Decision
2014, Vol. 1, No. 1, 35–58

© 2013 American Psychological Association

2325-9965/14/$12.00 DOI: 10.1037/dec0000001

Explaining Strategic Coordination: Cognitive Hierarchy Theory, Strong Stackelberg Reasoning, and Team Reasoning

Andrew M. Colman, Briony D. Pulford, and Catherine L. Lawrence
University of Leicester

In common interest games, players generally manage to coordinate their actions on mutually optimal outcomes, but orthodox game theory provides no reason for them to play their individual parts in these seemingly obvious solutions and no justification for choosing the corresponding strategies. A number of theories have been suggested to explain coordination, among the most prominent being versions of cognitive hierarchy theory, theories of team reasoning, and social projection theory (in symmetric games). Each of these theories provides a plausible explanation but is theoretically problematic. An improved theory of strong Stackelberg reasoning avoids these problems and explains coordination among players who care about their co-players' payoffs and who act as though their co-players can anticipate their choices. Two experiments designed to test cognitive hierarchy, team reasoning, and strong Stackelberg theories against one another in games without obvious, payoff-dominant solutions suggest that each of the theories provides part of the explanation. Cognitive hierarchy Level-1 reasoning, facilitated by a heuristic of avoiding the worst payoff, tended to predominate, especially in more complicated games, but strong Stackelberg reasoning occurred quite frequently in the simpler games and team reasoning in both the simpler and the more complicated games. Most players considered 2 or more of these reasoning processes before choosing their strategies.

Keywords: cognitive hierarchy theory, coordination, social projection theory, strong Stackelberg reasoning, team reasoning

Supplemental materials: http://dx.doi.org/10.1037/dec0000001.supp

As Schelling (1960, chap. 3) was the first to demonstrate empirically, human decision makers are remarkably adept at coordinating their actions and their expectations of one another's actions, but it is surprisingly difficult to explain how they achieve this. Game theory provides the appropriate framework for formulating coordination problems rigorously, but although this helps to clarify the problem, it also exposes what is arguably game theory's most conspicuous limitation, namely its indeterminacy—its failure to generate determinate solutions to many games, the prime examples being games of common interest.

Consider the following problem of coordination between doubles partners in a tennis match. The server has to choose between aiming the service wide or down the center line, and the server's partner has to prepare to intercept a service return from either a wide or center line service. Let us assume that the partners have not discussed the service but are twice as likely to win the point if both choose the wide rather than the center option and have no chance of winning the point if they choose different options, given the particular circumstances at the time. If both players know all this, then they are involved in a coordination problem with a strategic struc-

---

This article was published Online First December 2, 2013.

Andrew M. Colman, Briony D. Pulford, and Catherine L. Lawrence, School of Psychology, University of Leicester, Leicester, United Kingdom.

Catherine L. Lawrence is now at Bangor University.

We are grateful to the Leicester Judgment and Decision Making Endowment Fund (Grant RM43G0176) for support in the preparation of this article, to Joachim Krueger and Jerome Busemeyer for helpful comments on earlier drafts, and to the University of Leicester for granting study leave to the second author.

Correspondence concerning this article should be addressed to Andrew M. Colman, School of Psychology, University of Leicester, Leicester LE1 7RH, United Kingdom. E-mail: amc@le.ac.uk

ture corresponding to the Hi-Lo game shown in Figure 1, where *wide* corresponds to *H* and *center* to *L*.

In the payoff matrix depicted in Figure 1, Player 1 chooses a strategy represented by a row, either *H* or *L* and, independently of Player 1, Player 2 chooses a strategy represented by a column, either *H* or *L*. The outcome of the game is one of the four cells where the chosen strategies intersect. The numbers in each cell represent the payoffs to Player 1 and Player 2 in that order. In this game, it is in both players' interests to coordinate their strategy choices on either the (*H*, *H*) or the (*L*, *L*) outcome, and (*H*, *H*) is obviously better for both than (*L*, *L*); this is represented by the payoffs of 2 to each player in (*H*, *H*), 1 to each (*L*, *L*), and 0 to each in outcomes in which they fail to coordinate. More generally, any 2 × 2 game with payoffs of (a, a) and (b, b) in the main diagonal and zero elsewhere is a Hi-Lo game, provided that $0 < b < a$, and all such games are strategically equivalent.

