over time or space (by mail, Internet, or for future delivery). In these cases, both buyer and seller have to trust each other to deliver and not to cheat on the deal to their own advantage and the disadvantage of the other exchange party. We can thus say that we often (perhaps even daily) face such a prisoner's dilemma situation. But institutions or habits of exchange and trust, those very habits that economists since Hume have thought essential to the mechanism of the market, enable us to reach the mutually beneficial outcome rather than the mutually bad one. Here we take it for granted that our market institutions and exchange habits are backed by the law of contract, which does, in effect, legislate against the PD game outcome, namely, to curtail double-crossing.

The PD game has often been used to characterize situations of competition between large firms. In this context, it has also been commonplace to reinterpret the findings of famous past economists in such terms. For example, Kreps is one of many who use the PD game to reestablish Antoine Cournot's early nineteenth-century arguments about the behavior of rival mineral water companies:

> While the [PD] story is fanciful, the basic structure of options and payoffs that characterize this game occur over and over in economics. In this basic structure players can cooperate to greater or to lesser extent. If one player unilaterally decreases the level of her cooperation, she benefits and her rival is made worse off. Consider, for example, the case of Cournot duopolists [imagine Evian and Perrier] each (independently) choosing a quantity level to bring to the market. Typically, if one firm increases its production (which is a less cooperative strategy), its profits increase, at least for a while, and the profits of its rival decrease. But (past the monopoly level of output) if both firms increase their levels of output, both do worse.[42]

Notice how Kreps moves seamlessly from the game between two players into the competition between two Cournot firms, and how, at the same time, the game rules moved from a PD game (where no cooperation is allowed) to one of possible cooperation. As Kreps continues some pages later:

> With collusion, identical firms could each supply half the monopoly quantity, and together they would obtain the monopoly profits. But this isn't an equilibrium; if one side provides half the monopoly quantity, the

> other side has the incentive to supply more. . . . This isn't identical to the prisoners' dilemma game, since there we had a strictly dominant strategy for each side. But here, as there, we have (in equilibrium) each side taking actions that leave both worse off than if they could collude.[43]

Now we see this is not quite a PD game, but the characteristic outcome that characterizes the PD equilibrium remains. This, of course, remakes the exemplary point: that the self-interest outcome is not the best one. Once again the narrative matches the situation of the PD game with the economic case and enables a smooth transition in reasoning between the two while allowing a subtle change in the game specification.

Another example brings Marx into the game theory fold with his characterization of capitalists as paying their own workers low wages to maximize their own profits while hoping that all other capitalists will pay their workers high wages, thus increasing consumption demand.[44] This of course constitutes a version of the "free-rider" problem endemic in economic situations ranging from environmental pollution to labor supply. But here, the commentators have used the narration to move the situation surreptitiously from a two-person game to an *n*-person PD game.

This exactly reproduces the move made by Luce and Raiffa. Having discussed the prisoner's dilemma matrix and text, they give the following "alternative interpretation" for the PD game in a familiar narrative about farmers:

> As an n-person analogy to the prisoner's dilemma, consider the case of many wheat farmers where each farmer has, as an idealization, two strategies: "restricted production" and "full production." If all farmers use restricted production the price is high and individually they fare rather well; if all use full production the price is low and individually they fare rather poorly. The strategy of a given farmer, however, does not significantly affect the price level—this is the assumption of a competitive market—so that regardless of the strategies of the other farmers, he is better off in all circumstances with full production. Thus full production dominates restricted production; yet if each acts rationally they all fare poorly.[45]

The classic PD result: the invisible hand of self-interest leads the farmers to ruin. As I have noted earlier, there is no way, within the logic of the PD game, to solve this dilemma. Again, the ways around it must be sought by altering the rules of the game or some other aspect of the game played, or by widening the notion of rationality. And once again, these ways around are driven by the situation narrative providing the interplay between game and economic case:

> In practice, the equilibrium [of full production] may not occur since the farmers can, and sometimes do, enter into some form of weak collusion. In addition, a farmer does not play this game just once. Rather it is repeated each year and this introduces . . . an element of collusion. Finally, sometimes the government feels as we do, steps in, and passes a law against such games. Of course, in this analysis we have neglected the consumer. When he is included collusion may not be socially desirable even if it is desirable for the farmer.[46]

