

Interdisciplinary Gap Filling: Game Theory and the Law

Eric Talley

DOUGLAS BAIRD, ROBERT GERTNER, AND RANDAL PICKER. *Game Theory and the Law*. Cambridge: Harvard University Press, 1994. Pp. 320. \$46.95.

More than a half century has passed since John Von Neumann and Oskar Morgenstern published their pioneering work that provided the foundations for the field now known as "game theory."¹ By redirecting analytical focus away from a static environment (where individuals merely "react" to their surroundings) and toward a strategic one (where each individual's very surroundings depend centrally on others' reactions), game theory offered a fresh account of human behavior that was descriptively richer and more satisfying than most of its predecessors. In the decades since, the emergent field has cultivated significant advances not only in economics and finance, but also in disciplines as diverse as psychology, biology, political science, and eventually, law. Indeed, by the beginning of this decade, prominent law reviews began to provide a steady stream of articles utilizing the game-theoretic approach.

It was therefore hardly surprising when in 1994, Douglas Baird, Robert Gertner, and Randal Picker (all of the University of Chicago) published *Game Theory and the Law*, which represents the first comprehensive manual

Eric Talley is associate professor of Law, University of Southern California Law School. The author offers many thanks to Jennifer Arlen, Jenna Bednar, Ken Dau-Schmidt, Ariela Gross, Gillian Lester, and Matt Spitzer for helpful comments. Kendall Evans and Suzy Bramzon provided excellent research assistance.

1. While Von Neumann and Morgenstern's work (1944) was the first comprehensive account of complete information games, a number of important papers had earlier established the rudiments of game theoretic analysis in more special cases (Aumann 1989).

for strategic analysis of legal rules. In fact, given the amount of academic work in this field since the early 1970s (Brown 1972) and continuing today, perhaps the greater surprise was that it had taken so long for such a book to arrive. Its release was widely anticipated by legal academics, game theorists, and those somewhere in between.

Shortly after its publication, I had what can best be described as an "immersion experience" with the content of *Game Theory and the Law*: I taught from it. During the Spring semester of 1996, I led a small but stout-hearted group of USC Law School students through a course entitled "Strategic Behavior and the Law," for which the book was our principal substantive text. (As I was rather inexperienced in law teaching, I even chose the *title* of the course strategically, fearing gross undersubscription from students wary of any new professor teaching a seminar whose title included the word "theory"). The course gave me a unique opportunity to examine the book from the alternating perspectives of a researcher already well versed in the approach *and* as a pupil, since it was I who paid the price when students had difficulties with the more challenging parts of the text. With this essay, I can reflect on the results of this (rather uncontrolled) experiment.

Because this book is now three years old, mine is far from the first review of it. Previous efforts (e.g., Dau-Schmidt et al. 1997; Huang 1995; Lucy 1995; Morriss 1995; Salant and Sims 1996) have made a number of lucid and subtle observations about the book with which for the most part I concur. Thus, while I shall devote *some* time to reflecting on the relative merits of the text, my ultimate aim is neither a tour through the contours of its pages nor a self-contained tutorial on the fundamentals of game theory. While both tasks are important, they have both been accomplished quite capably (Salant and Sims 1996). Rather, I shall train my focus on a different question: What now? More explicitly (if not succinctly), what strata of the intersection between game theory and legal rules does the book (and the current literature) leave relatively unexplored, and how might one go forth in uncovering them?

The answer I shall offer to this question emanates from my primary criticism of the text: its omission of a more systematic presentation of the concept known as "common knowledge." Explicitly, a statement or fact is said to be common knowledge among a group of individuals if each person knows it, each person knows that each person knows it, and so on, ad infinitum.² Assumptions of common knowledge (in one manifestation or another) are central to virtually every solution concept within game theory proper. Relaxing this assumption can have profound effects on the set of plausible "outcomes" that one might predict for a given strategic setting.

2. For the formal mathematical definition, see Aumann 1976 and Geanakopoulous 1992.

Appreciating the power of the common-knowledge assumption is not only a prerequisite for understanding game theory on its own merits, but—I shall contend—it seems *especially* critical for developing a comprehensive account of game theory *and* law. Indeed, the law plays an integral role in shaping and regulating the interaction between players who possess possibly divergent interests and beliefs. But the manner in which it fulfills this role is necessarily constrained by the players' individual and common understanding (or lack thereof) of the existence, content, and applicability of legal rules.

To animate this argument, I will use an example inspired by an early contracts case that helps illustrate the critical relationship between law, communication, and common knowledge. In particular, the example demonstrates that breakdowns in common knowledge can have rather dramatic effects on behavior; and that legal rules, while perhaps unable to “resolve” this problem, can frequently mitigate its consequences. While the immediate applications of the example are rather narrowly limited to “offer and acceptance” doctrines, the possibilities it suggests extend much further, and may be particularly intriguing for those interested in how the law helps to shape and express collective consciousness and in establishing norms for individual behavior.

Before plunging into this extension, however, a short synopsis and analysis of the book are in order.

I. A BRIEF ROAD MAP OF THE TEXT

The 267 substantive pages of *Game Theory and the Law* consist of a modest introduction, eight substantive chapters, and a brief conclusion. While reading its introductory chapters, one is struck by at least two enormous expository challenges the authors faced in drafting the book. First, the standard pedagogical approach in much of game theory is to use a set of stock canonical examples to illustrate more general conceptual and mathematical phenomena. While example-based analysis might seem perfectly suited to legally trained readers (indeed law school training is similar), a clear role for legal rules is absent in many of the extant canonical examples. Nevertheless, a satisfying account of “law and game theory” must incorporate the role of legal rules explicitly.

Second, the authors chose not to capture “one audience” for the book, but rather directed their efforts toward crafting a book that *both* “offers those interested in law a new way of thinking about legal rules” *and* “shows those interested in game theory a fertile and largely unexplored domain in which its tools have many applications” (pp. xi–xii.). I shall refer in what follows to these two audiences as Lawyers and Game Theorists.

With one notable exception (discussed below), Baird, Gertner, and Picker have overcome both these obstacles admirably. The book frequently makes use of examples that are specially tailored to have clear legal content, and its pitch capably draws in both the Lawyers and the Game Theorists (often simultaneously). For these accomplishments alone, it deserves substantial praise.

Chapter 1 lays out the fundamental components (or what are often called the primitives) of a noncooperative game. Because these rudiments are among the most difficult for newcomers to master, the authors begin sensibly with the simplest possible class of games: simultaneous move games of complete information—often called normal-form (or strategic-form) games. The fundamental structure of such a game amounts to a description of (1) the relevant players, (2) the strategies available to each player, and (3) the payoffs each player receives as a consequence of each combination of strategies chosen.

Consider, for example, a simple “coordination” game between player 1 and player 2, who must each decide simultaneously on which side of a two-lane street they will drive as they approach one another at high speed. The three primitives of this model can be captured in a matrix-like diagram, which might appear as follows:

		Player 2	
		Left	Right
Player 1	Left	1, 1	-9, -9
	Right	-9, -9	1, 1

The above matrix communicates enough information to identify the players (player 1 and player 2), their available strategies (left side and right side for each), and their payoffs (illustrated by the numerical payoffs for each player under each possible combination of strategies the players might adopt³).

After defining the primitives of a normal-form game, it is typical to make three further (and somewhat more controversial) assumptions: (4) that the players are self-interested (i.e., their goal is to maximize their payoff when playing the game); (5) that the players are “rational” (i.e., they behave in a way that is consistent with the goal of maximizing their pay-

3. The numerical payoffs within each cell for each player are typically set apart by a comma, in which the first entry is player 1's payoff and the second is player 2's. In the coordination game, the payoffs are 1 apiece if the players can “coordinate” and pass by each other without a collision, but are -9 if the players fail to coordinate and end up in a head-on collision.

off⁴); and finally (6) that both the primitives of the game—(1), (2), and (3)—and the behavioral assumptions of the game—(4) and (5)—are common knowledge.

As noted above, it is standard to illustrate the fundamental structure of a normal-form game through use of various canonical examples such as the coordination game (above), or some other well-known example such as the Prisoner's Dilemma.⁵ In a surprising pedagogical move, the first few pages of the book sidestep this conventional approach entirely, illustrating key concepts using a specially crafted example more often associated with standard law and economics literature than with classical game theory: the strategic analysis of tort law, and the standards of due care (Shavell 1987). Using this example, the authors identify the foundational elements that define a normal-form game, and then begin to introduce analytic arguments that narrow the range of plausible outcomes of the game.

In particular, the accident example introduces readers to the fundamental centerpiece of classical game theory: Nash equilibrium. Informally, a Nash equilibrium consists of a strategy for each player that is essentially "self-enforcing," in that no player has an affirmative incentive to abandon her prescribed strategy if the other players are employing their prescribed strategies.⁶ While one of the weakest (in a predictive sense) of the extant solution concepts within game theory, Nash equilibrium is an important one because it delineates the behavior in which rational, self-interested actors would plausibly engage. Using this and related concepts, the authors walk through an argument about how legal rules, in the guise of various tort regimes (such as negligence with contributory versus comparative negligence), can alter the payoff structure for the parties, thereby altering their

4. Note that this concept of rationality is means based rather than ends based in its ilk. Thus, a classic game theory analysis takes individual payoffs as given, and predicts behavior assuming that the actors are attempting to maximize them.

