Experiments on Cognition, Communication,

Coordination, and Cooperation in Relationships

Vincent P. Crawford[1,2]

Prepared for *Annual Review of Economics*, Volume 11 (2019); this version 12 October 2018

Abstract: This review discusses selected work in experimental game theory. My goals are to further the dialogue between theorists and empiricists that has driven progress in economics and game theory, and to guide future experimental work. I focus on experiments whose lessons are relevant to establishing and maintaining coordination and cooperation in human relationships, the role of communication in doing so, and the underlying cognition. These are questions of central importance, where the gap between theory and experience and the role of experiments in closing it both seem large. Humans appear to be unique in their ability to use language to communicate and manipulate mental models of the world and of other people, vital skills in relationships. Continuing the dialogue should help to explain why it matters for cooperation that we can communicate, and why and how it matters whether we communicate via natural language or abstract signals.

Keywords: experimental game theory, behavioral game theory, strategic thinking, communication, coordination, cooperation

***JEL* codes:** C72, C92, C51

[1] Department of Economics and All Souls College, University of Oxford, Oxford OX1 3UQ, UK; e-mail: vincent.crawford@economics.ox.ac.uk.

[2] Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA.

1. INTRODUCTION

This review discusses selected work in experimental game theory. My goals are to further the dialogue between theorists and empiricists that has driven progress in economics and game theory, and to guide future experimental work. I focus on experiments whose lessons are particularly relevant to establishing and maintaining cooperation in relationships, the role of communication in doing so, and the cognition that underlies both; but I include some relevant work that is not labeled as about relationships. Communication, coordination, and cooperation in relationships are questions of central importance, where the gap between theory and experience and the role of experiments in closing it both seem large. Existing theory falls particularly short in explaining why and how communication matters as much as it does. Humans appear to be unique in their ability to use language to build and counterfactually manipulate mental models of their environments, and to communicate their conclusions to each other, skills that are vital in relationships. Yet we still have a great deal to learn about why it matters as much as it does that we can communicate, why it matters whether we can do so via natural language or abstract signals, and how the possibilities for communication shape how we can cooperate.

Experiments have played an essential role in the game-theoretic revolution of the past four decades, which has transformed the landscape of economics. Most modern economics studies systems in which individuals interact strategically to determine outcomes—systems for which analysis was once thought impossible simply because it was game-theoretic. Microeconomic examples include models of agency, contracts, bargaining, coordination, collusion, network or platform competition, market signaling or screening, auctions, social choice, public goods, and market design. Microeconomic examples even extend to "behavioral" models of individual decisions, such as static choice with reference-dependence or intertemporal choice with present bias, in which expectations about one's own future decisions play a central role. Macroeconomic examples include models of effective demand failures, rules versus discretion in policy, macroeconomic policy evaluation, the roles of information flows, sunspots, or "animal spirits" in the business cycle, speculative bubbles, and the vulnerability of financial markets to crises.

The game-theoretic revolution required much more than importing game theory into economics. The disciplines co-evolved, in a dialogue that has transformed game theory as much as economics. Economists contributed a rich set of questions and intuitions about behavior and outcomes, against which game theory's assumptions and predictions could be tested. Game theorists offered a rich language for describing strategic interactions and behavioral assumptions, which could be tested in economic applications. The dialogue began as conversations among theorists, with empirical content limited to introspection and casual empiricism. The unstated goal (in positive game-theoretic analyses) was to predict behavior and outcomes entirely by theory. The high-water mark of this approach was probably Harsanyi & Selten (1988), and Schelling (1960) was probably its most important critic. This phase of the dialogue clarified how to model economic interactions as games and to get the logic of strategic rationality right—no small accomplishments. Even so, pure theory often failed to yield convincing arguments for definite predictions, and those it did yield often deviated systematically from observation. As a result, the focus gradually shifted to relaxing unrealistic assumptions and building more credible models of observed behavior, endeavors for which empirical input played an essential role.

Experiments have long been the leading source of empirical information about the principles of strategic behavior, and they are likely to remain so despite the increased availability of large, high-quality observational datasets. Theories of strategic behavior are information omnivores, whose predictions are notoriously sensitive to the details of how people interact, what they know, what they know about what others know, and so on. There are a few field settings where some such details can be observed, such as auctions and matching markets; and some of them can be controlled in field experiments. But for most questions, laboratory experiments have a decisive advantage in the control needed for precise tests of theories of strategic behavior.

Experiments have another, less obvious advantage, which stems from the central role of cognition in modern economics. Experiments have long been used to study cognition indirectly, via designs in which different kinds of strategic thinking lead to different decisions. Some have even argued that because economic theory was intended to explain decisions, *only* decisions should be used to test it (Gul & Pesendorfer 2008; but see Camerer 2008 and Crawford 2008 for counter-arguments). However, in behavioral game theory, as in behavioral economics more generally, decisions and cognition are inextricably linked. And modern experimental methods make it possible to study cognition much more deeply, at modest incremental cost, by observing subjects' decisions while monitoring their searches for hidden information about the game, by monitoring chats between subject teams, or by eliciting subjects' beliefs about others' decisions. With such methods readily available, black-boxing cognition is often wasteful, and can be misleading. The cost of analyzing process or other nonstandard data is far from negligible, but the incremental benefits are large.

Space limitations require a sharp focus. I make no attempt at comprehensive coverage of the literature, even in the areas I consider, with apologies to those whose work is thereby omitted. I discuss details of experimental designs and methods only as needed to explain my points; see Davis & Holt (1993) and Camerer (2003) for more comprehensive accounts. Crawford (1998) and the lecture slides at Crawford (2017b) discuss the experimental literature on communication in games in more detail than is possible here. I omit econometric methods; for more detail and further references, see Camerer (2003), Costa-Gomes et al. (2001), and Costa-Gomes & Crawford (2006) (henceforth "CGC"). I omit machine learning and artificial intelligence methods of estimation; for introductions and further references, see Mullainathan & Spiess (2017), Fudenberg & Liang (in press 2019), and Gentzkow et al. (in press 2019). Finally, I omit neuroeconomics; for introductions and references see Camerer et al. (2005) and Camerer (2013).

Section 2 reviews terminology and assumptions. Section 3 considers *tacit* interactions—those without communication—including normal- or extensive-form games played once and repeated games. Section 3 also discusses some alternatives to eliciting decisions for studying strategic thinking: monitoring information search, eliciting beliefs, and monitoring subject-team chats. Section 4 considers interactions with communication, focusing mainly on the repeated Prisoner's Dilemma. Section 5 is a brief conclusion.

2. TERMINOLOGY AND ASSUMPTIONS

I focus mainly on two-person interactions, modeled as *noncooperative games*.[3] A noncooperative game is defined by its *structure*: its decision-makers or *players*, their feasible *decisions*, how their decisions interact to determine outcomes, and their preferences over outcomes, represented by von Neumann-Morgenstern utilities called *payoffs*. I distinguish *normal-form* games, in which players make single, simultaneous decisions, from *extensive-form* games, in which players make series of simultaneous and/or sequential decisions, with the order usually represented by a game tree. A *one-shot* (normal- or extensive-form) game is one that is played only once. A *repeated* game is one where the same *stage game* is played, finitely or infinitely often, by a fixed group of

[3] The term is a misnomer, in that noncooperative game theory is used to explain cooperative as well as other outcomes. I omit important experimental work on cooperative game theory, which sidesteps most details of the structure and seeks to characterize possible outcomes of frictionless bargaining without commitment among rational players.

players. Extensive-form or repeated games require generalizing the idea of a decision to that of a *strategy*, a contingent decision rule that specifies a player's decisions at each point s/he is called upon to make one, as a function of what s/he can then observe.

Games with communication are those in which players can send messages. I assume that any messages are *cheap talk*, with no direct payoff consequences. If such messages are in a language that makes lying a meaningful concept, cheap talk implies that lying has no direct payoff cost. More generally, cheap talk messages must have no positive or negative connotations. I distinguish between the communication phase of an interaction and the *underlying* game or stage games, the parts in which decisions other than messages directly determine outcomes. The communication phase may be structured or unstructured, limited or unlimited in duration, and messages may be via abstract signals or in natural language. In general, messages might concern players' private information; their non-binding *intentions* of future decisions; and possibly also their explanations, rationales, or motivations for decisions. Because I focus mainly on symmetric-information games, there is usually no private information for messages to convey. Otherwise messages are just another kind of decision, to which the above terminology applies.

The essential difficulty of game theory is that game outcomes depend not only on players' own decisions, but also on others' decisions that players cannot fully observe and must therefore predict. To focus on this difficulty, I assume, except where noted, the game-theoretic analogue of symmetric information: that the structure is *common knowledge*, in that players know it, know others know it, and so on ad infinitum. I also assume that people are at least approximately decision-theoretically *rational*, in that they best respond to *beliefs* about others' decisions, are logically omniscient, and make Bayesian probabilistic inferences. Experiments suggest that well-motivated subjects who understand the game make decisions consistent with rationality 80-90%

of the time. The evidence on logical omniscience and Bayesianism is less encouraging, but it's my working hypothesis—and that of most of the experimental literature—that it's useful to study how people form beliefs without considering interactions with failures of logical omniscience or of Bayesianism.[4] Otherwise I take an unrestricted, evidence-based view of players' beliefs.

Traditional game theory's benchmark model of players' beliefs is that they are determined in a *Nash equilibrium* (henceforth shortened to *equilibrium*), defined as a profile of strategies in which each player's strategy maximizes her/his own payoff, given the others' strategies.[5] This definition can be applied without regard to its rationale, but can be thought of as an *equilibrium in beliefs*—a rational-expectations notion in which players' beliefs about others' strategies are correct, given the rational choices they imply. The generality and tractability of equilibrium analysis have made it the method of choice in most applications. In extensive-form or repeated games, equilibrium is often strengthened to *subgame-perfect equilibrium*, defined as an equilibrium strategy profile that induces an equilibrium in every *subgame*, any part of a game that remains after part of it has been played. Thus subgame-perfect equilibrium requires a kind of time-consistency of the theory of behavior. A subgame-perfect equilibrium in a finite-horizon game can be found by backward induction. (In games with asymmetric information, the notion of subgame-perfect equilibrium must be generalized, to *sequential* or *perfect Bayesian* equilibrium.)