In this further narrative, the commentary has moved us from a single game to a repeated PD game, and this, as Luce and Raiffa already knew in 1957 from the experiments of the 1950s, moves us into a different game, one in which collusion is more than likely, particularly when, as in the farmers' case, the number of rounds remains unknown. But it is also, as we only learn later, a game where the so-called folk theorem (a theorem known from experience) holds, so that many equilibria may occur, so almost any outcome is possible.[47]

Narratives not only bind the PD game to the economic case and provide a means of reasoning about it but they also provide the means for probing the description of the case and hence the nature of the game, changing the latter if necessary to fit the former.

How Do Economists Reason about Situations?

The application of the PD game involves reasoning about situations. Here I make use of Karl Popper's method of "situational logic" or "situational analysis," which he characterized as *the* method of economic analysis and recommended as a methodological recipe for the social sciences in general.[48] Noretta Koertge formally described the basic elements of this recipe:[49]

1. Description of the Situation:	Agent A was in a situation of type C.
2. Analysis of the Situation:	In situations of type C, the appropriate thing to do is X.
3. Rationality Principle:	Agents always act appropriately to their situations.
4. Explanandum:	(Therefore) A did X.

The idea is that an analysis of the situation combined with a rational principle of action will define what it is logical to do in each type of situation and thus enable the social scientist to "explain" an action. Popper's 1967 discussion focuses our attention away from the rationality principle and onto the situation, the *typical* situation, as the important element in the analysis: "The theoretical

social sciences operate almost always by the method of constructing *typical* situations or conditions—that is, by the method of constructing models."[50]

Wade Hands has argued that this kind of logic is purpose built for economics and exemplifies exactly how economists do argue in the standard microeconomics of, for example, situations in which firms act rationally to maximize profits or consumers do to maximize utility.[51] Such situations are defined by, for example, the specific technologies of firms or the particular preferences of individuals, but in the analysis of deciding what is appropriate in such situations, microeconomists hold to some general behavioral theories or models providing standard outcomes. These theories are not very sensitive to the specifics of the situation, so that, in effect, the differences between situations hardly matter. Perhaps this is why both Hands and Bruce Caldwell,[52] who comment on Popper's ideas for economics and use Koertge's framework, nevertheless omit the important word "type" from their formulations. If there is no *type*, all situations appear either all the same or as all different individual cases. In the former case, general theory will do. In the latter, we need the kind of explanations we find in history—purpose built for each individual case. Situational analysis offers a middle level in between full generality and complete particularity only if there are different types of situations and an account of the "appropriate" thing to do for all situations of the same type.

If anything, game thinking seems much closer to fitting Popper's and Koertge's situational analysis and logic than does standard microeconomics, for here the choices or actions are very specific to the precise rules of the game and payoffs involved. In game theory and applications, the differences between situations really bite, so that even economists' thin but definite rationality principle does not always enable them to predict the outcome with ease. Yet defining the type of situation, categorizing a situation as a particular "type" of game, enables the economist to make limited generalizations about what kind of outcome will occur using the game type as the model situation.

Characterizing the analytical method in this way enables us to recognize, if we had not already done so from the examples treated above, how applications of game theory depend on the ability to match a description of an economic situation (real or hypothesized) to the description of a *type* of game in such a way that the appropriate behavior can be defined. Appropriate, for economists, as we have already seen, means not just using a rationality principle but locating some kind of an equilibrium outcome. Indeed, while much of economic game *theory* has concerned itself with defining the natures of different equilibrium concepts in different types of game situations, *applying* game theory depends on the matching of economic situations to those game types. As we have seen

above, the plausibility of this match is explored in the narrative sequences surrounding the application. If the match seems ill fitting, if the game is not appropriately specified, then the game specification is altered. As Popper argued for the social sciences, what economists do when they use game theory is "pack or cram our whole theoretical effort, our whole explanatory theory, in an analysis of the *situation*: into the *model*. . . . For in this field, the empirical explanatory theories or hypotheses are our various models, our various situational analyses. It is these which may be empirically more or less adequate."[53]

Popper's situational analysis offers only one of the elements I need to describe economists' reasoning with game theory.[54] Game thinking enables economists to maintain their thin rationality and yet give an account of what will happen in each particular type of situation: this is the domain of situational analysis.[55] Explanatory *depth* is provided by the narratives that match the situational analysis to events in the world. In the next section, we shall see how economists gain explanatory *breadth* across situations through the considerable possibilities for variation in the model situations or types.