5. The Prisoner's Dilemma is the numerical equivalent of the "tragedy of the commons." In it, two prisoners accused of criminal conduct are placed in separate rooms and asked simultaneously either to "confess" to the crime or to "remain silent." Although remaining silent would secure the best *joint* payoff for the parties combined, the payoffs of the game are structured so that each player *individually* does better by confessing than by remaining silent, no matter what the other player is doing (Fudenberg and Tirole 1991, 11; Osborne and Rubinstein 1994, 17; Van Damme 1996, 2).

6. In the coordination game pictured above, one such Nash equilibrium consists of both players choosing the right side of the road. Yet another consists of both players choosing left. As it turns out, the coordination game also has a third "mixed strategy" equilibrium, in which the players randomize between their various available strategies rather than playing one such strategy with certainty. In it, each player randomizes with a probability of $\frac{1}{2}$ whether to drive on the left hand side of the road. To see this, simply note that if player 2 were randomizing in such a fashion, player 1's expected, or "average," payoff from driving on the right hand side of the road would be -4 (i.e., $\frac{1}{2}(-9) + \frac{1}{2}(1)$). This is also the expected payoff player 1 would receive from driving on the left hand side of the road. Because her payoff is the same no matter what action she takes, player 1 would be willing to randomize over her strategies with probability $\frac{1}{2}$ on each. The same holds true for player 2. Thus, when each player is randomizing in this fashion, neither has a positive incentive to change her own strategy.

incentives, and in turn, the likely equilibria that emerge. Moreover, the authors point out, there need not always be one unique way of altering incentives to induce desired behavior: for each tort regime they analyze gives rise to equilibria in which both potential injurers and victims exercise the efficient levels of precaution.

While one can quibble with this nonstandard pedagogical approach of this introductory chapter (as indeed some have [Salant and Sims 1996, 1866]), it nonetheless serves a number of useful purposes, irrespective of the audience. First, it signals to the Lawyers that the examples in the text will be tailored to legal applications as much as possible. Second, it signals to the Game Theorists that these authors aspire to something greater than a simple summary review of standard game theory (with the words *plaintiff* and *defendant* substituted for *player 1* and *player 2*).⁷

An additional analytical step in the chapter comes with the introduction of the possibility that one might detect *multiple* Nash equilibria for a given game.⁸ The multiple equilibrium problem poses a sizeable challenge to game theory proper, especially if one's ultimate goal is to make accurate predictions about human behavior. If one desires to refine behavioral predictions in the presence of such indeterminacy, then she must develop a method for selecting among the equilibria either by (a) generating a more detailed account of behavioral norms, communication, and/or focal points for behavior or (b) refining the notion of Nash equilibrium to exclude some of the "less plausible" equilibria (particularly by exploring further the consequences of the rationality assumption). Following much of game theory literature of the last decade, the remainder of the book pushes the second agenda, with only a brief discussion of the first. This decision—while not surprising—is nonetheless unfortunate, for much of modern legal thought concerns the law's role in affecting the factors listed in (a); a more lengthy discussion of this intersection might have served the authors well, especially with respect to the Lawyers.⁹

Chapter 2 extends the analysis of complete information games beyond the "one shot" simultaneous-move environment, allowing for the fundamentals of the game to specify *sequential* rules about who can move and at what time. This added complexity facilitates the introduction of a number of dynamic elements of game theoretic reasoning, including the use of the "extensive form" rather than the normal form, backwards induction, the importance of credible threats, and the ability of players to commit to a particular action. Accounting for these phenomena in a dynamic setting

7. Moreover, so that the Lawyers are not robbed of the opportunity to become familiar with the canonical examples of normal form games, the authors present the more standard examples in the last part of the chapter.

8. For instance, recall that the coordination game has three Nash equilibria. See note 6.

9. The first agenda is related to the importance of "common knowledge," which I shall revisit in section 2.

makes it possible to introduce a significant refinement of the Nash equilibrium concept, called a subgame-perfect equilibrium. This refinement essentially excludes from consideration any Nash equilibrium that involves a player making threats that no one expects the player to carry out if called upon to do so. Like the first, this chapter also uses a number of legal examples, including debt contracting and bankruptcy, to demonstrate how the underlying structure and behavioral assumptions can be used to narrow one's prediction of the likely outcome of a game.

On balance, the opening chapters provide an introduction to the basic agendas of the book that will satisfy both the Lawyers and the Game Theorists. Most notable in this regard is the discussion of how legal rules can "matter" in a strategic environment. The authors adeptly illustrate that within complete-information games, legal rules can change behavior by altering either the payoffs to the parties or the permissible moves that the parties might make.¹⁰ This observation is of considerable value for those attempting to make contributions to the literature.

Beginning with the third chapter, the authors plunge into what is often perceived as the "cutting edge" of applied game theory: asymmetric information games, or AIGs. In their most common manifestation (and the one here), AIGs constitute a *particular type* of deviation from the common-knowledge assumption in complete-information games. No longer is the payoff structure of the parties commonly known by all parties. Now, certain players may have better information than others about their own (or even others') payoff structure. Notably, however, this deviation from common knowledge is a limited one. In fact, even though certain players have "private" information, the overall structure of the game (including the probability distribution from which the informed players learn their information) is still presumed to be common knowledge. Nevertheless, because of the informational environment, any solution concept must now *not only* account for the strategies adopted by players for each possible stage; it must *also* keep tabs on the beliefs that each player forms about others' information as the game progresses.

One central workhorse for analyzing these beliefs is what is known as a *Perfect Bayesian* equilibrium. This notion of equilibrium mandates that players must, at each stage of the game, behave in a manner that is rational, given the structure of the game, *and* given their beliefs about the "type" of player they are dealing with. The term *Bayesian* comes from a rule in probability theory called Bayes rule, which in the authors' words, "provides a means to capture formally the way rational people should update their beliefs in the wake of new information" (p. 83). The discussion touches on a number of applications for the perfect Bayesian solution concept, ranging

10. I might add to this list the notion that legal rules can also affect the number and identity of the players in a game.

from an entertaining account of the climactic scene in *The Maltese Falcon* to a more standard analysis of a commercial dispute. Unfortunately, however, *at no time in this chapter (or in the book for that matter) do the authors explain exactly what Bayes rule is, or how one who is otherwise unacquainted with it should use it.*¹¹

This omission is critical, for the Lawyers in particular. Indeed, throughout the first two chapters, the newcomer to game theory who was willing to read carefully and meticulously could work through all the numerical examples herself. In chapter 3, however, this reader is likely to feel like she has wandered into the climactic scene from a different Humphrey Bogart film—*The African Queen*—unable to propel her boat any further through a sun-drenched analytical swamp filled with underbrush, leeches, and thick Bayesian goo. In short, chapter 3 compels the Lawyer to become a spectator rather than a participant, and it threatens to leave her there for the balance of the journey.

In spite of this shortcoming, chapter 3 offers a number of helpful insights, especially in its carefully crafted analysis of how strategic behavior and prudent legal regulation depend crucially on the *nature* of the private information possessed by the parties. Particularly accessible and convincing is the analysis of *verifiable* information, in which a party can inexpensively reveal the content of her private information (sometimes called her “type”) to the other players. In such situations, parties with “good news” will, one by one, have a strong incentive to reveal this information (by say, volunteering to testify in one’s own defense in a criminal trial). By implication, then, silence may send a strong message of “bad news,” even in the presence of laws that attempt to minimize this inference (e.g., an instruction to the jury that no inference is to be made from the failure of the defendant to testify).

The authors then examine flaws in the unraveling argument, particularly when it becomes costly for the parties either to gather or to reveal

11. For the sake of clarity, Bayes rule states that the probability of an event A occurring, conditional on knowing that some other event B has occurred (or in short hand, $\Pr\{A|B\}$), can be derived from a combination of the respective “base rate” probabilities of A and B (or $\Pr\{A\}$ and $\Pr\{B\}$) and the “reverse conditional” probability that B occurs, conditional on knowing that A has occurred (or $\Pr\{B|A\}$). Explicitly, these four probabilities are related to one another according to the following formula:

$$\Pr\{A|B\} = \Pr\{B|A\} \cdot \Pr\{A\} / \Pr\{B\}$$

Thus, for example, suppose that one were attempting to use the outcome of a diagnostic test to infer whether an individual was carrying a deadly virus. Suppose further that when administered to people who are known to have the virus, the test yields a positive result 50% of the time. Moreover, among the general population, the test yields a positive result 25% of the time, and that 10% of the general population carry the virus. Using Bayes rule, it is possible to calculate the probability that an individual who has tested positive also carries the virus; the probability is equal to 20%:

$$\begin{aligned}\Pr\{\text{Virus} \mid \text{Positive Test}\} &= \Pr\{\text{Positive Test} \mid \text{Virus}\} \cdot \Pr\{\text{Virus}\} / \Pr\{\text{Positive Test}\} \\ &= (50\%) \cdot (10\%) / (25\%) \\ &= 20\%\end{aligned}$$

information to others. In such situations, mandatory disclosure laws may help induce changes in behavior, but only if courts are somehow better than players at discerning whether silence is due to the possession of bad news, or simply the possession of no information (pp. 95–97). An interesting flip side to this phenomenon, and one that the authors appear to miss, is that the ability of a court to verify what some players cannot is not always beneficial. Indeed, recent analyses of this problem demonstrate that a court's unique ability to verify information may *magnify* the incentive of the informed player to act strategically during settlement negotiations. For now, she also has private information about what a court is likely to do should a dispute ever get that far (Spier 1994; Ayres and Talley 1995).