Equilibrium, or subgame-perfect equilibrium, bundles the standard decision-theoretic notion of rationality with the far stronger assumption that players' beliefs and strategies are perfectly coordinated. This strengthening serves an essential purpose. Bernheim (1984) and Pearce (1984) have shown that even common knowledge of decision-theoretic rationality implies only that

[4] Eyster & Rabin's (2005) notion of cursed equilibrium focuses instead on violations of Bayesianism, but Crawford & Iriberri (2007) argue that much of the evidence they consider can be given a better and more unified explanation by considering Bayesian but non-equilibrium beliefs in level-*k* models.

[5] It is sometimes useful to allow *mixed* strategies, in which players randomize over their *pure*, or unrandomized strategies.

players' will choose *rationalizable* strategies—roughly, those that survive iterated elimination of strictly dominated strategies. Rationalizability, however, yields few or no restrictions on behavior in most games of economic interest (Crawford 2016, pp. 133-136, gives examples). Equilibrium or subgame-perfect equilibrium adds the precision that is needed in most economic analyses, but even equilibrium stops well short of making unique predictions in many games.

There are two main views of how players' strategies might come to be in equilibrium. (I omit some qualifications for subgame-perfect, sequential, or perfect Bayesian equilibrium in extensive-form or repeated games, and/or games with asymmetric information.) In the *learning* view, players play the same game repeatedly, either with varying partners or with negligible influence on others' future decisions, so that their strategies in the game that is repeated—I will abuse terminology by calling this the stage game—are the natural objects of choice. Otherwise, players may focus on their strategies in the repeated game that describes the entire learning process. Players adjust their stage-game strategies over time in ways that would increase their own stage-game payoffs if others' stage-game strategies continued unchanged. There are few general theoretical results; but even if players are naïve, such learning processes tend to converge to some equilibrium in the stage game—a prediction for which there is experimental support.

In the *thinking* view of equilibrium, players' beliefs and strategies are independently focused on the same outcome—even if they have no prior experience with the game—by reasoning based on a commonly understood principle that makes unique predictions (e.g. Brandenburger 1992). Given players' rationality, such an outcome must be an equilibrium. In all but the simplest games, particularly when the equilibrium is a nontrivial fixed point, such reasoning places heavy demands on players' cognition. Those demands are even more severe if the game has multiple equilibria, in which case players' beliefs must be independently focused by a commonly understood coordination refinement, such as Harsanyi & Selten's (1988) notion of payoff-dominance, favoring equilibria whose payoffs are not Pareto-inferior to other equilibria; or risk-dominance, favoring equilibria (roughly) with larger "basins of attraction," or sets of beliefs that support their strategies as best responses.

Economists have long been aware that people often lack enough experience with closely analogous games to make equilibrium's learning justification plausible. The learning justification seems especially strained for cooperation in relationships, because most important ones—marriages, careers, partnerships—afford little or no opportunity for practice.[6] I therefore focus on experiments that study subjects' strategic thinking in their initial responses to games, rather than learning in repeated play. In the experiments discussed below, thinking is studied in designs that repeatedly, anonymously pair subjects from a large population to play series of stage games, with new partners each period and without immediate feedback, to suppress learning from experience and repeated-game effects.[7] If each stage game is different, subjects' responses to each stage game can be treated as if it was played in isolation. If the stage games are all the same, subjects' first responses to the stage game can be treated as if it was played in isolation. I use "initial responses" to refer to games played as if in isolation in either sense.

3. TACIT INTERACTIONS

3.1. One-shot normal-form games

Economists have also long been aware that people often deviate systematically from equilibrium. Experimental subjects do avoid dominated strategies, usually with frequencies above 90%. But

[6] It is the rarity in everyday life of the perfect analogies assumed in learning models that makes the protagonist Phil Connors's use of unlimited practice in the film "Groundhog Day" so striking. Experimental analysis of learning from imperfect analogies is still in its infancy, but see Rankin et al. (2000), Van Huyck & Battalio (2002), and Cooper & Kagel (2008).

[7] "Eureka!" learning remains possible, and sometimes occurs, but can be tested for and is rare enough not to be a problem (e.g. Costa-Gomes et al. 2001, Cooper & Kagel 2005, or Costa-Gomes & Crawford 2006).

they rely on others' avoiding dominated strategies with much lower frequencies; still fewer subjects rely on more than one round of iterated dominance; and the presence of dominated strategies often affects outcomes even when rarely played. Systematic deviations are particularly apparent in responses to out-of-equilibrium payoffs (e.g. Goeree & Holt 2001 or Camerer 2003).

Adding payoff-sensitive noise to equilibrium predictions, which I will call "equilibrium-plus-noise", does not adequately account for people's systematic deviations from equilibrium. Until recently the only alternative was McKelvey & Palfrey's (1995) notion of quantal response equilibrium ("QRE"; see also Rosenthal 1993). In a QRE players' decisions are noisy with an assumed distribution, usually logit, and players best respond noisily, with a decision's density increasing in its expected payoff, taking the noisiness of others' decisions into account. This often allows QRE to fit observed behavior better than equilibrium-plus-noise. However, a QRE is a fixed point in a high-dimensional space of distributions, making its thinking justification cognitively far more demanding than for Nash equilibrium. Further, QRE's predictions are highly sensitive to its assumption about the error distribution; and the standard assumption that the distribution is logit is an untested, and probably untestable, modeling convention (Haile et al. 2008; but see Goeree et al. 2005 for counter-arguments).

Attempts to find alternatives to equilibrium-plus-noise or QRE models of strategic thinking were long held back by doubts that any model could systematically out-predict a rational-expectations notion like equilibrium (viewed as equilibrium in beliefs), and doubts about finding a credible basis to select one of the enormous number of logically possible nonequilibrium models. In an important example of the power of experiments to break theoretical roadblocks, experiments eliciting subjects' initial responses to games may answer both doubts.

Those experiments show that most subjects find fixed-point or indefinitely iterated dominance reasoning inaccessible. Some have suggested that this is a quirk of the laboratory, and that experienced decision-makers in the field will use such reasoning when the stakes are high enough; but I have yet to find even anecdotal evidence of quants or poker experts using fixed-point or indefinitely iterated dominance reasoning. Given the inaccessibility of such reasoning, if a game has equilibria that require it but the setting lacks the precedents for learning (which might enable them to converge to a fixed point without thinking), then subjects must find an alternative way to think about the game. The experiments I will now discuss suggest that subjects playing normal-form games tend to follow one of a simple class of "level-k" rules, described below, which imply systematic, predictable deviations from equilibrium—leading to a simple class of models that have the potential to answer both doubts for normal-form games.

The first hints that people's deviations from equilibrium might have a coherent structure came from studies of normal-form games by Nagel (1995) and Ho et al. (1998), inspired by Keynes's (1936, Chapter 12) "beauty contest" example. Nagel's and Ho et al.'s subjects played series of n-person guessing games, in which subjects made simultaneous guesses between common lower and upper limits and the subject who guessed closest to a target times the group average guess won a prize. The targets and limits varied only across treatments and the structures were publicly announced, to justify comparing results with symmetric-information equilibrium predictions. Subjects played the same games repeatedly in the same groups, so I omit Ho et al.'s small groups and focus on Nagel's and Ho et al.'s other subjects' first responses to the stage game.

Most of Nagel's and Ho et al.'s stage games are dominance-solvable, with unique equilibria. Yet subjects' guesses resembled neither equilibrium nor QRE, for any reasonable noise distribution. Most guesses respected 0 to 3 rounds of iterated dominance, when 3 to an infinite number are needed for equilibrium. And there were spikes in the continuous strategy spaces, which in games with limits [0, 100] tracked $50p^k$ for $k = 1$, 2, or 3 across the various targets p (Nagel 1995, Figure 1). In Crawford et al.'s (2013, p. 17) words, "Like the spectrograph peaks that foreshadow the existence of chemical elements, the spikes suggest that subjects' deviations from equilibrium have a coherent structure…." These results strongly reject equilibrium explanations, and suggest that subjects have heterogeneous, discrete levels of reasoning.

What were Nagel's and Ho et al.'s subjects doing? To evaluate the evidence, it is useful to define two families of nonequilibrium decision rules. Costa-Gomes et al.'s 2001 and CGC's 2006 rule *Dk* performs *k* rounds of iterated deletion of dominated strategies and then best responds to a uniform prior over the other player's surviving decisions in a two-person game (or their group average, in the *n*-person guessing games discussed below). The alternative rule *Lk* performs *k* iterated best responses to a uniform random "*L0*" prior over the other player's feasible decisions (or their group average, in *n*-person guessing games). Theorists often view Nagel's and Ho et al.'s spikes as prima facie evidence that subjects are performing finitely iterated dominance. But a spike at $50p^k$ could reflect either subjects following the rule *Dk*-1 or the rule *Lk* (CGC, p. 1739; Crawford et al. 2013, Section 3.1).[8]

More generally, Nagel's and Ho et al.'s designs, with large strategy spaces but only one initial response per subject, very weakly separate these and other decision rules. Stahl & Wilson (1994, 1995; henceforth "SW") and Costa-Gomes et al. (2001) found similar (though less visually interpretable) patterns of deviations in subjects' initial responses to series of 10 to 18 different matrix games. But their studies, with only two to four possible decisions per game, also had

[8] I ignore Breitmoser's (2012) qualifications regarding *Lk*'s decisions in *n*-person guessing games, as behaviourally irrelevant.

weak separation. As a result, all of these studies' data analyses had to rely heavily on a priori specifications of the possible rules and were highly sensitive to the econometric assumptions.[9]

The remedy was not more sophisticated econometrics using the same datasets (of which there are several examples), but more powerful experimental designs. CGC's design combined the strengths of Nagel's and Ho et al.'s large strategy spaces with SW's and Costa-Gomes et al.'s (2001) series of different games. CGC's subjects played 16 different two-person guessing games, invented for this purpose. In each guessing game, each player has her/his own lower and upper limit and target. A player's payoff is higher, the closer her/his guess is to her/his target times her/his partner's guess. The games have essentially unique equilibria and are finitely dominance-solvable, though often in large numbers of rounds. The targets and limits vary independently across players and games, so a subject's guess sequence creates a high-dimensional "fingerprint" of her/his strategic thinking. Another advantage is that two-person guessing games fully engage strategic thinking, in that subjects are not tempted to think others will treat subjects' own influences on outcomes as negligible, as they may be in *n*-person guessing games.