Why Does Game Theory Spawn Taxonomies?

Game theory manages to multiply the games it studies and so to generate more types of situation that can be characterized as games. This occurs via two mechanisms.

On the one hand, game theory has traditionally grown by filling in the holes in a taxonomy. Luce and Raiffa introduce the PD game in a chapter entitled "Two-Person Non-Zero-Sum Non-co-operative Games,"[56] thus designating the game as of a particular type in a taxonomy with six categories: two versus *n* people; zero-sum versus non–zero-sum; and cooperative versus noncooperative. Of course, faced with such a taxonomy, the natural theorist will find ways to fill the empty boxes by investigating extensions of particular games within a certain class: extending a two-person game to an *n*-person game; extending a game without cooperation to one with cooperation; games with one period to finite periods to infinite periods; games with zero-sums to non-zero sums; and so forth.

This method of extending theory produces taxonomies of games that themselves change over time. New classifications emerge out of old types; new questions generate new types. Thus the taxonomies have not just grown by adding cells over the past fifty years but they have also changed their categories. The categories recognized in Luce and Raiffa in 1957 had entirely changed by the time of Drew Fudenberg and Jean Tirole's text of the 1990s.[57] In the 1990s, static versus dynamic games with complete versus incomplete information provided

the basic four-cell taxonomy with subcategories of multistage and repeated games, games in normal or strategic versus extensive form (the form of representations, which are not fully equivalent), and so forth.

The labeling of a particular game also changes as class boundaries are revised, and these change in turn as the economists come to focus on different aspects and so analyze different features of the games. The PD game is classified in the 1991 text as a static game of complete information, and its repeated version as a dynamic game of complete information, whereas in the 1957 text, it is classified under two-person, non–zero-sum, noncooperative games, and its n-person version in an equivalent n-person cell.

On the other hand, game theory also extends by attempts to characterize particular economic situations, empirical or hypothetical ones, as game situations, and thus to type them as a particular category of game. Martin Shubik in 1953 produced a taxonomy matching economic situations to game types in a table with a grand title:[58]

GENERAL THEORY OF GAMES

Cooperative Games	Semi-Cooperative Games	Non-Cooperative Games
Duopoly, Duopsony	Oligopoly	Monopoly, Monopsony
Oligopoly	Cartel Theory	Pure Competition
Cartel Theory		Cournot, Bertrand Duopoly
Bilateral Monopoly		Macro-economics

More typically, new categories grow from the narratives, which as we have seen go through a process of matching the economic situation with the game situation and then exploring how and why it does not fit. When it does not fit, a new version of the game is developed with slight changes in the rules, payoffs, or information arrangements. Sometimes this revision turns out to be a different type of game, yet sometimes it is like the original type. We saw examples of this kind of reasoning with the PD game earlier. It has perhaps been most clearly evident in the industrial economics literature, where a serious tradition of using game reasoning to extend the economic theory of firm competition, and to understand the exact details of empirical cases, goes back to Shubik's seminal work of the 1950s.[59]

This integration of game theory into industrial economics by Shubik initially appeared highly effective in reexploring a number of classic results in the

field of the theory of the firm (dating from the nineteenth century and from the 1930s), in extending results for those cases and in allowing for the comparison of their "solutions." Game reasoning appeared to provide a new and constructive tool of analysis in the field: it offered the possibility of the analysis of strategic decisions based on situations offered by game theory in conjunction with more traditional microeconomic theory of the firm's profit-making possibilities. The approach seemed to combine the benefits of situation-based thinking with general theories of microeconomic behavior.