Nevertheless, the analysis provides a smooth transition into chapter 4, which discusses the case in which it is prohibitively costly to transmit verifiable information to either the other party or the courts. The first part of this chapter is dedicated to differentiating between “separating” and “pooling” equilibria. In the former, the informed player's actions make it possible for an uninformed player to deduce something about the informed player's information (e.g., healthy individuals are willing to accept insurance plans with higher deductibles and co-payments). In the latter, all informed players behave identically, and it is impossible to deduce any information from their behavior (e.g., new car salespeople all make a first offer equal to the sticker price for the car, regardless of their reservation price).

The remainder of the chapter is devoted to even more refinements of the Nash equilibrium concept, in very much the same spirit as subgame perfection in chapter 2.¹² While important, this discussion arguably manages to miss *both* its intended audiences. The Lawyers remain where they were at the end of chapter 3, somewhat befuddled by an opaque understanding of Bayesian games. The Game Theorists are also unsatisfied; for this section *does* lift verbatim from the classic papers in this field, doing little more than attaching employment law labels to a standard signaling model.¹³ The chapter finishes much more strongly, with a nice application to contractual default rules; however, one fears that many readers may have already checked out of the analysis by this time.

After a two-chapter foray into AIGs, chapter 5 returns to the world of complete information (for the most part), discussing infinitely repeated games. Particularly relevant here is the discussion of the “folk” theorems, which formalize the long-standing intuition that repeat players, by threaten-

12. This time, however, the refinements (with names such as the “intuitive criterion,” “sequential equilibrium,” “stable equilibrium,” and “divine equilibrium”) deal not only with eliminating implausible threats off the equilibrium path, but also with the plausibility of the *beliefs* that uninformed players might harbor when they are “surprised” by a move that was not called for in the hypothesized equilibrium.

13. The example used throughout the text is the so-called beer-quiche game (Cho and Kreps 1987).

ing "punishment" in later periods for deviant behavior today, may be able to implement certain "cooperative" outcomes that one-shot players cannot. The classic legal reference for this argument (and the one in which it was originally developed) is the formation and stability of price-fixing cartels among putative competitors. Interestingly, this type of cooperative behavior can occur without (and in fact in spite of) legal regulation.

The authors adeptly point out that the effectiveness of such "dissipative" punishment schemes¹⁴ depends on the credibility of their execution, and in particular the commitment of the cartel members to abstain from *renegotiating* the terms of a punishment phase once they have entered it. Indeed, the threat of a punishment may not have much deterrent effect if players expect to abandon the punishment strategies once invoked through renegotiation. While this discussion is thoughtful and accurate, it curiously sidesteps a central theme for the book: how the law can be designed so as to influence individual behavior. Explicitly, one might ask how or whether enforcement of antitrust laws increase or decrease the viability of tacit collusion. The answer to this question is far from obvious, and recent work suggests that enforcement of antitrust laws may actually *facilitate* the stability of a cartel, since renegotiation of the punishment scheme after an episode of cheating requires collaboration among competitors that risks detection by antitrust authorities (McCutcheon 1997).

The final part of chapter 5 briefly discusses repeated games in an environment of private information. The authors re-create the striking result that cooperative behavior may emerge even from *finitely* repeated games when one introduces even a small probability that one party has a "taste" for cooperation. This line of reasoning is tantamount to relaxing the assumption that rationality is common knowledge among players. In such situations, it may be optimal for even those players who do not have a taste for cooperation to "mimic" those who do (Kreps et al. 1982). Left unanswered (and even unasked) are the more interesting questions of how players may develop such tastes, what form such tastes are likely to assume, and whether it is completely satisfactory to model "irrationality" simply as a different type of payoff structure. These are questions that others in related fields have been addressing of late (Rose 1992; Sunstein 1993).

Chapters 7 and 8 cover noncooperative bargaining games with complete and private information. Extensive attention falls on "alternating-offer" bargaining games of complete information. Particularly interesting is the authors' introduction of "exit options" within this bargaining structure (i.e., the right to "quit" the bargaining game and seek a remedy in court),

14. The threatened punishments envisioned by the folk theorems are called dissipative because they necessitate inefficient or wasteful behavior from the players themselves in periods following a deviation. In contrast, when third parties (such as courts) are available to enforce agreements, penalties for deviant behavior can take the form of fines or other immediate-transfer payments, without requiring the threat of subsequent wasteful behavior.

and their application of the game to the classic contracts case of *Peevyhouse v. Garland*, 382 P.2d 109 (Okla. 1962). The analysis is both interesting and well executed; but one nonetheless wonders why the authors did not include this discussion as part of chapter 2, which provides all the fundamental tools for the analysis. Moreover, placing this material earlier might increase the chances of a careful read by Lawyers who might become discouraged upon entering the AIG chapters. In extending models of bargaining to asymmetric information environments, chapter 8 benefits from a wealth of law and economics literature on how various procedural rules (such as fee shifting, multi-stage trials, and discovery rules) affect negotiation. Once again, though, while cleanly executed, this section might be more prudently placed right after chapter 3.

Overall, the breadth and depth of analysis in this book are quite impressive. However, for one reason or another, other deserving and emergent areas of game theory with intriguing applications to law failed to make the cut. For instance, the book might have given a more extensive analysis of how well the predictions of game theoretic models fare in experimental settings. Though research in this area has proliferated in recent years (much of it inconsistent with standard solution concepts), very little of it appears in the text. Additional space might have been devoted to the problem of multiple equilibria and how players might (or might not) “learn” to play a particular equilibrium of a game. As mentioned above, the book might also have examined more thoroughly the appropriate way to model “perturbations” in player rationality and their consequences. Along these same lines, the book offers no analysis of so-called evolutionary models of strategic behavior, which explore the consequences of abandoning rationality altogether, and replacing it with a more Darwinian assumption that players behave in a payoff-enhancing way are more likely to thrive within a population (Weibull 1995).

Many of these omissions are perhaps justifiable on one of two grounds. First, as with any piece of writing, space considerations are *always* binding. And second, when faced with these choices, the authors chose to retain subjects in which they had the most expertise. Nonetheless, as I have mentioned previously, one omission from ongoing research is directly relevant to the arguments the book does present: a more systematic discussion of the effects of relaxing the common-knowledge assumption. It is to that critique that I now turn.

II. COMMON KNOWLEDGE AND LEGAL DOCTRINE

Central to virtually every applied model of game theory is an embedded assumption (in one manifestation or another) about “common knowledge.” As noted above, a statement or fact is said to be common knowledge

if every player involved in the game knows it, every player knows that every player knows it, and so on. Games of complete information (such as those studied in chapters 1 and 2 of the book) are rather special, for every important feature of the game is assumed to be common knowledge among the players. In contrast, typical AIG's (asymmetric information games, such as those studied in Chapters 3, 4, and 8) assume that the players are not completely informed about one another's payoffs. AIGs represent *one way* in which common knowledge assumptions can be relaxed (through the payoffs of the players), but they are likewise somewhat special. For even in a typical AIG, the *fact that there is an asymmetry* is itself common knowledge among the players, as are the players' *ex ante* beliefs about the forces that generate the privately informed player's information. However, such breakdowns in common knowledge might take a number of other manifestations, including knowledge about the structure or rules of the game, the rationality of the players,¹⁵ the number of players, or the effectiveness of communication between players.

It is somewhat perplexing that the authors chose not to devote part of the text to a broader, more systematic analysis of common knowledge.¹⁶ Indeed, while a principal normative thrust of game theory as applied to law involves designing rules and sanctions that influence the behavior of strategic players, it may be equally important to understand how (or whether) the

15. A classic example is often recounted to game theory students (in various degrees of taste and political sensitivity) to underscore the power of the "common knowledge of rationality" assumption, called the "teacher and the hats" allegory. It goes something like this: a teacher places hats on the heads of each of three students. No student can discern the color of her own hat, but she can observe the color of the other players' hats. In addition to the hats, the teacher gives the group one more piece of information: each hat is *either* red or green, and there is *at least* one red hat in the lot (but possibly more). The teacher offers a high grade for the first student who can both (1) state correctly the color of her hat and (2) prove deductively that she must be right. If a student fails in either respect, she will instead be given a failing grade.

Upon giving the three students this information, the teacher proceeds to ask each student *ad seriatim* if she knows the color of her hat. Student 1 looks around nervously and says no. Student 2 examines her fellow students in a similar matter, and also says no. On hearing their responses, student 3 states that she knows her hat is red, and proceeds to explain why. Upon hearing her explanation, the teacher gives her an A+. It turns out that the student's explanation is based squarely within an assumption of common knowledge of rationality. For the content of her response, see Geanakopoulous 1992.