The main difficulty in analyzing data from experiments on strategic thinking is identifying subjects' decision rules within the enormous, unstructured set of logically possible rules. CGC, building on Costa-Gomes et al. (2001), addressed this difficulty in two ways. To avoid misspecification, they created a list of possible decision rules, which included all the generalizable rules that earlier studies suggested might be empirically relevant. Those rules included the *Lk* and *Dk* rules defined above plus *Equilibrium*, which makes its equilibrium decision; and *Sophisticated*, which best responds to the observed distribution of potential partners' guesses, to test whether any subject understands others' behavior better than the

[9] Compare SW's (1994, 1995) or Costa-Gomes et al.'s (1998, 2001) different estimates on closely related or the same data.

specified rules.[10] CGC's econometric model assumed that each subject follows one of those rules in all 16 games, up to logit errors, and estimated by maximum likelihood which rule best fit each subject's decisions. Finally, CGC conducted a specification test, designed to detect any rules not on the specified list that describe more than one subject's guesses better than any rule on the list.

This procedure led CGC to classify 63% of the subjects as following one of the specified rules, leaving the other 37% unclassified (CGC, Table 1, p. 1741). Only four rules were helpful in explaining subjects' guesses. Of the 88 subjects in CGC's main treatments, the guesses of 43 (49%) complied *exactly* (within 0.5) with one rule's guesses in 7 to 16 of the games: 20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*. Given how strongly the rules are separated (CGC, Table 5, p. 1751) and that guesses could take from 200 to 800 rounded values in each game, such clear fingerprints could not plausibly happen by chance. For those 43 subjects, one can intuitively accept (pace Popper) the hypothesis that they followed their apparent rules, with error. Moreover, those rules build in risk-neutral, self-interested rationality and perfect models of the game. Thus, the high rates of exact compliance rule out alternative interpretations of behavior; and the deviations from equilibrium of the 35 subjects whose apparent rules are *L1*, *L2*, or *L3* can be confidently attributed to level-*k* beliefs rather than irrationality, risk aversion, altruism, spite, or confusion. (By contrast, with SW's or Crawford et al.'s (2001) coarse strategy spaces, even a perfect fit would not distinguish a subject's estimated rule from nearby omitted rules and there would be no way to rule out misspecification. In Nagel's and HCW's designs, the ambiguity is more severe.)

For CGC's other 45 subjects, noisier fits require econometrics. But 31 of those 45 violated dominance less than 20% of the time (versus 38% for random guesses), suggesting that their

[10] The uniform random *L0* is meant to be the starting point of *L1*'s and higher levels' thinking, not to describe anyone's actual decisions. It seldom shows up in econometric estimates unless the econometric model is given no other way to explain errors.

behavior was coherent; and their estimated rules are also concentrated on *L1*, *L2*, *L3*, and *Equilibrium*, in roughly the same proportions as for the 43 subjects with clear fingerprints.

The resulting level-*k* model is simple, tractable, applicable to any normal-form game, and, given a calibrated or estimated distribution of *k*, makes precise probabilistic predictions. For a given level distribution, a level-*k* model predicts not only that deviations from equilibrium will sometimes occur, but also which games are likely to evoke them, the forms they are likely to take, and their frequencies. Level-*k* rules are decision-theoretically rational; their only departure from *Equilibrium* is in replacing its perfect model of others with a simplified non-equilibrium model. *Lk*'s decisions survive *k* rounds of iterated elimination of dominated decisions and so in two-person games make decisions that are *k*-rationalizable (Bernheim 1984), just as *Dk*-1's are, but generally with a different selection. Thus, a level-*k* model can be viewed as a heterogeneity-tolerant refinement of *k*-rationalizability. (Aradillas-Lopez & Tamer 2008 show how weak the restrictions implied by unrefined *k*-rationalizability with low values of *k* can be.) Given the prevalence of low levels, level-*k* models mimic equilibrium in games that are dominance-solvable in small numbers of rounds but imply systematic deviations in most more complex games. The model explains those deviations as responses to out-of-equilibrium and other payoffs via its deterministic structure, without untested distributional assumptions. Thus, the model shares most of the desiderata that favor equilibrium analysis, while avoiding the cognitive complexity of fixed-point or indefinitely iterated dominance reasoning. (In games with multiple equilibria the model probabilistically predicts equilibrium selection—though not always an equilibrium—while avoiding the greater complexity of risk- or payoff-dominance refinements.)

Almost all of what I say here is also true of Camerer et al.'s (2004) cognitive hierarchy model, in which level-*k* best responds to an estimated mixture of lower levels and the population level distribution is assumed to be Poisson with an estimated parameter τ (the population average *k*). But in a cognitive hierarchy model, *Lk* for $k \geq 2$ need not make *k*-rationalizable decisions.

CGC's conclusions are generally consistent with those of other careful studies of initial responses to normal-form games. Fragiadakis et al. (2016), for instance, stayed close to CGC's instructions and procedures and used mostly CGC's games, adding some similar guessing games. Requiring 8/20 exact guesses rather than CGC's 7/16 (with such large strategy spaces, requiring that many exact decisions is far more stringent than usual), they classify 30% of their subjects as following one of CGC's *Equilibrium*, *Lk*, or *Dk* rules, in roughly the same proportions but with lower frequencies of *L1* and *L2* and a higher frequency of *Equilibrium*. To test which subjects were following other, not-yet-identified rules, they added treatments in which subjects were asked, after their initial series of decisions, to either replicate their own past choices in one treatment, or best respond to them in another. Most of their unclassified subjects fail to replicate or best respond, but their classified subjects are far more likely to do so. Conversely, their classified subjects are only 40% of those who replicate, but 68% of those who best respond. Fragiadakis et al. also leave a large number of subjects unclassified, much as in CGC's Table 1; and they conclude that there could be significant groups of subjects who best respond but are not following one of CGC's specified rules, or who are following completely non-strategic rules.

Kneeland (2015) introduces a novel design using "ring games", which allows her to cleanly test for subjects' levels of reasoning—roughly, the highest *k*s for which their decisions respect *k*-rationalizability—without specifying an *L0*. In two-person games, varying *L0* allows *Lk* to sweep out the set of *k*-rationalizable decisions, so her tests can also be viewed as tests of level-*k* models with unrestricted *L0*s. Her estimated distributions of levels of reasoning are consistent with the estimated distributions of *k* in previous level-*k* analyses.

We now have tests supporting some form of level-*k* model in several different classes of two-person normal-form games: SW and Costa-Gomes et al. (2001) in small matrix games; Camerer et al. (2004) in a variety of games; Arad & Rubinstein (2012) in larger matrix games; Kneeland (2015) in ring games; Fudenberg & Liang (in press 2019) in a range of matrix games, using machine learning; and Brocas et al. (2014) in zero-sum betting games, generalizing the level-*k* model to asymmetric information (Camerer et al. 2004), with a data analysis based on clustering that avoids many of other studies' econometric assumptions. The consistency of results across varying classes of games is reassuring, but each subject played only a series of games within a given class, and the comparisons all concern aggregate level distributions, between subjects.

A more stringent test of the level-*k* model's portability is whether it can predict behavior within subjects across games in different classes. Georganas et al. (2015) (see also Hyndman et al. 2013) suggested that if a level-*k* model describes a person's behavior, her/his level should be constant across games or at least co-vary positively across classes of games. One might expect constancy because a level-*k* rule is generalizable and meant to describe an aspect of cognition. Or, if the level reflects a cost-benefit trade-off, as discussed below, a person's level might be linked to cognitive ability but vary with the complexity of the game, in which case it would co-vary positively across classes. Georganas et al. elicited intial responses within subjects across six of CGC's guessing games and four "undercutting games" like Arad & Rubinstein's (2012). Unlike Fragiadakis et al. (2016), Georganas et al. drastically shortened CGC's instructions and omitted CGC's understanding test; and for Georganas et al.'s games like CGC's, their results are far noisier. Even so, their results support a level-*k* model of aggregate behavior within each class of games, but they only very weakly support positive co-variation of levels across classes.

It is plausible that a person's level of reasoning is positively associated with her/his intelligence or sophistication, and estimated level distributions do seem to vary with intuitive guesses about subjects' intelligence and the complexity of the game. Camerer et al. (2004, Table II, p. 875) report estimates of their cognitive hierarchy model's average level for subject groups with different demographics or occupations playing Nagel's (1995) guessing game. On the assumption of positive association, the estimates roughly confirm intuitive inferences about the sophistication of various groups—although I would have expected CEOs or Caltech trustees to have higher levels than U.S. high school students. Agranov et al. (2012), Georganas et al. (2015), and Gill & Prowse (2016) used intuitive inferences or intelligence tests to create groupings with asymmetric expected sophistication to play Nagel's guessing games or other games. Average levels increased with own sophistication and partner's perceived sophistication as expected.

None of these studies answer the basic question of what determines the level that drives a level-*k* person's decisions: her/his own cognitive limitations, her/his perception of the partners' limitations, or both? Untangling those factors is challenging because the partner's revealed level is also subject to both influences, creating a nontrivial feedback loop. Friedenberg et al. (2018) create a theoretical, epistemic framework that addresses this identification problem. In their model, a player may be *strategic* in the sense of having some theory about how to play the game; rational in the full epistemic sense (including belief formation; see e.g. Brandenburger 1992); or both. A rational player must be strategic, because rationality and beliefs derived from knowledge of rationality dictate play; but a strategic player need not be rational. Players are assumed to have exogenous rationality and strategic bounds, the former never higher than the latter—the model's unexplained fundamentals. In previous analyses those two bounds were implicitly set equal, but Friedenberg et al.'s theory allows them to be identified separately. Assuming that all players are rational (but not necessarily known to be), Friedenberg et al. reanalyze Kneeland's (2015) data, finding that nearly half of the subjects she identified as having rationality (and thus strategic) bounds of 1 or 2 are now estimated to have higher strategic than rationality bounds. Friedenberg et al.'s analysis implies a significant difference in the model inferred from Kneeland's data.

Alaoui & Penta (2016) propose another approach to endogenizing the level distribution, in which the model's unexplained fundamentals are cognitive costs, which play a role like Friedenberg et al.'s cognitive bounds. (Rationality-based explanations of people's levels that abstract from cognitive costs or bounds tend to end up at equilibrium, which is behaviorally a dead end.) Alaoui & Penta study the tradeoff between the costs and benefits of higher levels while distinguishing a player's cost from her/his beliefs about her/his partner's cost. Assuming that costs are the same across "cognitively equivalent" games, their theory delivers comparative statics predictions about the effects of changing payoffs, or beliefs about partners' costs. Those predictions generally fit the facts from both previous work and Alaoui & Penta's experiments.