But by the early 1990s, the outcome, in industrial economics at least, was found less rewarding than promised. First of all, as became clearer, there were very many possible ways of characterizing economic situations as games as each game depended on many detailed specifications. For example, Sam Peltzman wrote down a "non-exhaustive list" of twenty "questions that arise in formulating and solving game-theoretic models—questions whose answers can crucially affect results."[60] These ranged from such simple ones as how many players there are and who moves first to more difficult ones concerning the nature of the equilibria in the model. The answers to these many questions characterize the rules and institutions of either a hypothetical situation imagined in the model or an empirical situation under study. Peltzman was pessimistic that "the interminable series of special cases" generated by theorists had been of any help in analyzing empirical cases.[61] Franklin Fisher was equally sharp about the way in which theoretical cases multiplied and about how they were able to provide little reliable help for an analysis of oligopoly in his comment on the folk theorem: "Anything that one might imagine as sensible can turn out to be the answer. . . . This is a case in which theory is poverty-stricken by an embarrassment of riches."[62]

Nevertheless, Fisher did not accept that finding out about the series of special cases was pointless. He interpreted the exercise not as one of a failed general theory, but as a good example of exemplifying theory, particularly useful in thinking about cartels and oligopolies, a field in which general theories had been least effective:

> When well handled, exemplifying theory can be very illuminating indeed, suggestively revealing the possibility of certain phenomena. What such theory lacks, of course, is generality. . . . The status of the theory of oligopoly is that of exemplifying theory. We know that a lot of different things *can* happen. We do not have a full, coherent, formal theory of what *must* happen or a theory that tells us how what happens depends on well-defined, measurable variables. . . . At present, oligopoly theory consists of

> a large number of stories, each one an anecdote describing what might happen in some particular situation. Such stories can be very interesting indeed. Elie Wiesel . . . has said that "God made man because He loves stories," and economists (not merely game theorists) are plainly made in the divine image in this respect.[63]

Game theories exemplify typical situations or cases, and these typical cases are used to characterize empirical situations. In this sense, both theoretical work and empirical work proceed in the same way in this field, as examples of case-based reasoning. But whereas at the end of the 1930s industrial economists had four case situations, four typical situations, in their box of exemplars (perfect competition, monopoly, and two types of imperfect competition), by the 1990s game theory had filled their box of exemplifying theories for different types of situation to overflowing. For Fisher and Peltzman, typical situations had degenerated into a series of special cases or particular stories. While Fisher found this liberating, Sutton appeared more critical. He suggested that the flexibility of game theory to capture various situations was embarrassing because "given any form of behaviour observed in the market, we are now quite likely to have on hand at least one model which 'explains' it—in the sense of deriving that form of behaviour as the outcome of individually rational decisions."[64] But rather than producing an embarrassment of riches, Sutton saw it as embarrassing for game theory. As he so bluntly stated it: "In 'explaining' everything, have we explained nothing? What do these models exclude?"[65] With every economic situation potentially matched by more than one candidate model from game theory, the possibilities of using game theory for explanations in terms of types of situation—the middle-level explanatory power of situational analysis—is lost.[66] Explanatory breadth, obtained by the development of further typical situations, and therefore cells in a taxonomy, appears to have drowned in a sea of one-off individual cases and anecdotes.

EPILOGUE: A SCIENCE WITHOUT LAWS?

This article has argued through a series of four dilemmas. It began with the strategic dilemma of the Cold War and the parallel development of dilemma games in the economics of game theory, suggesting that these also posed a dilemma for historians writing about these events. It continued by discussing the way in which a particular game, the Prisoner's Dilemma game, embodied the dilemma for individuals portrayed in economic models, a dilemma that follows from economists' assumption of individual rationality and the requirement for equi-

librium outcomes. Rather than a one-off case of little interest, the Prisoner's Dilemma game turned out to have exemplary qualities.