Like any other game, the Hi-Lo game is an idealized abstraction of a potentially unlimited number of social interactions sharing a common strategic structure. To illustrate this point, one further radically different example of a Hi-Lo interaction will suffice. Consider a scenario in which three children are trapped in a burning building, two of them in one room and the third in a second room some distance away. A neighbor breaks in and has just enough time to rescue either the two children in the first room or the single child in the second room, but the rescue can succeed only if another neighbor with a fire extinguisher, who has found a different point of entry, heads straight for the same room. If both neighbors go to the first room, then the two children in it will be rescued, and if both go to the second room, then the single child in that room will be rescued; but if each neighbor goes to a different room, then none of the children will be rescued. If both neighbors know all this, then the strategic structure is once again the Hi-Lo game shown in Figure 1, with the first room corresponding to *H* and the second room to *L*. As in the previous example, it is in both players' interests to coordinate their strategy choices on either the (*H*, *H*) or the (*L*, *L*) outcome, and both prefer (*H*, *H*) to (*L*, *L*) if their objective is to save as many lives as possible. The game model strips out complicating factors such as dynamic visual monitoring and adjustment in tennis or communication via cell phones in a fire emergency. It is not difficult to think of other lifelike examples of the Hi-Lo game; see Bacharach (2006, pp. 36–42) and Sugden (2005, pp. 181–182) for several further suggestions.

		Player 2	
		H	*L*
Player 1	*H*	2, 2	0, 0
	L	0, 0	1, 1

Figure 1. The Hi-Lo game, with a payoff-dominant Nash equilibrium at (*H, H*) and a dominated equilibrium at (*L, L*).

The Hi-Lo game is the simplest example of a *common interest game*, a class of games that illustrate most starkly the inability of game theory to explain strategic coordination. A common interest game is one in which a single strategy profile or outcome strongly *payoff-dominates* all others, in the sense that it yields strictly better payoffs to every player than any other outcome and is, therefore, jointly optimal for all (Anderlini, 1999; Aumann & Sorin, 1989). It seems intuitively obvious that rational players will choose *H* in the Hi-Lo game, resulting in the payoff-dominant outcome (*H*, *H*), and it is therefore surprising that game theory provides a player with no reason or justification for choosing *H*. To see why this is so, consider the following standard *common knowledge and rationality* assumptions of game theory:

1. Common knowledge: The specification of the game, represented in a simple two-player case by its payoff matrix, together with everything that can validly be deduced from it, is *common knowledge* in the game, in the sense that both players know it, know that both know it, know that both know that both know it, and so on.
2. Rationality: The players are *instrumentally rational* in the sense of always choosing strategies that maximize their own individual payoffs, relative to their knowledge and beliefs, and this too is common knowledge in the game.

On the basis of these assumptions, Player 1, for example, has a reason to choose *H* in the Hi-Lo game if and only if there is a reason to expect Player 2 to choose *H*. The "only if" condition arises from the fact that *H* is not an *unconditionally* best strategy, because if Player 2 were to choose *L*, then Player 1 would do better by also choosing *L*. The crucial question is, therefore, whether Player 1 has a reason to expect Player 2 to choose *H*. The answer is *no*, because the symmetry of the game ensures that Player 2 faces exactly the same dilemma, having no reason to choose *H* in the absence of a reason to expect Player 1 to choose it.

This problem of coordination cannot be solved by pointing out that the (*H*, *H*) outcome is more salient or prominent than (*L*, *L*), because of its superior payoffs, or in other words that it is a *focal point* in the sense of Schelling (1960), although it obviously is. Gilbert (1989) provided a rigorous and conclusive argument showing that "mere salience is *not* enough to provide rational agents with a reason for action (though it would obviously be nice, from the point of view of rational agency, if it did)" (p. 69, italics in original). She showed that a player has no reason to choose a salient focal point in the absence of a reason to expect the co-player to choose it. Any attempt to derive a reason for choosing *H* from the standard common knowledge and rationality assumptions generates a vicious circle that loops back to the starting point without reaching any definite conclusion (Anderlini, 1999; Aumann & Sorin, 1989; Bacharach, 2006, chap. 1; Bardsley, Mehta, Starmer, & Sugden, 2010; Cooper, DeJong, Forsythe, & Ross, 1990; Crawford & Haller, 1990; Harsanyi & Selten, 1988; Janssen, 2001). The problem of coordination is of central importance to our understanding of human social behavior, and it arises frequently in everyday strategic interactions, but it is poorly understood, and both psychologists and decision scientists have failed to pay it as much attention as it deserves.