Further work on strategic thinking in one-shot normal-form games should strive for clearer experimental identification of subjects' decision rules and how they vary across games with the wide range of structures in applications. This will require more tests of replicability and portability, and tests of portability that build on replicability would allow clearer diagnoses of the causes of failures of portability. It will be helpful to continue substituting better experimental design for more sophisticated econometrics, by studying games for which decisions have more power to reveal cognition than in small matrix games or simple guessing games. Finally, field relevance seems to require designs that ensure that subjects understand the games they are playing, so that they will focus sharply on the problem of predicting others' responses. This is not to suggest that confusion does not happen in the field, just that uncontrolled confusion in the laboratory is not conducive to informative experiments about strategic thinking.

3.2. One-shot extensive-form games

Economists have also long been aware of deviations from subgame-perfect equilibrium in extensive-form games. For instance, Beard & Beil (1994) studied simple two-person extensive-form games in which one player has a dominated strategy, varying payoffs across treatments. Their subjects normally made undominated decisions themselves, but relied much less on dominance for others, at rates that varied sensibly with changes in own and partner's payoffs, and with the cost a subject's decision imposed on her/his partner. Again, adding noise to subgame-perfect equilibrium predictions does not adequately account for observed behavior.

A number of experimental papers report results suggesting that, far from performing full backward induction, subjects in extensive-form games often look ahead only a small number of periods, with heterogeneous horizons. Camerer et al.'s (1993) and Johnson et al.'s (2002; henceforth collectively "CJ") results for alternating-offers bargaining games (Section 3.3) provide evidence for small, heterogeneous horizons. Binmore et al. (2002) also report experiments with alternating-offers bargaining games. They dissect backward induction into three components: rationality, *subgame consistency* (play in a subgame is independent of its position in the larger game), and *truncation consistency* (replacing a subgame with its equilibrium payoffs does not affect play elsewhere in the game); and create a powerful design to test each component. Their subjects violated truncation consistency (among other things), responding systematically more to variations in the expected value of playing a subgame than to equivalent variations in its terminal payoffs—again suggesting small, heterogeneous horizons.

CJ's and Binmore et al.'s analyses suggest that their subjects may have used something like a dynamic version of level-*k* thinking, where *k* corresponds to the number of periods looked ahead. García-Pola et al. (2018) consider more explicit structural explanations of non-subgame-perfect equilibrium behavior along those lines. In a powerful design, they elicit initial responses to 16 different Centipede games, using the strategy method as is necessary to cleanly study strategic thinking in extensive-form games. They used some games like those in previous work, but also varied the structures to separate dynamic level-*k* models, QRE, and altruistic preferences. As usual, subgame-perfect equilibrium play is rare. There is no sign of altruistic preferences, but subjects' behavior is explained by a roughly equal mix of the level-*k* model and QRE.[11]

Other related work includes Kawagoe & Takizawa (2012) and Ho & Su (2013), who define dynamic level-k-style learning models for extensive-form games, reminiscent of Selten's (1991) anticipatory learning and Stahl's (1996) rule learning; and show that they give credible accounts of the decision dynamics in previous Centipede and sequential bargaining experiments.

3.3. Monitoring information search, eliciting beliefs, and monitoring subject-team chats

Returning to CJ's and CGC's analyses, I now consider their uses of process data, in the form of searches for hidden but freely accessible information about the game, to study strategic thinking. Crawford (2008) gives more detail. CJ were the first to use such methods in games, adapting Payne et al.'s (1993) MouseLab methods for studying individual decisions. Other noteworthy applications to games include Costa-Gomes et al. (2001); Camerer & Johnson (2004); Wang et al. (2010), using eye-tracking instead of Mouse-tracking; and Brocas et al. (2014).

[11] McKelvey & Palfrey (1992) originally explained their results, in the spirit of Kreps et al. (1982), via a structural model with asymmetric information about whether players are altruistic. McKelvey & Palfrey (1998) suggested a model that applies QRE to the agent normal form of an extensive-form game, which is less ad hoc and fits better than their original model.

CJ's analyses are another example of how experiments with powerful designs can break theoretical roadblocks. They studied initial responses to three-period alternating-offers bargaining games, with time-discounting simulated by shrinking the size of the "pie" to be divided each period. These games have unique subgame-perfect equilibria, from which subjects' decisions deviated systematically. When they wrote, the deviations were variously attributed to subjects' cognitive limitations or to social preferences that make subjects willing to take revenge after "unfair" offers by rejecting them, even at a pecuniary cost. It was clear that both factors were important, but there was no agreement on their relative importance.

CJ's innovation was to present the games to subjects using a MouseLab interface that conceals each period's pie but allows subjects to look them up as often as desired, one at a time, while recording their searches. With low search costs, free access made the entire structure effectively public knowledge, except for privately known revenge motives that made most subjects willing to take pecuniary losses to punish unfair offers, which CJ controlled for. Given a specification of the distribution of private revenge motives, the games are simple enough to be solved by backward induction even with this asymmetry of information.

Intuitively, backward induction in CJ's design has a characteristic search pattern, in which subjects first look up the last-period pie, then the second-last, and so on, with most transitions from later to earlier periods. This claim generated pushback from theorists, because it is possible to scan the pies in any order, memorize them, and only then process them into decisions, in which case the *order* of look-ups will have no necessary relation to the rule being implemented. Empirically, however, subjects using MouseLab have a strong tendency to rely on the interface as an aid to thinking, and seldom memorize. As a result, a subject's look-up *order* is closely related to the decision rule s/he is trying to implement. CJ supported this empirical claim by showing that a separate control group of subjects, trained in backward induction and rewarded for making their subgame-perfect equilibrium decisions, exhibited just the claimed pattern.

By contrast, untrained subjects, rewarded for their game payoffs against other untrained subjects, spent 60-75% of their time checking the first-period pie, 20-30% checking the second, and only 5-10% checking the third pie, with most transitions from earlier to later pies: the opposite of the backward-induction pattern. All subjects looked at the first-period pie at least once, but 19% never looked at the second and 10% never looked at the third. Even if those 19% or 10% made subgame-perfect equilibrium offer or acceptance decisions—which would trigger a strong "as-if" reflex in theorists who believe economic theory should only be tested by observing decisions—they could not have been making equilibrium decisions for equilibrium reasons. And without those reasons, their equilibrium decisions in this setting give essentially no evidence that they will make equilibrium decisions in other settings.

To show that subjects' search patterns can reveal their cognition, CJ (2002) conducted a rule-based analysis of a dynamic level-*k* model, in which some subjects are assumed not to look ahead at all, while treating the bargaining game as an ultimatum game; others to look ahead one round, while treating it as a truncated two-stage bargaining game; and still others to look ahead the full two rounds and play their subgame-perfect equilibrium strategies. (All levels' decisions were calculated taking responders' estimated average rejection behavior into account.) CJ's subjects' estimated levels were heterogeneous, with a strong positive association between the closeness of their look-up patterns to the backward-induction pattern and the closeness of their offer and/or acceptance decisions to the subgame-perfect equilibrium. CJ's results allow an evidence-based assessment of the relative importance of limited cognition and social preferences.

Their data analysis assigned them roughly equal weights—significantly more weight on limited cognition than analyses of decisions alone in alternating-offers bargaining had suggested.

CGC, within their 16 guessing games' common publicly announced structure, presented the games one by one to subjects via MouseLab, which hid both players' targets and limits but allowed subjects to look them up as often as desired, one at a time (CGC, Figure 6, p. 1753). Publicly announced free access to the hidden payoff parameters again made the structure effectively public knowledge; and the analysis of decisions (Section 3.1) suggests that there were no revenge motives or other distortions, so the results can be used to test theories of behavior in symmetric-information versions of the games.

CGC's search design combines the strengths of CJ's presentation of games as functions of a small number of hidden parameters within an intuitive common structure and Costa-Gomes et al.'s (2001, Figure 1, p. 1201) multi-dimensional search spaces. CJ's design elicited searches that varied mainly in one dimension, backward versus forward, enough for their purpose. But CGC's goal, like Costa-Gomes et al.'s, was to use search to separate a large number of rules that vary across their games in more nuanced ways. CGC's higher-dimensional search spaces allow then to strongly and independently separate leading decision rules' implications for search as well as guesses, with search implications that can be characterized in a tractable way.

Almost all studies of cognition via search have followed psychology's tradition of focusing on response times and the numbers and durations of look-ups, usually aggregated across subjects and over time (e.g. Rubinstein 2007). Like CJ and Costa-Gomes et al., CGC focused instead mainly on which look-ups subjects made at all, in which order. Because search behavior, like cognition, is highly heterogeneous, they also studied search and decisions at the individual level.

CGC's analysis of search is based on a procedural model of cognition, in which a subject's decision rule first determines her/his searches, and the rule and searches then determine her/his decision. Each of the leading decision rules is naturally associated with one or more algorithms that process payoff information into decisions. CGC derived a rule's search implications under conservative, empirically motivated assumptions about how cognition drives search. (I am unaware of any comparable theory of how a rule should be related to response times.) First, if a subject's rule's decision depends on a payoff parameter, it must appear in her/his search sequence. Second, if a subject's rule's decision depends on a basic operation on two parameters, they must appear adjacently in the sequence. More precisely, under CGC's assumptions each decision rule has a characteristic search sequence, the most efficient way to look up and combine the parameters needed to identify its decision (CGC, Table 4, p. 1751). A subject's search compliance for a given rule is measured by the density of the rule's characteristic sequence in the subject's observed search sequence. As with CJ's assumptions about how cognition drives search, the search data for subjects trained and rewarded for following the leading rules make it clear that these assumptions are at least approximately satisfied for almost all subjects (e.g. Crawford 2008, Table 10.2, p. 267; compare with Table 10.1, p. 265, or CGC, Table 4, p. 1751).