In his 1984 book on the evolution of cooperation, Axelrod remarked on the infectious quality of the PD in social psychology: "The iterated Prisoner's Dilemma has become the *E. coli* of social psychology."[67] The PD infection was equally invasive in economics, for economists came to use the case of the PD whenever they wished to discuss particular situations in which individual rationality leads to a supposedly irrational outcome.[68] Once economists started thinking about the nasty outcomes in economics that might be described in terms of a PD game, they began to find many such PD situations in the economy. For example, Michael Carter and Rodney Maddock used the two-person game to model the inflationary outcome from the interaction between government and unions.[69] Hargreaves Heap has surveyed the ways in which the PD game can be used as a model for understanding the institutional backgrounds for macroeconomic performance.[70] It has been used to model Gresham's Law (that bad money drives out good),[71] international fishing wars, productivity problems,[72] and so forth. But, and this is the dilemma explored in the third section of my essay, in developing and using the Prisoner's Dilemma game as exemplary for many such situations across many subfields of economics, economists undercut their most general lawlike claim—namely, that the social outcome of self-interested individual actions will be a good one. As a result of the infection, economics founded a second law of economics, namely, that the first law will not necessarily hold.

The final dilemma was the problem of how to characterize game reasoning in economics given that it did not fit easily into economists' perceptions of their science as a deductive, mathematical science applying general theories. Here the dilemma was resolved by an account in which narrative, situational analysis, and types of cases were fit together to portray reasoning with game theory as a kind of case-based reasoning. Situational analysis focuses the explanatory power into accurate descriptions of typical situations—into the model—so that even with a thin notion of rational behavior, specific outcomes can be deduced from the game theory for an economic situation of that type. Narrative plays the important role of enabling the economist to check that the chosen model type matches the economic situation, that is, the economic situation is accurately described in the model. By exploring the features of that match, narrative provides a sense of explanatory depth to the specific cases discussed. Explanatory breadth is derived from the development of a full taxonomy of typical cases, so that different model situations span the various empirical situations in such a way that real-world cases can be categorized in terms of a type of

case. This combination should give local explanatory power at the level of types of cases. But at least in industrial economics, economists reached the point at which each case seemed different. And since explanatory power resides in the accurate description of the model situation, not in the general but thin rationality that animates the models, game theory claims both explanatory depth and breadth yet does not achieve generalized explanations. Sutton's article title "Explaining Everything, Explaining Nothing" is indicative of a science full of models but without laws.

In this process of case-based reasoning, of arguing from model situation to real case and back again, economics has undergone the same kind of transition that we observed in discussing our first dilemma, the problem of how to characterize and discuss nuclear strategy. There, we saw that part of the "civilizing" of discussions about war during the Cold War meant that nuclear war became nuclear "exchange," a reduction to economic language; and that game theory, originally a field of strategy that might be applied to Cold War problems, became the way commentators saw the Cold War world. The models of game theory began as models of, and perhaps a guide to action for, war strategies, but then the war situations themselves became seen as games. The same happened in economics. The Prisoner's Dilemma entered the *OED* in 1989, and I hazard a guess that it is around this time that economists stopped seeing the PD as a model of, and/or for, strategic behavior in economic situations and began to see PD situations directly in the economy. The PD game began as the lens through which economists observed less-than-happy outcomes in the economy. But they are now the things economists see in the economy, and they apparently see them almost everywhere.

NOTES

This article was written for one of the Princeton Workshops in the History of Science titled "Model Systems, Cases, and Exemplary Narratives," February 10, 2001 (and became Research Memoranda in History and Methodology of Economics 01-7, University of Amsterdam, 2001). A section of the essay was redrafted for the INEM (International Network for Economic Method) conference at Stirling in September 2002 under the title "Game Theory Explanations in Economics: Reasoning about Model Situations," and subsequent versions were given at the History of Science Society meeting in Austin, Texas, and the Second Siena Workshop on the History of Economics, both of which took place in November 2004.

I thank the Princeton workshop organisers: Angela Creager, Elizabeth Lunbeck, and Norton Wise for their invitation and responses, and Suman Seth for his commentary. I am grateful to my London School of Economics (LSE) research

assistants, Till Gruene and Gabriel Molteni, who showed incredible patience with my library requests. My thanks go also to the British Academy for funding this research, to Bruce Caldwell, Robert Leonard, John Sutton, Margaret Bray, Nancy Cartwright, Ned McClennen, and my colleagues at the University of Amsterdam and at the LSE for their helpful comments.