16. In fact, the term *common knowledge* does not make its first appearance (and a cameo one at that) until chapter 3—about 40% of the way through the text (p. 108). This is somewhat later than other leading game theory texts (Fudenberg and Tirole 1991, 4; Myerson 1991, 62 [the first 36 pages are dedicated to mathematical preliminaries]). It should be noted, however, that this criticism of the book should not be understood to be an accusation that the authors "missed" the importance of common knowledge. Indeed, they are keenly aware of the strength of the assumption (e.g., p. 165 [noting that the assumption is a "strong one," is "quite powerful," and that care must be taken in applying it]; p. 108 [noting the strength of the common knowledge assumption, but justifying it on the basis that it is a modeling tool, like any other in game theory]). Rather, my criticism is that the legal environment is a particularly ripe one to explore the consequences of relaxing this assumption at greater length than do the authors.

existence, content, and applicability of such rules become commonly known among the players. This question may be of particular importance to those interested in exploring the role played by law in influencing and/or facilitating the formation of collective understanding and behavioral norms among groups (or, perhaps, in mitigating the consequences when collective accord cannot be achieved) (Cooter 1997).

This section explores an example illustrating how common knowledge defects in pre-play communication can affect behavior, and how legal rules might optimally respond. Explicitly, I will examine at length another type of coordination game, this one involving contract formation through "signals" sent between two parties. The example also helps demonstrate that even when law cannot undo the common-knowledge deficiencies, legal rules may well play a mitigating role. While this example is rather limited in its immediate doctrinal implications, the intuition it conveys may be generalized to encompass broader questions of how laws and common knowledge can (or should) interact.

A. Motivating Case: *Crook v. Cowan*

To motivate the discussion, consider the classic nineteenth-century case of *Crook v. Cowan*, 64 N.C. 743 (1870), a case taught in a number of first-year contracts classes. In *Crook*, the plaintiff seller had held itself out to the general public as a purveyor of handmade carpets. The defendant buyer ordered two carpets from plaintiff through a correspondence, asking the seller either to ship them COD or to advise the buyer of the price in order to arrange prepayment. After a few days, the defendant, having received no reply, sent an additional telegram, inquiring why the seller had not confirmed receipt of the initial order. After receiving no reply once again, the defendant apparently procured his carpets elsewhere. Only some weeks later did the defendant learn that the seller *had* in fact received the first order, and had produced and sent the carpets COD. When the defendant refused to pay for the carpets, the seller sued for breach of contract.

In affirming a holding below for the seller, the North Carolina Supreme Court found that the parties' actions constituted a valid offer and acceptance of a contract for carpets. The court gave two alternative interpretations of its holding, each having the same result (i.e., the plaintiff wins). Under the first interpretation, the seller's act of holding itself out to the public as a merchant constituted an offer, and the buyer accepted this offer through its order. Under the alternative interpretation, the original letter from the buyer constituted an offer authorizing acceptance through performance, and the seller's act of constructing and shipping the carpets within a reasonable time constituted just such an acceptance.

Regardless of the interpretation, it is interesting to note that the court adhered rather stingily to a "one offer, one acceptance" rule for contract formation. In an oft-cited portion of the holding, the court rejected the defendant's argument that it would be reasonable to require additional confirmations before contractual rights and duties attach:

[I]f an offer and acceptance — an unconditional and specific order, and an exact fulfillment, as in this case, does not complete the contract, how would it be possible to complete a contract by mail? A sends an unconditional order to B, and instead of B's filling the order, he writes back that he accepts the order and will fill it, but in the meantime, A may have changed his mind, and lest he has, he must write back to B and so on, for ever. . . . [If we were to require such serial re-confirmations], no contract could ever be completed by post. For if the defendant was not bound by his offer, when accepted by the plaintiff, until the answer was received, then the plaintiff ought not to be bound until after he had received the notification that the defendant had received his answer and assented to it. And so it might go on *ad infinitum*. (747)

Law students frequently disagree with this holding and, in particular, with the court's interpretation of the defendant's advertisement as an "offer." When teaching the case, however, I redirect questions to this passage, asking students to evaluate the general proposition that the necessary and sufficient communications for creation of a contract consist of one offer followed by one acceptance. Why, for instance, does this court adhere to *this* particular rule, rather than, say, one requiring only an offer; or perhaps one that requires an offer, an acceptance, *and* one additional confirmation by the initial offeror? Neither alternative rule, I point out, would lead to the infinite regress that the opinion predicts any more than would a rule requiring two communications.

A possible answer to this question, as well as an interesting analysis of the reasoning of the court, is provided by further investigation of the relationship between contract formation and common knowledge. The fact pattern of *Crook v. Cowan* bears a striking resemblance to a well-known game of common knowledge, attributable originally to Ariel Rubinstein, and popularly called the "Electronic Mail" game (1989). In the pages below, I will examine an example based on the facts of *Crook* that resembles Rubinstein's principal argument: When the success of pre-play communication between the contracting players is not common knowledge (even when it is "almost" common knowledge), their efforts to coordinate with one another (i.e., enter a contract) may fail. In such an environment, legal rules on offer and acceptance can help mitigate this contracting problem, *not* by "resolving" the problem of common knowledge failure, but by mitigating its consequences.

B. Common Knowledge and Pre-play Communication: An Example

Imagine a potential buyer (player 1) and a potential seller (player 2) of some specialty good or service who communicate with each other through an indirect medium such as the post or electronic mail. The players are *potential* rather than *actual* buyers and sellers because it is not clear whether there are gains from trade in making a purchase. Explicitly, buyers come in one of two flavors: low valuers (who value the good at \$0), or high valuers (who value the good at \$300). Suppose that low valuers constitute 60% of the buyer population and high valuers constitute the remaining 40%. In order to take delivery of the good, the buyer must expend (in addition to the price) \$100 on preparatory expenses—which does not go to the seller. Because of these preparatory expenses, even a high-valuing buyer would be willing to pay *at most* a price of \$200 for the object, which represents her net value after incurring preparatory costs.

Sellers can also come in one of two types: low-cost producers, who can produce and deliver the object for \$100; and high-cost producers, who can produce and deliver the object for no less than \$500. Suppose that low-cost sellers constitute 40% of the seller population and high-cost sellers constitute the remaining 60%.¹⁷

The above framework suggests that a transaction would only increase aggregate economic welfare when the buyer is a high valuer and the seller is a low-cost producer, for only then does the buyer's net willingness to pay (\$200) exceed the seller's cost of production (\$100). For simplicity, suppose that in such a situation, the seller and buyer will reach a price of \$150, effectively splitting the difference between their valuations.¹⁸ In all other situations (high-cost seller with either type of buyer, or low-valuation buyer with either type of seller), the transaction would give rise to a net social loss.¹⁹

One important aspect of the scenario just described is that (unlike many other examples of contracting with private information), neither the buyer nor the seller has any strategic reason to hide her knowledge from the other (i.e., there is no room for either player to use her private information

17. I shall assume that the buyer's and seller's respective valuations are independent of one another, and are verifiable in court.

18. Though the price might often depend on the underlying legal rule governing offer and acceptance (Katz 1990), in this example the dependence is not crucial, since my focus is on the communication between the parties and not the terms. Moreover, as the example will demonstrate, the common knowledge problem is still present even when (as in this case) neither party has a strategic reason to misrepresent her private information.

19. In making this statement, I am assuming implicitly that the buyer's preparation costs and the seller's production costs are irreversible. It is not possible for either party to recoup any "salvage" value from her expenditures. Introducing partial reversibility of investments does not change the qualitative results, so long as some component of the investment is lost if the other side does not show up to consummate the transaction.

in securing a better price for the object). To illustrate this point, if a low-cost seller were successful in "bluffing" the buyer into believing that the seller faces high costs, the successful bluff would not cause the buyer to make concessions on price; rather, it would cause the buyer simply to walk away from the transaction, believing that she could pay no price that would make production worthwhile to a high-cost seller. Analogously, the high-valuing buyer has no incentive to represent herself as placing a low value on the good. On the contrary, in this setting both the low-cost seller and the high-valuing buyer have strong incentives to signal that information to the other party.

Unfortunately, however, as in *Crook v. Cowan*, the parties cannot communicate face-to-face. Instead, they must use a technology for indirect communication, such as the post or electronic mail (email). The key problem posed by this indirect mode of communication is that email, while generally reliable, can sporadically lose messages. Suppose that each time a player sends a message to the other player, there is a 10% chance that it will get lost somewhere in cyberspace, never to return. This risk creates problems for a low-cost seller and high-valuing buyer who wish to convey invitational signals to one another. Thus, an efficiency objective for law within this strategic setting is to *structure rules of offer and acceptance so as to enhance the chance that high-valuing buyers and low-cost sellers will recognize one another and trade, while simultaneously minimizing the chances that other buyer-seller combinations will incur performance costs when there are no gains from trade available.*

For the sake of exposition, assume that communications between the potential seller and potential buyer are governed by the following simple protocol: If the seller, who is the first mover, has a high cost (and is therefore uninterested in a transaction with either type of buyer), his computer is programmed to do nothing; conversely, if the seller has a low cost (and is therefore interested), his computer is programmed to *automatically* transmit an email message to the buyer indicating his willingness to make a transaction at \$150, and it will *automatically* confirm any and all positive replies received from the buyer. If the buyer—who is the second mover—has a low valuation (and is therefore uninterested in a transaction with either type of seller), her computer likewise does nothing (even if an initial message from the seller arrives); conversely, if the buyer has a high valuation (and is therefore interested), her computer will *automatically* send positive replies to any and all invitational messages received from the seller (initial or subsequent), affirming a mutual interest in transacting.²⁰

20. It is possible to hypothesize more elaborate protocols for communication (such as the parties continue to send confirmations for as long as the offer-acceptance rule requires), but this one suffices to convey the primary intuition behind the Rubinstein result.