CGC (Sections II.C-E) describe the econometrics of subjects' guesses and searches in detail. The search analysis extends CGC's (and Costa-Gomes et al.'s 2001) maximum likelihood error-rate models of decisions to explain search compliance along with guesses, estimating each subject's most likely rule. Importantly, the list of specified rules and their characteristic search sequences create a kind of basis for the enormous spaces of possible guess and search sequences. This makes it meaningful to say whether the directions in which a given subject's guesses and searches deviated from equilibrium guesses and searches are positively associated—just as it was possible to say in CJ's space of possible searches whether a deviation from backward-induction search was positively associated with a deviation from subgame-perfect equilibrium decisions.

In CGC's econometric analysis, subjects' rules can be more precisely identified by guesses and search than by guesses or search alone (CGC, Table 7B, pp. 1760-1761). Most guesses-and-search estimates of subjects' rules refine and sharpen their guesses-only estimates. For some subjects the guesses-and-search estimate resolves a tension between guesses-only and search-only estimates, sometimes in favor of the search-only estimate (as also happens in CJ's and Costa-Gomes et al.'s 2001 analyses). However, the guess part of the likelihood has six times the weight of the search part, because CGC's theory of guesses makes precise predictions that are very unlikely to be satisfied by chance, but their theory of search is conservative given the lack of theory or evidence, imposing only weak restrictions. (Thus, refining our understanding of how cognition drives search should make search even more useful.) Even so, most subjects' decision rules can be identified from search alone, which sometimes directly reveals a subject's rule's algorithm or distinguishes her/his intended guesses from errors (CGC, Table 7A, pp. 1758-1759).

Costa-Gomes & Weizsäcker (2008) (see also Hyndman et al. 2013) reported experiments that elicited subjects' initial responses, including stated beliefs about others' decisions as well as own decisions, in a series of 14 asymmetric two-person 3×3 games. Subjects were rewarded for the correctness of their stated beliefs via a quadratic scoring rule, which is incentive compatible when the decision-maker is risk-neutral. Their econometric analysis views stated beliefs as just another kind of decision, with an error structure like that for decisions. Elicited beliefs provide another, complementary lens through which to examine strategic thinking. Most subjects' stated beliefs were close to *L2*'s, i.e. believing others are *L1*s; but most subjects' decisions were close to *L1*'s, i.e. believing others are *L0*s This suggests that subjects treated decisions and stated beliefs as unrelated tasks. It thereby challenges the widely accepted view that the game causes players' beliefs, which then cause their decisions. Instead the results suggest that a subject's decision rule is the fundamental, which in conjunction with the setting causes both her/his beliefs and decisions. Imposing an untested restriction that decisions best respond to beliefs is risky.[12]

Burchardi & Penczynski (2014) (see also Penczynski 2016, who used similar methods in hide-and-seek games) elicited responses to one of Nagel's (1995) beauty contest games. Following Cooper & Kagel (2005), each of a game's player roles was filled by an anonymously paired team, whose members could simultaneously send each other one message, with a suggested decision and a natural-language justifying text. Team members then read each other's messages and simultaneously submitted proposed team decisions, each equally likely to be implemented. (Although teammates communicate, the game between teams is played without communication.) This design gives each subject an incentive to convey her/his own thinking to her/his teammate, and then to submit the decision s/he thought was optimal.

The messages were classified by research assistants, first as either level-*k*, equilibrium, or iterated dominance; and then, if level-*k*, by lower and upper bounds on the level of reasoning and the *L0*, if any. (Possible alternative methods for classification are discussed in Houser & Xiao 2011 and Gentzkow et al. in press 2019.) Almost half of the subjects are classified as non-strategic, and Burchardi & Penczynski describe them as *L0*s. (Even with the chat data, those subjects might be more accurately described as "unclassified", as in CGC's and Fragiadakis et al.'s 2016 analyses.) The others are concentrated on *L1* and to a lesser extent *L2*. Nonstrategic and strategic subjects' *L0*s are insignificantly higher than the uniform mean of 50, with a spike at

[12] Rey-Biel (2009) partially replicates Costa-Gomes & Weizsäcker's results, while suggesting that in constant-sum games, where equilibrium strategies can be identified without fixed-point reasoning, *Equilibrium* may fit better than level-*k* or other decision rules. See also Manski & Neri's (2013) analysis of hide-and-seek games.

50 for strategic subjects. Except for *L0*s, the estimated level distribution is close to other estimates.

The studies discussed in this subsection are a proof of concept that extending the domains of both theory and experimental design to include measures of cognition more direct than conventional decisions, can greatly increase the power of experiments to identify subjects' thinking. This will require both new theory and refinements of experimental and data analysis methods, but the rewards seem likely to repay the costs.

3.4. Repeated games

Recall that a repeated game is one where a given stage game (normal- or extensive-form) is played repeatedly by a fixed pair or group of players. I discuss repeated games separately from other extensive-form games because their special structure and the possibility of infinite repetition raise issues that have led to partly separate theoretical and experimental literatures. The analyses I discuss here all have subjects repeatedly, randomly paired from a large population to play (finitely or infinitely) repeated symmetric Prisoner's Dilemmas, with symmetric information. Recall that in this case, subjects' responses to their first play of the repeated Prisoner's Dilemma can be treated as initial responses, as if the game was played in isolation.

Since Roth & Murnighan (1978), infinite repetition has been implemented in experiments by publicly announcing a constant probability that the interaction will terminate after any given period, so it never becomes public knowledge that the next period is the last. If the termination probability is low enough (or equivalently, if the implied discount factor is high enough), it is theoretically possible to support cooperation in subgame-perfect equilibrium even if the stage game has short-run incentives to cheat. By contrast, if a stage game with a unique equilibrium, such as the Prisoner's Dilemma, is repeated a finite, commonly known number of times, then the only subgame-perfect equilibrium of the repeated game is for players to play the stage-game equilibrium every period independent of the history, however inefficient that outcome may be.

Subgame-perfect equilibria that support cooperation with infinite repetition imply an "implicit contract" that defines what it means to cooperate. Cooperation is usually enforced via threats to end the relationship if either player fails to cooperate—threats made credible by the subgame-perfect equilibrium. Under weak assumptions there is a very large "Folk Theorem" set of the cooperative outcomes that are supportable in some subgame-perfect equilibrium.

As in any equilibrium analysis, players' beliefs are assumed to be focused on a particular subgame-perfect equilibrium with certainty. If the relationship lacks close precedents, as I have argued is typical in applications, players must both reach the same equilibrium via independent strategic thinking—usually without the symmetry of our favorite examples to distinguish an ideal compromise. In most equilibrium analyses there is no tolerance for error; and even trivial failures to coordinate end the relationship (exceptions include Porter 1983, van Damme 1989, Friedman & Samuelson 1994, and Embrey et al. 2013). It is clear that to be effective in real relationships, strategies must be far more robust than such equilibrium strategies. Yet theoretical work in this area focuses on characterizing the Folk Theorem set under more and more general informational assumptions, assuming subgame-perfect, sequential, or perfect Bayesian equilibrium.

Experimental work on cooperation in repeated games in economics began with Roth & Murnighan (1978) on the infinitely repeated Prisoner's Dilemma, and Selten & Stoecker (1986) on the finitely repeated Prisoner's Dilemma. The literature is very large, and I will discuss only some promising recent work. Dal Bó & Fréchette (2018) and Embrey et al. (2018) survey the experimental literatures on the infinitely and finitely repeated Prisoner's Dilemma respectively.

The papers I discuss all focus on the symmetric Prisoner's Dilemma and consider learning as well as initial responses, although I focus on the latter. They are distinguished by their use of datasets and/or meta-datasets from experiments with widely varying payoff structures, and also by their openness to the need for strategies to be robust. Their analyses replace the customary assumption that any subgame-perfect equilibrium is a possible outcome with simple criteria that use the tools of equilibrium analysis but do more justice to the strategic uncertainty subjects face in their need to coordinate on some cooperative outcome that is consistent with their incentives.

In the infinitely repeated Prisoner's Dilemma, Dal Bó & Fréchette (2011) and Blonski et al. (2011) (see also Dal Bó 2005 and Blonski & Spagnolo 2015) report new experiments, finding that the consistency of Cooperation with subgame-perfect equilibrium is normally necessary but far from sufficient for cooperation. Dal Bó & Fréchette (2011) also find that subjects are more likely to Cooperate when Cooperative strategies are risk-dominant (Harsanyi & Selten 1988) in a reduced repeated Prisoner's Dilemma stage game with its strategies restricted to the Grim trigger strategy, which supports Cooperation in every period via threats to terminate the relationship if either player ever fails to Cooperate, and Always-Defect. Blonski et al. (2011) derive similar criteria from plausible axioms, including one that requires that players Defect when the payoff for Cooperating while one's partner Defects is very low, and Cooperate when that payoff is very high. This leads Blonski et al. to a measure of the risk-dominance of cooperation like Dal Bó & Fréchette's, which with their other axioms implies a critical discount factor—higher than the smallest one that makes Cooperation consistent with subgame-perfect equilibrium—above which Cooperation is predicted to prevail. Their experiments, together with a meta-analysis of previous data, confirm that discount factors above the critical level tend to yield higher rates of cooperation. Dal Bó & Fréchette's (2011) and Blonski et al.'s (2011) conclusions are generally confirmed and strengthened in Dal Bó & Fréchette's (2018) comprehensive meta-analysis.

That said, these studies' criteria leave a substantial fraction of cooperation unexplained. Further, although the symmetric Prisoner's Dilemma is a sensible place to start, it may make coordination on cooperative outcomes unrepresentatively easy. Future work should try to extend these analyses to asymmetric repeated games without obvious ideal solutions, seeking criteria that reflect the difficulties asymmetrically situated subjects face.

Finally, establishing and maintaining cooperation is cognitively taxing, and a full understanding will require studying cognition at the individual level. Breitmoser (2015) took a first step in this direction by analyzing individual subjects' strategies in several experiments, including Dal Bó & Fréchette's (2011) and Blonski et al.'s (2011). He finds that if the discount factor is above Blonski et al.'s critical level, subjects' behaviors after their first responses are well summarized by strategies he calls Semi-Grim: Cooperate after both Cooperate, Defect after both Defect, and otherwise (no matter who Cooperated) randomize. Combining this with a model of subjects' first responses might yield a simple criterion that maps the distribution of individual behaviors into a prediction of whether cooperation will prevail. However, Dal Bó & Fréchette's (2018, Result 6) meta-study of individual subjects' behavior reached a more nuanced conclusion, in which the strategies All-Defect, Grim, and Tit-for-Tat account for most subjects' behavior.