1. Jacob Bronowski, *The Listener*, July 1, 1954, quoted in W. Poundstone, *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb* (New York: Doubleday, 1992), 265.
2. I note also how the language of nuclear war has been "civilized" into one of "exchange." David Hume argued that trade (exchange) was a civilizing influence on nations' relationships with each other; trade would replace war. Here war has become trade.
3. Anatol Rapoport, *Two-Person Game Theory: The Essential Ideas* (Ann Arbor: University of Michigan Press, 1966), 189–91.
4. He had aired these difficulties in a more public space in A. Rapoport, "The Use and Misuse of Game Theory," *Scientific American*, December 1962, 108–18.
5. Poundstone, *Prisoner's Dilemma*, 1.
6. Ibid.
7. To provide another example from Poundstone: game theorists might use the game of Chicken (two young men driving toward each other in the middle of the road to see who gives way first) to analyze the Cuban missile crisis. But, if they were Rapoport, and retained a strong sense of how far such games were from representing individual behavior, or Martin Shubik, who retained a sense of how difficult it was to represent the world in the matrix, they would warn strongly against politicians acting on the assumption that the crisis was a simple game of Chicken. See Rapoport, *Two-Person Game Theory*; and M. Shubik, "Game Theory, Behavior, and the Paradox of the Prisoner's Dilemma: Three Solutions," *Conflict Resolution* 14 (1970): 181–93.
8. Another dilemma for the historian lies in marrying the serious context and content of game thinking and the avowed purpose of its funding with the anecdotal accounts by those involved. The tenor of the anecdotes stress the playroom aspects: groups of primarily young and male mathematicians and social scientists, at least those at RAND and Princeton, had enormous fun, devising and playing games that spilled over from work into social times. As is well known, their work was funded by the military, but the accounts give little sense of the immediacy of Cold War threats, or indeed of the fact that their activities might have anything at all to do with Cold War interests. Poundstone, *Prisoner's Dilemma*, does an excellent job of laying out these disparate elements.
9. See E. Roy Weintraub, *Toward a History of Game Theory* (Durham, N.C.: Duke University Press, 1992); Sylvia Nasar, *A Beautiful Mind: A Biography of John Forbes Nash, Jr., Winner of the Nobel Prize in Economics, 1994* (New York: Simon and Schuster, 1998); Philip Mirowski, *Machine Dreams: Economics Becomes a Cyborg Science* (Cambridge: Cambridge University Press, 2001); N. Giacoli, *Mod-*

eling Rational Agents: From Interwar Economics to Early Modern Game Theory (Cheltenham, UK: Edward Elgar, 2003); R. J. Leonard, *From Red Vienna to Santa Monica* (New York: Cambridge University Press, forthcoming). Each of these works covers various aspects of the history of game theory. See entries on the PD and game theory in *The New Palgrave: A Dictionary of Economics*, ed. John Eatwell, Murray Milgate, and Peter Newman (London: Macmillan, 1998) for practitioner histories. M. S. Morgan, "Economics," in *The Modern Social Sciences*, vol. 7 of *The Cambridge History of Science*, ed. T. Porter and D. Ross (Cambridge: Cambridge University Press, 2003), 275–305, provides a more general background to the history of economics in the twentieth century.

10. M. Flood, "Some Experimental Games," in Research Memorandum RM-789, RAND Corporation, Santa Monica, 1952.
11. H. Raiffa, "Game Theory at the University of Michigan, 1948–52," in Weintraub, *Toward a History of Game Theory*, 165–75, see especially 171–73.
12. R. Duncan Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* (New York: Wiley, 1957), 95.
13. For example, the Battle of the Sexes and Chicken are two other classic two-person dilemma games usually discussed alongside the PD game.
14. The game may also be represented and described in its "extensive form" (a branching tree diagram showing the choices and payoffs), or even, in the early days, in terms of its "characteristic function" (the individual and combined maximum rewards possible).
15. For example, Eric Rasmusen, *Games and Information: An Introduction to Game Theory* (Oxford: Blackwell, 1989), 30. Also see R. Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984), 9–10.
16. Poundstone, *Prisoner's Dilemma*, 116–18.
17. The *OED* entry also notes Albert Tucker's more general claim to the game's provenance on the grounds of his narrative about the prisoners. Raiffa, "Game Theory at the University of Michigan, 1948-52," 173, gives an account of the disputed provenance (with RAND) of the type of game, suggesting that it was "folk knowledge" of that time.
18. Luce and Raiffa, *Games and Decisions*, 95.
19. See Poundstone, *Prisoner's Dilemma*, 2, chap. 6.
20. Luce and Raiffa, *Games and Decisions*, 96.
21. J. S. Mill, "On the Definition of Political Economy," in *Collected Works of John Stuart Mill: Essays on Economics and Society*, ed. J. M. Robson (Toronto: University of Toronto Press, 1836).
22. See M. S. Morgan, "Economic Man as Model Man: Ideal types, Idealization and Caricatures," *Journal of the History of Economic Thought* 28, no. 1 (2006): 1–27.
23. Thought experiments—or "inductive modelling," as Rapoport termed the process in which he imagined how the possible, or hypothetical, reasoning of the individuals might result in different kinds of behavior—nevertheless led to the same conclusions. See Rapoport, *Two-Person Game Theory*, chap. 10.