Should either player be uninterested in a transaction, the chain of communications will be quite short-lived, consisting of either no messages, or of one initial message from the seller, which goes unanswered. On the other hand, if both players are interested, then the chain of communications will vary in length, depending on when the first communication between the players becomes lost in cyberspace, thereby breaking the chain. (Recall the stipulation that every time a communication is sent from one party to the other, there is a 10% chance it will become lost. Because of this constant rate of loss, the chain of communications must end after some number of exchanged messages.)

Using this framework, it is possible to analyze the facts in *Crook v. Cowan* and to suppose that the court had issued a different holding (i.e., that the number of communications necessary and sufficient to create an enforceable contract was something other than two). To begin this analysis, let us start at an (admittedly absurd) extreme, supposing that the court had, contrary to its holding, decided that the seller's confirmation was necessary, as was a reconfirmation by the buyer, as was a *re-reconfirmation* by the seller, *ad infinitum*. In other words, suppose the court had adopted a rule stating that an enforceable contract could not come into existence if at any point of correspondences the chain of confirmations between the buyer and seller's email accounts was ever broken. For explanatory ease, I shall call this the *infinite communication* (or $N = \infty$) rule for contract formation.

An immediate implication of such a rule, as the court noticed in *Crook*, is that an enforceable contract can never be created through delayed serial correspondence. Indeed, because each message has at least a 10% chance of being waylaid, a confirmation *will* get lost at some point, and one party will not be able to reconfirm. Hence, under this rule, the parties cannot establish an enforceable contract through email communication.

Nonetheless, does the lack of enforceability necessarily imply that a high-valuing buyer and a low-cost seller would fail to coordinate? Not necessarily. One could easily imagine that somewhere down the line (say, for instance, after the seller had sent 27 emails to the buyer offering a trade, and the buyer had sent 26 acceptances and confirmations), the legal consequences of the communications would be irrelevant. At this point, one might argue, each party would have gained enough information about the other's willingness to trade that they would each show up at the time and place for delivery, expecting to engage in a spot transaction (notwithstanding the absence of an enforceable contract).

To analyze this proposition, it is helpful to specify the payoffs for each party. The normal-form matrix representations of the game (similar to that illustrated in the previous section) specifying the players, strategies, and payoffs provide a convenient mechanism for doing so. Because there are four buyer-seller combinations, however, it is necessary to consider the nor-

mal form of the game for each possible combination. In each of these representations, once the sequence of email transmissions (if any) has taken place, the players must *simultaneously* decide whether to go about other business (i.e., stay home) or prepare for spot transaction at a price of \$150 (i.e., show up).

Consider first the situation in which there are gains from trade—that is, where the buyer has a high value and the seller faces a low cost (a combination that occurs 16% of the time²¹). The normal form for this combination is as follows:

		Seller	
		Stay home	Show up
Buyer	Stay home	0, 0	0, -100
	Show up	-100, 0	50, 50

The above matrix embodies all the possible payoffs of the players. If both stay home (the northwest cell), then neither incurs preparatory costs, and their payoffs are both zero. Conversely, if both show up and transact at a price of \$150 (the southeast cell), then each will net (after accounting for costs) \$50 in surplus or profits. However, if only one party shows up and the other stays home (the southwest and northeast cells), then the party showing up must surrender \$100 in preparatory costs (getting nothing in return), while the party that stays at home receives a payoff of zero.

Now consider the payoffs when the buyer is a high valuer and the seller faces a high cost (a combination occurring 24% of the time²²). The payoffs in this situation differ from the previous one only if the seller shows up, incurring \$500 in production costs rather than \$100. Thus, the normal form for this situation is as follows:

		Seller	
		Stay home	Show up
Buyer	Stay home	0, 0	0, -500
	Show up	-100, 0	50, -350

21. To see this, simply note that the buyer is a high valuer with probability .4, and the seller faces a low cost also with probability .4. Since the buyer's valuation and the seller's cost are independent, the joint probability of a high-valuing buyer with a low-valuing seller is $(0.4) \times (0.4) = 0.16 = 16\%$.

22. $(0.6) \times (0.4) = 0.24 = 24\%$.

The payoffs change in the opposite manner if the buyer is a low valuer and the seller faces a low cost (which also occurs 24% of the time):

		Seller	
		Stay home	Show up
Buyer	Stay home	0, 0	0, -100
	Show up	-100, 0	-250, 50

Finally, the payoffs below reflect the situation in which the buyer is a low valuer and the seller faces a high cost (a combination occurring 36% of the time²³):

		Seller	
		Stay home	Show up
Buyer	Stay home	0, 0	0, -500
	Show up	-100, 0	-250, -350

Given these four possible combinations of buyer and seller valuations, does communication between the parties facilitate coordinated trade when and only when there are gains from such a trade? To answer this question, consider first the benchmark question of how the parties would behave under a non-enforcement rule if they were unable to communicate with one another at all. It turns out that there is only one Nash equilibrium of this game: both the buyer and the seller stay home. The intuition behind this equilibrium is as follows. Since neither party has knowledge of the other's cost/value other than the original probabilities, each has to act in a way that is profit maximizing *on average*. Consider a buyer who has a high valuation (and thus has the potential to gain from trade). Is this buyer willing to spend the \$100 in preparation costs, hoping that the seller will be a low-cost seller and show up? Probably not: For the buyer knows that if the seller faced a high cost (which occurs, remember, 60% of the time), then such a seller would never rationally show, even with a willing buyer waiting at the market. Thus, three times out of five (at least), the buyer would show up at the place for delivery to be sadly disappointed, having sunk \$100 in preparatory costs.

Even so, couldn't the buyer gamble that the seller faces a low cost and will therefore show up? Perhaps, but this gamble imposes a large downside

23. $(0.6) \times (0.6) = 0.36 = 36\%$.

(– \$100) that occurs at least 60% of the time, and a more modest upside (\$50) that occurs at most 40% of the time.²⁴ Thus, this gamble loses \$40 on average, and it would not be worth pursuing for a rational buyer (unless she were an extreme risk taker). The rational buyer will therefore stay home, regardless of whether she is a high or low valuer. An identical argument applies to the seller: even a low-cost seller would probably not be willing to take the gamble that a high-valuing buyer would show up. Thus, the only equilibrium of the “no communication” game is for both players to stay home, regardless of their respective types.

Moving away from the “no-communication” benchmark, we can now inquire into the effects of the cascade of emails hypothesized above. As has already been mentioned, allowing such communication—even without a legal rule to back it up—might be sufficient for a buyer-high; seller-low combination to coordinate with one another. In keeping with the convention that the seller starts the series of emails (if any), one might ask how the behavioral strategies of the players would change as one increases incrementally the number of successful communications between the parties.

To start the process, suppose that the seller were a low-cost type, and therefore sends an invitational email to the potential buyer. What would happen if this first email were lost somewhere in cyberspace (and thus the buyer never received it, never sent an answer, and so on)? To answer this question, one must inquire into the buyer’s beliefs under such a contingency. Because the buyer has heard nothing, she must believe that one of two things has happened. Either (1) the seller is a high-cost type, and thus never sent a message or (2) the seller is a low-cost type, and sent a message that was somehow lost. (The buyer has no way to know which of the two possibilities is correct.) In order to determine whether she should “gamble” on the second hypothesis rather than the first, the buyer might attempt to assign probabilities on the validity of hypothesis 1 versus hypothesis 2. Using Bayes rule (described above), the buyer can infer that the probability the seller sent no message is approximately 0.94, while the probability that a message was sent but lost is approximately 0.06.²⁵ Because these inferences

24. I use the terms “at least” and “at most” to reflect the fact that the described gamble is a best-case scenario. For the buyer knows that a high-cost seller will never show up, but there may even be a chance that a low-cost seller will fail to show up as well (and thus the possibility that the downside of the gamble occurs more than 60% of the time).

25. The derivation of these estimates is as follows: The buyer is attempting to deduce the probability that the seller faces a high cost given that she received no communication. Mathematically, this is:

$$\Pr\{\text{High Cost} \mid \text{No Signal}\} = \Pr\{\text{No Signal} \mid \text{High Cost}\} \cdot \Pr\{\text{High Cost}\} / \Pr\{\text{No Signal}\}$$

The term in the denominator (i.e., the probability that the buyer never receives a signal) is reflected by the sum of the probability that the seller is a high-cost type (0.6) plus the probability the seller is a low-cost type but that the signal was lost (or $0.4 \times 0.10 = .04$). Thus, the ex ante probability of not receiving a signal at all is equal to 0.64. The two terms of the numerator are easy to calculate. First, the probability that the seller is a high-cost type is

so overwhelmingly support the former hypothesis, a rational buyer would likely be unwilling to "gamble" on the latter.²⁶ Consequently, if the first signal fails to make it to the buyer, the buyer will stay home regardless of whether she is a high or low valuer.