Turning to the finitely repeated Prisoner's Dilemma, Embrey et al. (2018) provide a unifying meta-analysis of experimental research and report a new experiment. Having a publicly known last period adds a new behavioral pattern, in that subjects invariably Cooperate until some threshold near the end, and Defect from then on. The threshold varies across individuals, and before subjects start to reach it they are in a situation much like the infinitely repeated Prisoner's

Dilemma. Embrey et al. find, like others before them, that a longer finite horizon increases the rate of initial cooperation. They suggest that this occurs not because of the cognitive difficulty of a longer backward induction, but because a longer horizon increases the value of conditionally cooperative strategies. They propose to capture this trade-off by the size of the basin of attraction of the Always-Defect strategy, defined as above.

Embrey et al.'s (2018) and previous analyses of subjects' thresholds are reminiscent of some of the analyses for other kinds of extensive-form game (Section 3.2). They again suggest the possibility of a dynamic level-*k* model, in which it's taken as given that subjects look ahead only a few periods, but also respond to the value of conditionally cooperative strategies (see Kawagoe & Takizawa 2009, Ho & Su 2013, and García-Pola et al. 2018). A possible ultimate goal is a model that uses insights like Dal Bó & Fréchette's (2011, 2018) and Blonski et al.'s (2011) to give similar explanations of subjects' behavior in the infinitely repeated Prisoner's Dilemma and the finitely repeated Prisoner's Dilemma well before the end, while combining Embrey et al.'s insights with those from other analyses of extensive-form games to capture the pattern of unraveling at subjects' heterogeneous thresholds near the end.

These analyses have only begun to gather the empirical knowledge needed for a better theory of cooperation in repeated games.[13] But this may well be another case where experiments help to break a theoretical roadblock. Methods like those discussed in Section 3.3 may also be useful.

[13] Crawford (2002, pp. 8-10; 2016, p. 142) discusses another puzzle, which highlights the use of repeated-game strategies for what Camerer et al. (2002) called "strategic teaching", whereby subjects who interact repeatedly with the same partners deviate from their myopically optimal decisions to try to influence their partners in ways that will benefit them in the future.

4. INTERACTIONS WITH COMMUNICATION

I now consider interactions with communication. The role of communication in relationships underlies many insights about fostering cooperation (Ostrom 2010) and also some about preventing it, perhaps most notably the prohibition of conspiracies in restraint of trade in antitrust law (Genesove & Mullin 2001, Andersson & Wengström 2007). In each case casual empiricism and field evidence provide clear rationales, but it seems likely that a more systematic game-theoretic analysis would help us better understand how communication affects outcomes.

Some of this section's arguments draw on Crawford (2017b), my unpublished Nancy Schwartz Lecture. Recall (Section 2) that I distinguish between the communication phase and the *underlying* game or stage games in which decisions directly determine outcomes. Communication is via cheap talk messages, and may be structured or unstructured, limited or unlimited in duration, and via abstract signals or natural language. I avoid treatments in which communication is face-to-face, because they blend its purely strategic effects, which are my focus, with the uncontrolled social preferences that face-to-face treatments commonly evoke.

In experiments, as in the field, communication can make an enormous difference, even with symmetric information. Crawford (1998, 2017b), Camerer (2003, Ch. 7), and Blume & Ortmann (2007) survey experiments showing that communication via abstract signals is often effective in promoting coordination and cooperation (see also Embrey et al. 2013 and Arechar et al. 2017). And a number of recent papers, includng Valley et al. (2002), McGinn et al. (2003), Weber & Camerer (2003), Ellingsen & Johannesson (2004), Charness & Dufwenberg (2006), Brandts & Cooper (2007), McGinn et al. (2012), Cooper & Kühn (2014), Awaya & Krishna (2016), and Dugar & Shahriar (2018) (see also Charness's 2012 survey), suggest that communication via natural-language messages, even limited in duration, is even more effective.

The size and robustness of these effects, particularly of natural-language messages, compel our attention. Moreover, even if all that is needed is communication of intentions, which in most theory and experiments is modeled by taking signals to be pre-defined abstract decision labels, not all applications come with obvious signals. Consider what may be the most famous example of abstract signals in American history, or at least American literature, Longfellow's (1861) "one if by land and two if by sea" code in his poem "Raul Revere's Ride". In theory the non-obvious meanings would be part of equilibrium beliefs (or not, in a babbling equilibrium), but the poem is about practical men, who defined them in the only way they could have, in natural language.

Despite the importance of communication, particularly via natural language, standard theoretical frameworks, equilibrium or nonequilibrium, do not tell us how to model them. In games with symmetric information, subgame-perfect equilibrium limits communication to a symmetry-breaking role, as in Farrell's (1987) analysis of unlimited preplay communication via non-binding announcements of intentions before a one-shot Battle of the Sexes game. To model the difficulty of coordination, Farrell restricted attention to the *symmetric* mixed-strategy subgame-perfect equilibrium of the entire game, including the communication phase. He assumed that announcements of coordinated decision labels would lead players to play the labelled equilibrium in the underlying game. Symmetric equilibrium communication yields more efficient coordination than the symmetric mixed-strategy equilibrium of the underlying game without communication. But surprisingly, even unlimited communications does not lead to fully efficient coordination.[14] Aside from such symmetry-breaking, in an equilibrium analysis with

[14] Crawford & Haller (1990) justify Farrell's symmetry assumption for pure coordination games. Farrell (1988) showed, assuming rationalizability (the natural assumption to avoid question-begging) with a refinement to ensure that messages are understood, that preplay communication need not assure equilibrium in the underlying game. Ellingsen & Östling (2010), Crawford (2017a), and Ellingsen et al. (2018) reconsider different aspects of Farrell's analysis using level-*k* models.

symmetric information there is nothing players can accomplish via communication that they could not also accomplish without it: They have no private information to communicate, and subgame-perfect equilibrium's rational expectations make anything else they can say redundant.

With regard to natural-language messages versus abstract signals, in equilibrium models, with or without symmetric information, the meanings of cheap talk messages are determined by the beliefs that support an equilibrium, and an equilibrium's possible substantive outcomes are independent of the messages that lead to them, as long as there are enough different possible messages (Crawford & Sobel 1982). Even in nonequilibrium models like Crawford's (2003) or Ellingsen & Östling's (2010) level-*k* analyses of communication of intentions, the meanings of messages are anchored in an assumed truthful *L0*, with truthfulness defined by an unmodeled pre-existing common language. Thus, existing models of communication have nowhere to "plug in" the difference between natural-language messages and abstract signals.

It might be that natural-language messages are just inherently more credible, even if not face-to-face. If so they are not truly cheap talk: just another kind of costly signaling.

But let me suggest additional sources of the effectiveness of natural-language messages that do not depend on invoking such social credibility effects. Crawford (2016, pp. 143-5) gives an example in which natural-language messages are better at conveying an understanding of strategic issues. In a two-person Stag Hunt game, each player does weakly better if his partner chooses high effort (Stag), without regard to his own intentions. Aumann (1990) argued that a message of intent to play Stag is not credible on the grounds that it is not *self-signaling*, in Farrell & Rabin's (1996) sense that it is sent when and only when the sender intends to do as he said. Crawford (2016, p. 144) argued that Aumann's exclusive reliance on the logic of equilibrium in this game is inappropriate: "When uncertainty about others' thinking is of the essence, it is unlikely that intelligent people will interpret a sender's message as if there were no chance whatsoever that it would influence equilibrium selection or whether players' choices are even in equilibrium."[15] In fact even one-sided communication via abstract signals of intent to play Stag is usually effective in Stag Hunt (Cooper et al. 1992, Charness 2000, Ellingsen et al. 2018; but see Clark et al. 2001).[16] But Crawford (2016, p. 145) argued further that it would be even more effective to send a natural-language message conveying an understanding of the strategic issues, such as: "I can see, as I'm sure you can, that the best possible outcome would be for both of us to play Stag. I realize that Stag is risky for you, as it is for me. But despite the risk, I think Stag's higher potential payoff makes it a better bet. I therefore plan to play Stag, and I hope you will too." And there is now experimental evidence, culminating with Dugar & Shahriar (2018), that in Stag Hunt natural-language messages are even more effective than abstract messages.

This example highlights another problem with modeling the effects of natural-language communication in games. The logical omniscience normally assumed in equilibrium or nonequilibrium analyses makes it theoretically possible for players to mentally simulate any natural-language message or dialogue that does not rely on others' private information, and to form the same beliefs that a real message from another player would have led to. Behaviorally, however, such mental simulations are no substitute for real communication. Existing theory has no way to model such differences, but in experiments they are reflected in subjects' behavior.

[15] Farrell (1988) and Baliga & Morris (2002) discuss this issue in much more detail.

[16] In Charness's (2000) experiments, the apparent credibility of one-sided self-committing messages in Stag Hunt is much higher when a subject's message precedes her/his action than when the timing is reversed, even though the time-sequencing does not affect what her/his partner observes; see also Farrell (1988, p. 213).

Finally, natural-language messages are more effective in repairing broken relationships. Crawford (2017b) gives an example of a long-term relationship, governed by an implicit contract you believe that both you and your partner understand. What do you do if your partner does something that goes against what you thought the agreement was? Without language, all you can do, short of ending the relationship, is to signal your displeasure via tit-for-tat, hoping that your partner will understand and return to cooperating. Such a tactic might work if the ideal agreement is clear, as in Crawford's (2016, p. 142) example; but otherwise it would be hopeless. If instead you and your partner can communicate, but only via abstract messages from a pre-set list, you might do somewhat better; but only if good agreements are simple and reaching them doesn't require complex adjustments. But finally, if you and your partner can communicate via a natural-language message or dialogue, you have a genuine chance to restore cooperation, perhaps starting like this: "I value our relationship, and I believe you are trying to cooperate. But what you just did was inconsistent with what I thought we had agreed. [Elaborates….] Please help me understand your thinking." A successful dialogue needs to reaffirm parties' wish to continue the relationship and show them how to adjust the contract and atone for the breach, both easy with natural language, but hard to imagine doing via abstract messages.