24. This result does not necessarily hold for the game repeated an infinite (or perhaps unknown) number of times, but it is not clear that theory tells us this.
25. David M. Kreps, *Game Theory and Economic Modelling* (Oxford: Clarendon, 1990). Also see David M. Kreps, *A Course in Microeconomic Theory* (New York: Harvester Wheatsheaf, 1990).
26. For example, G. Tullock, "Adam Smith and the Prisoners' Dilemma," *Quarterly Journal of Economics* 100 (1985): 1073–81.
27. See, for example, Shubik, "Game Theory, Behaviour, and the Paradox of the Prisoner's Dilemma."
28. See Shaun Hargreaves Heap and Yanis Varoufakis, *Game Theory: A Critical Introduction* (London: Routledge, 1995); and R. Campbell and L. Sowden, *Paradoxes of Rationality and Cooperation* (Vancouver: University of British Columbia Press, 1985).
29. The two disciplines do not necessarily make the same inferences from the same experiments, as R. J. Leonard, "Laboratory Strife: Higgling as Experimental Science in Economics and Social Psychology," in *Higgling: Transactors and Their Markets in the History of Economics*, ed. Neil de Marchi and Mary S. Morgan (Durham, N.C.: Duke University Press, 1994), 343–69, shows in the context of experiments with bargaining games.
30. A. Rapoport and A. M. Chammah, *Prisoner's Dilemma* (Ann Arbor: University of Michigan Press, 1965).
31. Axelrod, *The Evolution of Cooperation*.
32. Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, 1776, ed. R. H. Campbell and A. S. Skinner (Oxford: Oxford University Press, 1976), Book 1, Chapter 2.
33. Bernard Mandeville, *The Fable of the Bees, or, Private Vices, Publick Benefits*, ed. F. B. Kaye (Oxford: Clarendon, 1927 [1705/1724]); E. Ullman-Margalit, "Invisible-Hand Explanations," *Synthese* 39 (1978): 263–91. Ullman-Margalit's insightful discussion of invisible-hand explanations points out that on hearing them, we should find them both plausible and surprising. The PD game, I think, fits this description, and it is often associated with invisible-hand explanations, but it does not fit the other main element of her description: that they involve aggregate outcomes.
34. Margaret Bray, "Micro-economic principles," an intermediate-level course at the London School of Economics, Fall 2000.
35. See M. S. Morgan, "Models, Stories and the Economic World," *Journal of Economic Methodology* 8 (2001): 361–84.
36. Ursula Le Guin, "It Was a Dark and Stormy Night; or, Why Are We Huddling about the Campfire?" in *On Narrative*, ed. W. J. T. Mitchell (Chicago: University of Chicago Press, 1980), 194, gives an example of a minimalist middle that she is prepared to label "a whole story," one carved in runes in Carlisle Cathedral: "Tolfink carved these runes in the stone." This might be sufficient of a story to narrate a situation and outcomes as we will see in my argument's continuation (though

it is not sufficient to be a game situation unless the runes have been misread and two people were there: Tol and Fink!).