Moving incrementally, suppose now that the first message successfully reaches the buyer, and that the buyer places a high value on the transaction (thereby answering immediately), *but* that the buyer's response is lost on the way back to the seller. One now must inquire into the *seller's* beliefs at this juncture, based on her observation that she sent an initial invitation but never received an answer. The seller must believe one of three things has happened. Either (1) the buyer failed to receive the seller's original missive, and never had the opportunity to respond; (2) the buyer did receive the message, but was uninterested and declined to reply; or (3) the buyer received the seller's original message, replied favorably, and her reply was lost. Under hypothesis 1 as we have seen above, the seller knows that the buyer will stay at home—for failing to receive the initial message makes the buyer nearly certain that the seller is not interested. Under hypothesis 2 the buyer will also stay at home—for the buyer has no interest in a trade. Only under hypothesis 3 is it even possible that the buyer will show up. Thus, the operative question for the seller is whether she should gamble on hypothesis 3 being true rather than hypotheses 1 or 2. Using Bayes rule again, it is possible to demonstrate that such a gamble would pay off only 5% of the time.²⁷ Consequently, if a seller who has sent one message fails to receive a return message from the buyer, the seller would find it optimal to stay home.

Although the direction of this argument is now becoming clear, it will prove useful to work through one more iteration before summing up. Suppose now that (1) the willing seller's original message goes through; that (2) it meets a willing buyer who replies favorably; that (3) the buyer's favorable reply successfully reaches the seller; and (4) that the seller promptly confirms, but the confirmation is lost. This is a crucial juncture of the game because it marks a point at which each party has received at least one notice of her counterpart's interest in a trade. One might conjecture, then, that at least by this point, the parties will show up voluntarily at the time and place for delivery regardless of whether they have a legally binding agreement.

This conjecture, as it turns out, is off course. Although the players have *mutual knowledge* of existing gains from trade at this juncture (i.e., each party knows of her counterpart's interest), the gains from trade are not *common knowledge*, and anything short of common knowledge is insufficient for

simply 0.6. The probability that a high-cost seller does not send a signal is simply 1.0. Thus, the right hand side of the above expression is $(1) \times (0.6)/(0.64) = 0.94$.

26. In fact, the "gamble" of showing up (i.e., performing) is even less attractive than if no communication were allowed.

27. Explicitly, the probability that hypothesis 1 is true is 0.15 (or 15%); that hypothesis 2 is true, 0.80 (or 80%); and that hypothesis 3 is true, 0.05 (or 5%).

the parties to overcome the coordination problem they face. Here's why: Suppose we put ourselves in the shoes of the high-valuing buyer who has received the original communiqué from the seller, has replied favorably, but never received a confirmation from the seller. One possibility that this buyer can rule out is that the seller is not interested; for the original message from the seller signaled unambiguously that the seller has a low cost. However, the buyer cannot rule out two remaining hypotheses: (1) that the seller never received the buyer's favorable reply; or (2) that the seller received the buyer's reply and confirmed, but that the confirmation itself was lost.

This observation is the key. For if hypothesis 1 were true, as demonstrated above, the seller will not show up: for failure to receive a reply from the buyer signals to the seller (with overwhelming probability) that the buyer is not interested. Therefore, only with hypothesis 2 is there *even a chance* that the seller will show up. The buyer is put once again in a position of deciding whether to stay home, or to gamble that hypothesis 2 is true. And once again, this gamble hinges crucially on the probabilities assigned to each hypothesis. From Bayes rule, this gamble will pay off only 47% of the time and will yield an average payoff of $-\$29.50$ at most, quite obviously less than the $\$0$ payoff the buyer could obtain (with certainty) by simply staying home.²⁸ This was the precise position of Mr. Cowan in *Crook v. Cowan* (although slightly different prices and procedures were involved).

Consequently, when there is no enforceable contract, if a high-valuing buyer fails to receive an additional confirmation from the seller, it is not in her interests to show up for the spot transaction. This result holds *even though* the parties are mutually aware of each other's interest. The problem that causes the spot market to fail is one that emanates from an imbalance of common knowledge between the parties. For even though the buyer becomes aware of the seller's interest through the first communiqué, she can not be sure whether the seller has also been made aware of the buyer's interest. Alternatively, even though the seller becomes aware of the buyer's interest, the seller cannot be sure whether the buyer *knows* that the seller is aware of her interest. With this imbalance in "knowing that your opponent knows that you know . . . etc.," both parties stay home, even when they are mutually aware that they could gain from the trade.

Moreover, this iterative reasoning can be repeated, *no matter how many times*, with no change in result. At some point, the confirmations and re-confirmations *must* cease: say, for instance, when the seller has sent 62 messages and confirmations, and the buyer has sent 61 in return. Even in

28. Explicitly, probability that hypothesis 1 is true is $0.10 / (.10 + .90 \times .10) = 0.10 / 0.19 = 0.53$, and thus the complementary probability that hypothesis 2 is true is approximately equal to 0.47 (or 47%). Using this probability, the expected payoff of gambling is at most $(0.53) \times (-100) + (0.47) \times (50) = -29.50$.

this instance, the lack of common knowledge of the transmissions gives rise to just one Nash equilibrium: Both parties stay home, even when they could gain from trade.

This market-failure result is undesirable under numerous normative criteria. From an economic efficiency perspective, where an important goal of offer and acceptance doctrine is to promote the consummation of value-increasing trades, the infinite-communication rule performs in a decidedly poor fashion: it ensures that no trades—be they value increasing or value decreasing—ever come to fruition. Alternatively, under a normative criterion that ranks distributional equity above efficiency, the market-failure result is also unattractive. While the failure to trade ensures that both parties receive a payoff of zero, it is possible to maintain equal distributions while increasing all players' welfare by encouraging even some trades when there are potential gains.

Given the rather poor performance of the infinite-communication rule, one might inquire whether "finite" communication rules do any better. In other words, the legal rule might instruct courts to find that a contract exists when at least N serial communications are sent by the parties, where $N \geq 0$.²⁹ Once an enforceable contract for trade exists, then a breaching party whose breach makes her counterpart worse off becomes liable for expectation damages that are sufficiently large to put the aggrieved party in the same position as if the contract had been fully performed.³⁰

Imposing contractual liability after N communications effectively changes the normal-form payoffs for the parties. When the buyer is a high valuer and the seller faces a low cost, the parties' payoffs *under an enforceable contract* are now as follows:

		Seller	
		Stay home	Show up
Buyer	Stay home	0, 0	-150, 50
	Show up	50, -150	50, 50

When the buyer is a high valuer and the seller faces a high cost, the payoffs are as follows:

29. It would be possible to designate N based not on the number of serial communications sent, but rather the number received. Because the communication protocol calls for a player's computer to send automatic responses to each message, a rule predicated on N sent messages would correspond identically with one that required $(N-1)$ received messages.

30. Thus, when an enforceable contract exists, a low-cost seller who has performed may recover the difference between the contract price (\$150) and the salvaged value of her performance (\$0), for a damages award of \$150. The same is true for the high-cost seller. A high-valuing buyer may recover the difference between the value she places on the good (\$300) less the contract price (\$150), for a damages award of \$150. A low-valuing buyer, however, places no value on the good, and as such can recover nothing in expectation damages.

*Buyer high; Seller low***Seller***Stay home**Show up*

Buyer	<i>Stay home</i>	0, 0	-150, -350
	<i>Show up</i>	50, -150	50, -350

When the buyer is a low valuer and the seller faces a low cost, the payoffs are as follows:

*Buyer high; Seller low***Seller***Stay home**Show up*

Buyer	<i>Stay home</i>	0, 0	-150, 50
	<i>Show up</i>	-100, 0	-250, 50

Finally, when the buyer is a low valuer and the seller faces a high cost, the payoffs look like this:

*Buyer high; Seller low***Seller***Stay home**Show up*

Buyer	<i>Stay home</i>	0, 0	-150, -350
	<i>Show up</i>	-100, 0	-250, -350

Notice that once the right to expectation damages obtains, the high-valuing buyer and the low-cost seller will always show up at the prescribed time and date. In such a case, these parties are *fully insured* against the reliance-cost losses they incur by preparing to perform.

Consider first a rule of $N = 0$. In this case nothing has to occur before an enforceable contract can exist. As noted above, both the high-valuing buyer and the low-cost seller will decide to show up, with the others remaining at home and risking the chance that they will have to pay damages. While this rule seems nearly as absurd as the infinite-communication rule, it does produce mutual gains whenever both parties are well-suited to a transaction. Thus, when a high-valuing buyer meets a low-cost seller (which occurs 16% of the time), the $N = 0$ rule produces a social welfare gain of \$100. On the other hand, if one party is not well suited to transact, the rule does little to deter the well-suited party from sinking the preparatory costs of contracting despite the eventual nonconsummation of the trade. Consequently, for every other buyer-seller combination, the $N = 0$ produces net

losses. All tolled, the $N = 0$ rule creates an expected welfare loss of \$32,³¹ making it even worse from an efficiency standpoint than the infinite-communication rule studied above.