More generally, natural-language communication appears to play essential roles in allowing people to economize on communication costs—compare trying to persuade a superior with decision authority to make the obviously-to-you right decision by filing a single all-or-nothing memo versus by an open-ended back-and-forth conversation—and in crafting robust responses to the difficulty of coordination. Fully understanding the role of communication in establishing and maintaining cooperation in relationships will require creating new theory in which the difference between natural-language and abstract messages matters, rather than "black-boxing" it. Relaxing equilibrium assumptions in empirically grounded ways should yield theoretical frameworks that create more realistic roles for communication and distinguish when and how natural-language communication is more effective than abstract signaling. Models in which flexible dialogues economize on communication costs or models in which the form of language matters (as in Selten & Warglien 2007, Hong et al. 2016, and Gibbons et al. 2017) seem especially promising.

Fully understanding the role of communication in relationships will also require new experiments. An obvious next step would be more analyses of the repeated Prisoner's Dilemma with natural-language communication, building on those that have begun for extensive-form games (Section 3.2) and tacit repeated Prisoner's Dilemmas (Section 3.4), seeking criteria that indicate the likelihood of cooperation with communication along the lines of Dal Bó & Fréchette's (2011, 2018) and Blonski et al.'s (2011). Helpful adjuncts will include enhanced experimental methods to study cognition like those discussed in Section 3.3 and methods for classifying natural-language texts such as Cooper & Kagel's (2005), Houser & Xiao's (2011), Burchardi & Penczynski's (2014), and Gentzkow et al.'s (in press 2019).[17]

5. CONCLUDING REMARKS

This review has discussed selected recent work in experimental game theory, with particular attention to experiments whose lessons are relevant to establishing and maintaining cooperation in human relationships and the role of communication in doing so. These are questions of central importance, where the gap between theory and experience and the role of experiments in closing it both seem large. I hope that continuing the dialogue between theory and experiment in this

[17] Gentzkow et al. (in press 2019) discuss how analyzing natural-language text differs from analyzing more conventional kinds of data and gives an overview of econometric methods and applications. A typical analysis using the methods they discuss begins by editing the data to reduce dimensionality, then using the data to predict the unknown variables of interest, and then using the predictions much as one would use standard data. Their discussion includes causal inference as well as prediction.

area will help to explain why and how observed patterns of cooperation vary so much with the possibilities for communication, and why and how communication via natural-language messages or dialogues is more effective than communication via abstract signals.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I owe thanks for valuable advice to many people, especially Johannes Abeler, Colin Camerer, Gary Charness, Miguel Costa-Gomes, Tore Ellingsen, Guillaume Fréchette, Drew Fudenberg, Nagore Iriberri, Willemien Kets, Juanjuan Meng, Stephen Morris, Muriel Niederle, Robert Östling, Adam Sanjurjo, Joel Sobel, Séverine Toussaert, and John Vickers. The research leading to these results received primary funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339179. The contents reflect only the authors' views and not the views of the ERC or the European Commission, and the European Union is not liable for any use that may be made of the information contained therein. The University of Oxford, All Souls College, and the University of California, San Diego also provided research support.

LITERATURE CITED

Agranov M, Potamites E, Schotter A, Tergiman C. 2012. Beliefs and endogenous cognitive levels: An experimental study," *Games and Econ. Behav.* 75(2), 449-63

Alaoui L, Penta A. 2016. Endogenous depth of reasoning. *Rev. of Econ. Stud.* 83(4), 1297–33

Andersson O, Wengström E. 2007. Do antitrust laws facilitate collusion? Experimental evidence on costly communication in duopolies. *Scand. J. of Econ.* 109(2): 321–39

Arad A, Rubinstein A. 2012. The 11-20 money request game: A level-*k* reasoning study. *American Econ. Rev.* 102(7), 3561-73

Aradillas-Lopez A, Tamer E. 2008. The identification power of equilibrium in simple games. *J. of Bus. & Econ. Statistics* 26(3): 261–83

Arechar AA, Dreber A, Fudenberg D, Rand DG. 2017. I'm just a soul whose intentions are good: The role of communication in noisy repeated games. *Games and Economic Behavior* 104(1): 726-43

Aumann R. 1990. Nash-equilibria are not self-enforcing. In *Economic Decision-Making: Games, Econometrics and Optimisation*, ed. JJ Gabszewicz, JF Richard, L Wolsey, pp. 201-06. Amsterdam and New York: North Holland

Awaya Y, Krishna V. 2016. On communication and collusion. *American Econ. Rev.*106(2): 285–315.

Baliga S, Morris S. 2002. Co-ordination, spillovers, and cheap talk. *J. of Econ. Theory* 105(2): 450-468

Bernheim BD. 1984. Rationalizable strategic behavior. *Econometrica* 52(4): 1007–28

Binmore K, McCarthy J, Ponti G, Samuelson L, Shaked A. 2002. A backward induction experiment. *J. of Econ. Theory* 104(1): 48–88

Blonski M, Ockenfels P, Spagnolo G. 2011. Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Econ. Journal: Microeconomics* 3(3): 164–92

Blonski M, Spagnolo G. 2015. Prisoners' other dilemma. *International J. of Game Theory* 44(1): 61–81

Blume A, Ortmann A. 2007. The effects of costless pre-play communication: Experimental evidence from games with pareto-ranked equilibria. *J. Econ. Theory* 132(1): 274–90.

Brandenburger A. 1992. Knowledge and equilibrium in games. *J. of Econ. Perspectives* 6(4): 83–101

Brandts J, Cooper DJ. 2007. It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. J. of the European Econ. Association 5(6): 1223–68

Breitmoser Y. 2012. Strategic reasoning in p-beauty contests. *Games and Econ. Behav*. 75(2): 555-569

Breitmoser Y. 2015. Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma. *American Econ. Rev.* 105(9): 2882–2910

Brocas I, Carrillo JD, Camerer CF, Wang SW. 2014. Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Rev. of Econ. Studies* 81(3): 944–70

Burchardi KB, Penczynski SP. 2014. Out of your mind: Eliciting individual reasoning in one shot games. *Games and Econ. Behav.* 84(1): 39–57.

Camerer CF. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press

Camerer CF. 2008. The case for mindful economics. In *Perspectives on the Future of Economics: Positive and Normative Foundations*, ed. A Caplin, A Schotter, Chapter 2. Oxford, UK: Oxford Univ. Press.

Camerer CF. 2013. A review essay about “Foundations of Neuroeconomic Analysis” by Paul Glimcher. *J. of Econ. Lit.* 51(4): 1155–82

Camerer CF, Ho T-H, Chong J-K. 2002. “Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *J. of Econ. Theory* 104(1): 137–88

Camerer CF, Ho T-H, Chong J-K. 2004. A cognitive hierarchy model of games. *Quarterly J. of Econ.* 119(3): 861–98

Camerer CF, Johnson EJ. 2004. Thinking about attention in games: Backward and forward induction. In *The Psychology of Economic Decisions, Volume Two: Reasons and Choices*, ed. I Brocas, J Carrillo, pp. 111-29. Oxford, UK: Oxford Univ. Press

Camerer CF, Johnson EJ, Rymon T, Sen, S. 1993. Cognition and framing in sequential bargaining for gains and losses. In *Frontiers of Game Theory*, ed K Binmore, A Kirman, P Tani, pp. 27-47. Cambridge, MA: MIT Press

Camerer CF, Loewenstein G, Prelec, D. 2005. Neuroeconomics: How neuroscience can inform economics. *J. of Econ. Lit.* 43(1): 9-64.

Charness G. 2000. Self-serving cheap talk: A test of Aumann’s conjecture. *Games and Econ. Behav.* 33(2): 177–94

Charness G. 2012. Communication in Bargaining Experiments. In *The Oxford Handbook of Economic Conflict Resolution*, ed R Croson and GE Bolton, pp. 7-18. Oxford, UK: Oxford Univ. Press

Charness G, Dufwenberg, M. 2006. Promises and partnership. *Econometrica* 74(6): 1579–1601

Clark K, Kay S, Sefton M. 2001. When Are Nash Equilibria Self-Enforcing? An Experimental Analysis. *International J. of Game Theory* 29(4): 495–515

Cooper DJ, Kagel JH. 2005. Are two heads better than one? Team versus individual play in signaling games. *American Econ. Rev.* 95(3): 477-508

Cooper DJ, Kagel JH. 2008. Learning and transfer in signaling games. *Econ. Theory* 34(3): 415–39

Cooper DJ, Kühn K-U. 2014. Communication, renegotiation, and the scope for collusion. *American Econ. J.: Microeconomics* 6(2): 247–78

Cooper R, DeJong DV, Forsythe R, Ross TW. 1992. Communication in Coordination Games. *Quarterly J. of Econ.* 107(2): 739–71

Costa-Gomes MA, Crawford VP. 2006. Cognition and behavior in two-person guessing games: An experimental study. *American Econ. Rev.* 96(5): 1737–68

Costa-Gomes MA, Crawford VP, Broseta B. 1998. Cognition and behavior in normal-form games: An experimental study. University of California San Diego Discussion Paper 98–22

Costa-Gomes MA, Crawford VP, Broseta B. 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69(5): 1193-1235

Costa-Gomes MA, Weizsäcker G. 2008. "Stated beliefs and play in normal-form games. *Rev. of Econ. Studies* 75(3): 729-62

Crawford VP. 1998. A survey of experiments on communication via cheap talk. *J. of Econ. Theory* 78(2): 286-98

Crawford VP. 2002. Introduction to experimental game theory. *J. of Econ. Theory* 104(1): 1–15

Crawford VP. 2003. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Econ. Rev.* 93(1): 133–49

Crawford VP. 2008. Look-ups as the windows of the strategic soul: Studying cognition via information search in game experiments. In *Perspectives on the Future of Economics: Positive* *and Normative Foundations*, ed. A Caplin, A Schotter, Chapter 10. Oxford, UK: Oxford Univ. Press

Crawford VP. 2016. New directions for modelling strategic behavior: Game-theoretic models of communication, coordination, and cooperation in economic relationships. *J. of Econ. Perspectives* 30(4): 131–50

Crawford VP. 2017a. Let's talk it over: Coordination via preplay communication with level-*k* thinking. *Research in Econ*. 71: 20–31

Crawford VP. 2017b. Modeling strategic communication: From rendezvous and reassurance to trickery and puffery. Nancy L Schwartz Memorial Lecture; slides at http://econweb.ucsd.edu/~v2crawford/SchwartzLecture17.pdf ; video at https://www.kellogg.northwestern.edu/news-events/lecture/schwartz/speakers/2010-2019/2017.aspx