37. In contrast to L. O. Mink, "History and Fiction as Models of Comprehension," *New Literary History* 1 (1970): 541–58, who presents scientific and narrative explanations as competing, I see them as complementary here, as in my previous discussion on the role of narrative in economics. See Morgan, "Models, Stories and the Economic World."
38. Luce and Raiffa, *Games and Decisions*, 95.
39. Ibid., 96
40. Ibid., 96–97.
41. Hargreaves Heap and Varoufakis, *Game Theory*, 149.
42. Kreps, *A Course in Microeconomic Theory*, 504.
43. Ibid., 524.
44. Hargreaves Heap and Varoufakis, *Game Theory*, 154 (they quote from Marx).
45. Luce and Raiffa, *Games and Decisions*, 97
46. Ibid.
47. See D. Fudenberg and J. Tirole, *Game Theory* (Cambridge: MIT Press, 1991).
48. Karl Popper, "The Autonomy of Sociology," 1945 in *Popper Selections*, ed. D. Miller (Princeton: Princeton University Press, 1985), 345–56.
49. N. Koertge, "The Methodological Status of Popper's Rationality Principle," *Theory and Decision* 10 (1979): 87.
50. Karl Popper, "The Rationality Principle," 1967, in Miller, *Popper Selections*, 357–58; emphasis original.
51. D. W. Hands, "Falsification, Situational Analysis and Scientific Research Programs," in *Post-Popperian Methodology of Economics: Recovering Practice*, ed. Neil de Marchi (Dordrecht, the Netherlands: Kluwer, 1992), 27–31.
52. B. J. Caldwell, "Clarifying Popper," *Journal of Economic Literature* 29 (1991): 1–33.
53. Popper, "The Rationality Principle," 359–60; emphasis original.
54. It seems more than likely that my characterization will not transfer to other social sciences. For example, Rapoport, a social scientist whose base field is psychology, defines the failure of the thin rationality to provide plausible explanatory devices in game situations as one of the most important aspects of game theory, even one of its most important achievements: game theory reveals all too clearly what social scientists do not know about how humans behave.
55. This is consistent with my earlier argument about the way narratives explore the possibilities within the model. See Morgan, "Models, Stories and the Economic World."
56. Luce and Raiffa, *Games and Decisions*, 88–113.
57. Fudenberg and Tirole, *Game Theory*.
58. M. Shubik, "The Role of Game Theory in Economics," *Kyklos* 6 (1953): 27.
59. M. Shubik, *Strategy and Market Structure* (New York: Wiley, 1959).
60. S. Peltzman, "The Handbook of Industrial Organisation: A Review Article," *Journal of Political Economy* 99 (1991): 201–7.

61. Ibid., 206.
62. F. M. Fisher, "Games Economists Play: A Noncooperative View," *RAND Journal of Economics* 20 (1989): 116.
63. Ibid., 118.
64. J. Sutton, "Explaining Everything, Explaining Nothing?" *European Economic Review* 34 (1990): 506.
65. Ibid., 507.
66. In Sutton's case, this outcome has lead to the development of his so-called class-of-models approach which, in my view, reestablishes the middle level of explanatory power, not at the level of typical cases, but at the empirical level of industry characteristics. See John Sutton, *Marshall's Tendencies: What Can Economists Know?* (Cambridge: MIT Press, 2000); and my commentary on that article, "How Models Help Economists to Know," *Economics and Philosophy* 18 (2002), 5–16.
67. Axelrod, *The Evolution of Cooperation*, 28.
68. At the point when a TV quiz show can be labeled a PD game show, as happened during the period I was writing this article, it is apparent that the PD "infection" has spread beyond the social scientists' world.
69. M. Carter and R. Maddock, "Inflation: The Invisible Foot of Macroeconomics," *Economic Record* 63 (1987): 120–28.
70. S. Hargreaves Heap, "Institutions and (Short-Run) Macroeconomic Performance," *Journal of Economic Surveys* 8 (1994): 35–56.
71. G. Selgin, "Salvaging Gresham's Law: The Good, the Bad and the Illegal," *Journal of Money, Credit and Banking* 28 (1996): 637–49.
72. H. Leibenstein, "The Prisoners' Dilemma in the Invisible Hand: An Analysis of Intrafirm Productivity," *American Economic Review* 72 (1982): 92–97.