Under a single-offer rule, or $N = 1$ (also known as “acceptance by silence”), the potential seller can create a binding contract simply by sending an initial message. No response is required, nor is the buyer off the hook even when the initial message is lost. As noted above, the seller (who is the first mover) will only send the message if she has a low cost of production. (Indeed, a high-cost seller who communicates a willingness to transact immediately risks putting herself in the hole by \$150). Moreover, on sending the message, the low-cost seller will have created a contractual right and will therefore certainly show up.

For the low-valuation buyer, it is irrelevant whether an enforceable contract exists, for she will choose to stay home regardless (risking damages if the seller has a low cost). For the high-valuation buyer, in contrast, things are a bit more complex. If she receives the first message, then she knows that a contractual right exists, and she will show up at the time and place for performance. However, if she does not receive the first message, it is because either the seller is a high-cost type, and thus never sent a message; or the seller is a low-cost type who sent a message, thereby creating an enforceable contract, but the message was somehow lost. Using an analysis similar to that above, it is possible to demonstrate that “gambling” on the second hypothesis is not an optimal decision for the high-valuing buyer, and thus she will show up only if she first receives an invitational initial message from the seller.³²

The $N = 1$ rule allows the high-cost sellers to “opt out” of inefficient contracting by failing to send an initial message. However, because the rule fails to deter the expenditure of production costs by the seller when there are no gains from trade, the $N = 1$ rule still gives rise to an expected welfare loss of \$11.20.³³ Thus, while the $N = 1$ rule fares better than the $N = 0$ rule

31. With a high-buyer/high-seller combination (which occurs 24% of the time), the $N = 0$ rule produces a social welfare loss of \$100, due to the wasteful preparation expenditures by the high-value buyer. With a low-buyer/low-seller combination (which occurs 24% of the time), the $N = 0$ rule also produces a social welfare loss of \$100, due to wasteful production expenditures by the low-cost seller. Finally, with a low-buyer/high-seller combination (which occurs 36% of the time), the rule produces neither welfare gains nor losses.

32. As demonstrated above, the buyer can infer from the lack of message that hypothesis 1 is true with a probability of approximately 0.94 (or 94% of the time), while hypothesis 2 is true with a probability of 0.06 (or 6% of the time). Although the costs are now greater for an incorrect inference that hypothesis 1 is correct (since the buyer runs the risk of having to pay damages to the seller if the seller was a low cost type and sent a message), the odds are still large enough to dissuade the high-valuing buyer from showing up if she does not first receive a message.

33. With a high-buyer/low-seller combination (which occurs 16% of the time), the $N = 1$ rule produces a successful contract and social gain of \$100 90% of the time, when the seller's initial message makes it to the buyer; for the remaining 10% of the time, the rule produces a social loss of \$100, when the seller's message never reaches the low cost buyer, but the seller

on efficiency grounds, it still fares far worse than the infinite-communication rule.

Consider now the rule coinciding with the usual common law offer-and-acceptance doctrine: $N = 2$. Under such a rule, contractual liability obtains only if the seller successfully sends an initial message to the buyer, and the buyer sends a positive response. As was the case with the $N = 1$ rule, under an $N = 2$ rule, a high-cost seller will never send the initial message, and thus, an enforceable contract cannot be created in such a contingency, due to self-selection among the sellers. Additionally, however, the $N = 2$ rule enables *buyers* to self-select in a similar fashion. Should a low-valuing buyer receive an initial message from a seller, the buyer can simply decline to reply, and no liability (or performance) will ensue. Consequently, under the $N = 2$ rule, a contract can only come into existence when *both* the seller and the buyer are willing trading partners.

By enabling *both* parties to self-select in this fashion, the $N = 2$ rule creates considerable efficiency gains. Indeed, much of the inefficiency of the $N = 0$ and $N = 1$ rules emanated from preparatory costs incurred by one party when the other had no intention of performing. Under the $N = 2$ rule, contractual formation is no longer "overinclusive." The $N = 2$ rule achieves this result, however, by imposing a risk that the additional communication required to create an enforceable contract will be lost, thereby frustrating any trades.³⁴ In this particular example, however, this risk is small compared to the benefits of facilitating pre-play self-selection by both unwilling buyers and unwilling sellers. All tolled, the $N = 2$ rule creates an expected welfare

incurs production costs anyway. With a high-buyer/high-seller combination (which occurs 24% of the time), the $N = 1$ rule produces no social gains nor losses, since the seller never sends the initial message. With a low-buyer/low-seller combination (which occurs 24% of the time), the $N = 1$ rule also produces a social welfare loss of \$100, due to wasteful production expenditures by the low-cost seller. With a low-buyer/high-seller combination (which occurs 36% of the time), the rule produces neither welfare gains nor losses, as once again the seller never sends the initial message.

34. Two distinct risks are actually imposed by the $N = 2$ rule. To see this, suppose that the seller faces a low cost, and the buyer attaches a high value. In such a situation, the seller will send an initial message. If a message reaches the buyer, then the buyer will reply favorably. Increasing N from 1 to 2 creates two distinct risks (rather than one) that a critical message will be lost. First, should the initial message fail to reach the high-valuing buyer, then the situation is very much like a lost initial message under the $N = 1$ rule: the buyer will infer (once again using Bayes rule) that the seller is most likely not interested, and thus, the buyer will not show up for the transaction. This time, however, by not showing up, the buyer no longer risks "hidden" liability for the lost initial message.

The second communication hazard under the $N = 2$ rule occurs when the original message successfully arrives, but the buyer's reply is lost. Once again, the beliefs of the seller become critical. For a seller who sends an initial message but receives no reply cannot rule out three hypotheses: (1) the buyer never received the seller's original message; (2) the buyer did receive the seller's initial message, but did not reply; or (3) the buyer received the seller's original message and replied favorably, but her reply was lost. Under the previous (nearly identical) analysis, the probabilities of these respective hypotheses are 0.15 (or 15%), 0.80 (or 80%), and 0.05 (or 5%). And as before, a seller in such a situation will also choose to stay home, this time risking a small (5%) chance of contractual liability.

gain of \$11.52.³⁵ Note, the $N = 2$ rule is the only rule analyzed thus far that creates strictly positive expected welfare gains.

From here it is possible to conduct same analysis for $N = 3$, but the same qualitative results hold: Namely, no contract ever forms unless both parties are willing trading partners, since low-value buyers and high-cost sellers have already disqualified themselves (by refusing to send a signal) by the second stage. Further, as with the previous case, when there are gains from trade, both parties will perform *only when* all required signals have been sent or received (three of them in this case). The aggregate level of social welfare under the $N = 3$ rule is equal to \$10.36.

Note that moving to $N = 3$ actually *reduces* (albeit modestly) the expected level of social welfare below that of $N = 2$.³⁶ The reason for this modest decline lies in considering the *crucial selection function* that the first two stages of communication alone play. When formation doctrine necessitates at least one communication ($N \geq 1$), unwilling sellers may protect themselves against future liability by failing to send a signal. When formation doctrine necessitates a second signal ($N \geq 2$), unwilling *buyers* may also remove themselves from the system by failing to respond. In each of these two cases, the marginal *benefits* from allowing the inefficient parties to remove themselves outweigh the marginal costs imposed by an increased risk of communication loss.

Once each player has a chance to reject unattractive trades, however, there are no more marginal gains from granting them additional opportunities to communicate with one another. Thus, moving from $N = 2$ to $N = 3$ presents no additional "screening" benefits. However, such a move does impose greater social costs by requiring an additional successful communication between the parties—a communication that may be lost.

This pattern is not only true for $N = 3$, but it continues for $N = 4, 5, 6$, and so on. And as N grows infinitely large (devolving into an infinite-communications rule), the ability to capture gains from trade eventually dissipates completely. Figure 1 illustrates how the expected level of social welfare changes with N . As the figure demonstrates, dramatic efficiency gains occur in moving from $N = 0$ through $N = 2$, and uniform decreases in expected welfare occur as N moves beyond $N = 2$. The downward trend for $N > 2$ in the figure is not limited to this particular example, so long as a

35. With a high-buyer/low-seller combination, the $N = 2$ rule produces a social gain of \$100 and an enforceable contract $81/100$ of the time, when neither the seller's nor the buyer's initial message is lost; a social loss of \$100 and an enforceable contract $9/100$ of the time, when the seller's message successfully arrives, but the buyer's message is lost; and neither a social gain nor a loss for the remaining $1/10$ of the time, when seller's initial message never reaches the low-cost buyer. In all other combinations (which occur jointly 84% of the time), the $N = 2$ rule produces no social gains nor losses, since either the high-cost seller self-selects by never sending the initial message, or the buyer self-selects by never responding.

36. Recall, for $N = 2$ the expected social welfare was \$11.52.

positive probability exists that serial messages will be lost.³⁷ Thus, the $N = 2$ rule appears to be the optimal one,³⁸ which essentially restates the holding in *Crook v. Cowan*.