Crawford VP, Costa-Gomes MA, Iriberri N. 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *J. of Econ. Literature* 51(1): 5–62

Crawford VP, Iriberri N. 2007. Level-*k* auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica* 75(6): 1721-70

Crawford VP, Sobel J. 1982. Strategic information transmission. *Econometrica* 50(6): 1431–51

Dal Bó P. 2005. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Econ. Rev.* 95 (5): 1591–604

Dal Bó P, Fréchette GR. 2011. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Econ. Rev.* 101(1): 411–29

Dal Bó P, Fréchette GR. 2018. On the determinants of cooperation in infinitely repeated games: A survey. *J. of Econ. Literature* 56(1): 60-114

Davis DD, Holt CA. 1993. *Experimental Economics*. Princeton, NJ: Princeton Univ. Press

Dugar S, Shahriar Q. 2018. Restricted and free-form cheap-talk and the scope for efficient coordination. *Games and Econ. Behav*.109: 294-310

Ellingsen T, Johannesson M. 2004. Promises, threats and fairness. *Econ. J.* 114(495), 397–420

Ellingsen T, Östling R. 2010. When does communication improve coordination. *American Econ. Rev.* 100(4): 1695–1724

Ellingsen T, Östling R, Wengström E. 2018. How does communication affect beliefs in one-shot games with complete information? *Games and Econ. Behav*. 107: 153-81

Embrey M, Fréchette GR, Stacchetti E. 2013. An experimental study of imperfect public monitoring: Efficiency versus renegotiation-proofness. http://cess.nyu.edu/frechette/print/Embrey_2013a.pdf

Embrey M, Fréchette GR, Yuksel S. 2018. Cooperation in the finitely repeated Prisoner's Dilemma. *Quarterly J. of Econ.* 133(1): 509-551

Eyster E, Rabin M. 2005. Cursed equilibrium. *Econometrica* 73(5): 1623-72

Farrell J. 1987. Cheap talk, coordination, and entry. *RAND J. of Econ.* 18(1): 34–39

Farrell J. 1988. Communication, coordination, and Nash equilibrium. *Econ. Letters* 27(3): 209–14

Farrell J, Rabin M. 1996. Cheap talk. *J. of Econ. Perspectives* 10(3): 103–18

Fragiadakis DE, Knoepfle dT, Niederle M. 2016. Who is Strategic? https://web.stanford.edu/~niederle/Who_Is_Strategic_FKN_09_12_16.pdf

Friedenberg A, Kneeland T, Kets W. 2018. Is Bounded Rationality Driven by Limited Ability?.

http://u.arizona.edu/~afriedenberg/car.pdf

Friedman JW, Samuelson L. 1994. The ‘Folk Theorem’ for repeated games and continuous decision rules. In *Problems of Coordination in Economic Activity*, ed JW Friedman, Chapter 6. Berlin: Springer Verlag

Fudenberg D, Liang A. In press 2019. Predicting and understanding initial play. *American Econ. Rev.* http://economics.mit.edu/files/14505

García-Pola B, Iriberri N, Kovářík J. 2018. Non-equilibrium play in Centipede games. https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxuYWdvcmVpcmliZXJyaXxneDozYzJmYzY4MTNjYzU3MGE5

Genesove D, Mullin WP. 2001. Rules, communication, and collusion: Narrative evidence from the Sugar Institute case. *American Econ. Rev.* 91(3): 379–98

Gentzkow M, Kelly BT, Taddy M. In press 2019. Text as Data. *J Econ. Literature* ??(?): ??-??

Georganas S, Healy PJ, Weber RA. 2015. On the persistence of strategic sophistication. *J. of Econ. Theory* 159A(1): 369-400

Gibbons R, LiCalzi M, Warglien M. 2017. What situation is this? Coarse cognition and behavior over a space of games. Working Papers 09, Department of Management, Università Ca' Foscari Venezia. https://ideas.repec.org/p/vnm/wpdman/142.html

Gill D, Prowse V. 2016. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *J. of Political Econ*. 124(6): 1619-76

Goeree JK, Holt CA. 2001. Ten little treasures of game theory and ten intuitive contradictions. *American Econ. Rev.* 91(5): 1402-22

Goeree JK, Holt CA, Palfrey TR. 2005. Regular quantal response equilibrium. *Experimental Econ* 8(4): 347–67

Gul F, Pesendorfer W. 2008. The case for mindless economics. In *Perspectives on the Future of Economics: Positive and Normative Foundations*, ed. A Caplin, A Schotter, Chapter 1. Oxford, UK: Oxford Univ. Press

Haile PA, Hortaçsu A, Kosenok G. 2008. On the empirical content of quantal response equilibrium. *American Econ. Rev.* 98(1): 180-200

Harsanyi JC, Selten R. 1988. *A General Theory of Equilibrium Selection in Games.* Cambridge, MA: MIT Press.

Ho T-H, Camerer CF, Weigelt K. 1998. Iterated dominance and iterated best response in experimental '*p*-Beauty Contests'. *American Econ. Rev.* 88(4): 947-69

Ho T-H, Su X. 2013. A dynamic level-k model in sequential games. *Management Science* 59(2), 452-69

Houser D, Xiao E. 2011. Classification of natural language messages using a coordination game. *Experimental Econ.* 14(1): 1–14

Hyndman K, Terracol A, Vaksmann J. 2013. Beliefs and (in)stability in normal-form games. https://www.hyndman-honhon.com/uploads/1/0/9/9/109960863/htv-stability.pdf

Johnson EJ, Camerer CF, Sen S, Rymon T. 2002. Detecting failures of backward induction: Monitoring information search in sequential bargaining. *J. of Econ. Theory* 104(1): 16-47

Kawagoe T, Takizawa H. 2010. Level-*k* analysis of experimental Centipede games. *J. of Econ. Behav. and Organization* 82(2-3): 548-66

Keynes JM. 1936. *The General Theory of Employment, Interest and Money*. New York: Harcourt Brace and Co.

Kneeland T. 2015. Identifying higher-order rationality. *Econometrica* 83(5): 2065–79

Kreps DM, Milgrom P, Roberts J, Wilson R. 1982. Rational cooperation in the finitely repeated prisoners' dilemma. *J. of Econ. Theory* 27(2): 245-52

Longfellow, HW. 1861. “Paul Revere's Ride”, *The Atlantic Monthly*. pp. 27ff https://www.theatlantic.com/magazine/archive/1861/01/paul-revere-s-ride/308349/

Manski CF, Neri C. 2013. First- and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Econ. Behav*. 81(1): 232-54

McGinn KL, Milkman KL, Nöth M. 2012. Walking the talk in multiparty bargaining: An experimental investigation. *J. of Econ. Psychology* 33(1): 278-91

McGinn KL, Thompson L, Bazerman MH. 2003. Dyadic processes of disclosure and reciprocity in bargaining with communication. *J. of Behav. Decision Making* 16(1): 17–34

McKelvey RS, Palfrey TR. 1992. An experimental study of the Centipede game. *Econometrica* 60(4): 803-36

McKelvey RS, Palfrey TR. 1995. Quantal response equilibria for normal-form games. *Games and Econ. Behav.* 10(1): 6-38

McKelvey RS, Palfrey TR. 1998. Quantal response equilibria for extensive form games. *Experimental Econ*. 1(1): 9–41

Mullainathan S, and Spiess J. 2017. Machine learning: An applied econometric approach. J. of Econ. Perspectives 31(2): 87-106

Nagel R. 1995. Unraveling in guessing games: An experimental study. *American Econ. Rev.* 85(5): 1313-26

Ostrom E. 2010. Beyond markets and states: Polycentric governance of complex economic systems. American Econ. Review 100(3): 641–72

Payne J, Bettman J, Johnson EJ. 1993. *The Adaptive Decision Maker.* Cambridge, U.K.: Cambridge Univ. Press

Pearce DG. 1984. Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52(4): 1029–50

Penczynski SP. 2016. Strategic thinking: The influence of the game. *J. of Econ. Behavior & Organization* 128: 72-84

Porter RH. 1983. Optimal cartel trigger price strategies. *J. of Econ. Theory* 29(2): 313–38

Rankin F, Van Huyck JB, Battalio RC. 2000. Strategic similarity and emergent conventions: Evidence from similar Stag Hunt games. *Games and Econ. Behavior* 32(2): 315–37

Rey-Biel, Pedro. 2009. “Equilibrium play and best response to (stated) beliefs in normal form games,” *Games and Econ. Behavior* 65(2): 572-85

Rosenthal, Robert W. 1993. Rules of thumb in games. *J. of Econ. Behav. and Organization* 22(1): 1-13

Roth AE, Murnighan JK 1978. Equilibrium behavior and repeated play of the Prisoner’s Dilemma. *J. of Math. Psychology* 17(2): 189–98

Rubinstein A. 2007. Instinctive and cognitive reasoning: A study of response times. *Econ. J.* 117(523): 1243-59

Schelling TC. 1960. *The Strategy of Conflict*. Cambridge, MA; Harvard Univ. Press

Selten R. 1991. Anticipatory learning in two-person games. In *Game Equilibrium Models*, ed R Selten. Berlin: Springer-Verlag

Selten R, Stoecker R. 1986. End behavior in sequences of finite Prisoner's Dilemma supergames: A learning theory approach. *J. of Econ. Behav. and Organization* 7(1): 47-70

Stahl DO, Wilson, PR. 1994. Experimental evidence on players' models of other players. *J. of Econ. Behav. and Organization* 25(3): 309-27

Stahl DO, Wilson, PR. 1995. On players' models of other players: Theory and experimental evidence. *Games and Econ. Behav.* 10(1): 218-54

Valley K, Thompson L, Gibbons R, Bazerman MH. 2002. How communication improves efficiency in bargaining games. *Games and Econ. Behav.* 38(1): 127–55

van Damme E. 1989. Stable equilibria and forward induction. *J. of Econ. Theory* 48(2): 476–96

Van Huyck JB, Battalio RC. 2002. Prudence, justice, benevolence, and sex: Evidence from similar bargaining games. *J. of Econ. Theory* 104(1): 227–46

Wang JT-Y, Spezio M, Camerer CF. 2010. Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth-telling and deception in sender-receiver games. *American Econ. Review* 100(3): 984–1007

Weber RA, Camerer CF. 2003. Cultural conflict and merger failure: An experimental approach. *Management Science* 49(4): 400–15