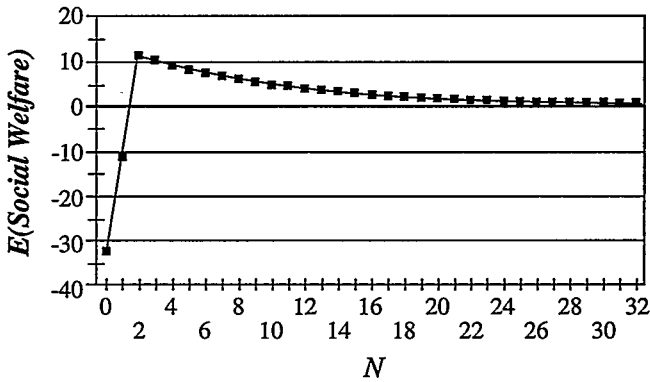


FIGURE 1
Expected Social Welfare as a Function of N

C. Reflections on the Model

Although the example above is lengthy and somewhat arduous, it suggests a number of interesting insights about how common-knowledge considerations can inform legal policy. On a narrow level, the example helps to elucidate an efficiency rationale behind the traditional default rule that both an offer and an acceptance are the necessary and sufficient conditions for the creation of an enforceable contract. As demonstrated above, such a rule is the only one that enables both sides of a potential transaction to avoid unwanted trades while simultaneously minimizing the costs imposed by superfluous communications. More simplified models of offer and acceptance using common knowledge and complete information are unable to provide such results (e.g., Katz 1990).

The example also provides additional insights into other offer-and-acceptance doctrines. Consider, for instance, the "mailbox rule," which deems an offer to be effective upon receipt and an acceptance to be effective upon dispatch. Under the optimal rule in the example ($N = 2$), a contract is

37. By setting the players' payoffs and communication protocol appropriately, it is possible to minimize the downward slope of the curve when $N > 2$. Nonetheless, the slope will always be at least weakly negative. In contrast, it may be possible to construct an example in which the cost of communication is so large that the optimal rule consists of $N = 0$ or $N = 1$ (e.g., Katz 1990).

38. This result contrasts with the model of Katz 1990, in which the costs of communication take a monetary (rather than probability) form. However, Katz's model is one in which all information is common knowledge, and it does not account for the beneficial effects of offeree self-selection.

formed when the buyer *sends* her reply to the seller, regardless of whether the seller ever receives it. Granted, such a rule admits the risk that a seller may never learn of the buyer's acceptance until it is too late to avoid a breach. However, as the model demonstrates, common-knowledge defects of this sort are often endemic to sequential communication: The most a legal rule can hope to do in such circumstances is to minimize the consequences of such asymmetries.³⁹ This last observation may also shed light on the apparent victory of "objective" over "subjective" theories of mutual assent within contract law during the last century. Indeed, if a "meeting of the minds" is the doctrinal equivalent of common knowledge, then a subjectivist formation rule suffers from many of the same weaknesses observed in the infinite-communication rule considered above.

On a broader level, the example suggests a number of implications that common-knowledge defects could have on a general normative account of law. Even outside the contract-negotiation process, legal rules may simply not be able to "resolve" imperfections in collective knowledge. Indeed, if individuals begin with divergent prior beliefs about the world, then their beliefs may never converge no matter how extensive their subsequent common experiences and dialogue with one another (Kalai and Lehrer 1994). But even when legal rules cannot remedy such nonconvergence, it may nonetheless be possible to craft rules that mitigate the costs imposed by such perpetual disagreement. Thus, for example, the Supreme Court's permissive approach toward the regulation of "fighting words" may reflect a type of regulatory damage control when the marketplace of ideas fails due to common-knowledge defects (*Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942); Spitzer 1997). Such a damage-control rationale may also support the argument that courts *appropriately* tend to couch their holdings in low-level principles so as to formulate "incompletely theorized agreements" among groups with irreconcilable beliefs (Sunstein 1996).

Moreover, common-knowledge considerations may play a key role in normative theories of interpretation, for both face-to-face communications and written texts. Communication breakdowns might occur not because of *failed* transmissions (as in the example), but rather because of either *garbled* transmissions (in the case of face-to-face bargaining) or because *meanings* attached to a particular text by each player are not common knowledge among all players. Thus, were one serious about pursuing a textualist agenda in enforcing the "plain meaning" of a statutory term, it may be insufficient to simply inquire whether an overwhelming majority of people mutually

39. If one were to alter the mailbox rule so that it would make a contract effective upon receipt of the offeree's acceptance, it would not overcome the knowledge asymmetry alluded to in the text; it would merely implement the equivalent of an $N = 3$ rule, shifting the risk due to lost messages from the buyer to the seller. As illustrated above, such a rule marginally decreases social welfare in equilibrium, because performance turns on one extra communication loop.

attach the same meaning to that term. Rather, the substantive benefits of a "plain meaning" rule may come about only when (1) legal actors share a mutual interpretation, (2) they are all aware of this shared interpretation, (3) they are aware of their awareness, and so on (*Smith v. United States*, 113 S. Ct. 2050 (1993); Polich 1994).

III. CONCLUSION

In this essay, I have used *Game Theory and the Law* as a stepping stone to speculate on one possible future course of work in the field. A refined appreciation of the nuances and defects in common knowledge, I have argued, is a prerequisite to a more complete normative account of legal policy in strategic settings. The example I have offered constitutes only a modest tip of a much larger analytical iceberg. While countless other conceivable legal applications exist for games of common knowledge, virtually all of them remain unexplored to date. Nevertheless, this iceberg is, in my view, eminently worth exploring, and *Game Theory and the Law* is a useful tool for the journey.

REFERENCES

- Aumann, Robert. 1976. Agreeing to Disagree. *Annals of Statistics* 4:1236–39.
- . 1989. Game Theory. In *The New Palgrave: Game Theory*, edited by John Eatwell, Murray Milgate, and Peter Newman. New York: W. W. Norton.
- Ayres, Ian, and Eric Talley. 1995. Solomonic Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade. *Yale Law Journal* 104:1027–1117.
- Brown, John. 1972. Toward an Economic Theory of Liability. *Journal of Legal Studies* 3:323–49.
- Cho, In-Koo, and David Kreps. 1987. Signalling Games and Stable Equilibria. *Quarterly Journal of Economics* 102:179–221.
- Cooter, Robert. 1997. Normative Failure Theory of Law. Working paper, University of California at Berkeley.
- Dau-Schmidt, Kenneth, Michael Alexeev, Eric Rasmussen, Jeff Stake, and Bob Heidt. 1997. Review Dialog: Game Theory and the Law. Working paper, Indiana University Law School, Bloomington.
- Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.
- Geanakopoulou, Jean. 1992. Common Knowledge. *Journal of Economic Perspectives* 6 (4):53–82.
- Huang, Peter. 1995. Strategic Behavior and the Law: A Guide for Legal Scholars to *Game Theory and the Law* and Other Game Theory Texts. Review. *Jurimetrics* 36:99–114.
- Kalai, Ehud, and E. Lehrer. 1994. Weak and Strong Merging of Opinions. *Journal of Mathematical Economics* 23:73–86.
- Katz, Avery. 1990. The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation. *Michigan Law Review* 89:216–95.

- Kreps, David, Paul Milgrom, John Roberts, and Robert Wilson. 1982. Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory* 27:245-52.
- Lucy, William. 1995. Review of *Game Theory and the Law*. *Cambridge Law Journal* 54:465-67.
- Osborne, Martin, and Ariel Rubinstein. 1994. *A Course in Game Theory*. Cambridge: MIT Press.
- McCutcheon, Barbara. 1997. Do Meetings in Smoke-filled Rooms Facilitate Collusion? *Journal of Political Economy* 105:330-50.
- Morriss, Andrew P. 1995. Review of *Game Theory and the Law*. *Journal of Legal Education* 45, (3):465-67.
- Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge: Harvard University Press.
- Polich, John. 1994. The Ambiguity of Plain Meaning: *Smith v. United States* and the New Textualism. *Southern California Law Review* 68:259-88.
- Rose, Carol. 1992. Women and Property: Gaining and Losing Ground. *Virginia Law Review* 78:421-59.
- Rubinstein, Ariel. 1989. The Electronic Mail Game: Strategic Behavior under Almost Common Knowledge. *American Economic Review* 79:385-91.
- Salant, Stephen, and Theodore Sims. 1996. Ready for Prime Time? Review of *Game Theory and the Law*. *Michigan Law Review* 94:1839-82.
- Shavell, Steven. 1987. *Economic Analysis of Accident Law*. Cambridge: Harvard University Press.
- Spier, Kathryn. 1994. Settlement Bargaining and the Design of Damage Awards. *Journal of Law, Economics, and Organization* 10:84-95.
- Spitzer, Matthew. 1997. Freedom of Expression. In *The New Palgrave: Law and Economics*, edited by Peter Newman. New York: W. W. Norton.
- Sunstein, Cass. 1996. *Legal Reasoning and Political Conflict*. New York: Oxford University Press.
- . 1993. Endogenous Preferences and Environmental Law. *Journal of Legal Studies* 22: 217-54.
- Van Damme, Eric. 1996. *Stability and Perfection of Nash Equilibrium*. 2d ed. Berlin: Springer-Verlag.
- Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.
- Weibull, Jorgen W. 1995. *Evolutionary Game Theory*. Cambridge: MIT Press.