The fundamental solution concept of game theory is *Nash equilibrium* (Nash, 1950, 1951). In a two-player game, a Nash equilibrium is a pair of strategies that are *best replies* to each other, a best reply being a strategy that maximizes the payoff of the player choosing it, given the strategy chosen by the co-player. It follows from this definition that any pair of strategies that are out of equilibrium constitute an outcome that is necessarily self-destabilizing, because at least one player could have done better by acting differently and, therefore, has a motive to avoid that outcome. In the Hi-Lo game, (*H*, *H*) is a Nash equilibrium, because if Player 1 chooses row *H*, then Player 2's best reply is column *H* (because $2 > 0$), and if Player 2 chooses column *H*, then Player 1's best reply is row *H* for the same reason. The *H* strategies are best replies to each other, but essentially the same argument establishes that (*L*, *L*) is also a Nash equilibrium, because *L* is a best reply to *L* for both players, although with lower payoffs of 1 to each. The game also has a mixed-strategy equilibrium, the details of which need not detain us here, in which both players choose randomly with probabilities of 1/3 *H* and 2/3 *L*. The outcomes (*H*, *L*) and (*L*, *H*) are both out of equilibrium, neither strategy being a best reply to the other. The (*H*, *H*) equilibrium payoff-dominates (or Pareto-dominates) all other outcomes of the game, including the (*L*, *L*) and mixed-strategy equilibria, in the sense that both players receive higher payoffs in (*H*, *H*) than in any other outcome of the game, and the Hi-Lo game is, therefore, a common interest game. However, there is no reason, based on the standard assumptions of game theory, for either player to choose *H*, and neither the influential *subgame-perfect* equilibrium introduced by Selten (1975) nor any of the other Nash equilibrium refinements that have been proposed solves this surprisingly tricky problem.

A celebrated *indirect argument*, first put forward by von Neumann and Morgenstern (1944, section 17.3.3, pp. 146–148) in relation to strictly competitive games and later generalized to other games by Luce and Raiffa (1957, pp. 63–65, 173) proves that if a game has a uniquely rational solution, then that solution must be a Nash equilibrium. The proof can be outlined for a two-player game as follows. The standard common knowledge and rationality assumptions imply that if a game has a uniquely rational solution—in a two-player game, if it is uniquely rational for Player 1 to choose a particular strategy s_1 and for Player 2 to choose a particular strategy s_2—then those strategies must be best replies to each other, because common knowledge ensures that each player can anticipate the other's strategy and will choose a best reply to it. Because s_1 and s_2 are best replies to each other, they are in Nash equilibrium by definition. However, although a uniquely rational solution must necessarily be a Nash equilibrium, the theory is indeterminate in games with two or more Nash equilibria, even in common interest games in which one equilibrium is better for both players than any other, because the standard common knowledge and rationality assumptions provide no reason for choosing strategies associated with a payoff-dominant equilibrium (Harsanyi & Selten, 1988). Nevertheless, in the Hi-Lo game, for example, experimental evidence confirms that, in practice, more than 96% of players manage without difficulty to coordinate on the obvious payoff-dominant (*H*, *H*) equilibrium (Bardsley et al., 2010).

Ad-Hoc Explanations

How can coordination in common interest games be explained? In particular, what accounts for the powerful intuition that it is rational to choose strategies associated with payoff-dominant Nash equilibria in the Hi-Lo game and other common interest games? Harsanyi and Selten (1988) called this the *payoff-dominance problem* and discussed it at length in relation to the Stag Hunt game shown in Figure 2. Like the Hi-Lo game, the Stag Hunt game is a common interest game with two Nash equilibria in the main diagonal, (*C*, *C*) payoff-dominating (*D*, *D*). It is named after Rousseau's (1755, Part 2, paragraph 9) example of two hunters who need to cooperate (*C*) to catch a stag, but who are both tempted to defect (*D*) and chase after a hare, which they can catch without each other's help. (In Harsanyi and Selten's version shown in Figure 2, if both players defect, then they are slightly less likely to catch hares, perhaps because they may chase after the same one.)

		Player 2	
		C	D
Player 1	C	9, 9	0, 8
	D	8, 0	7, 7

Figure 2. The Stag Hunt game, with a payoff-dominant Nash equilibrium at (*C, C*) and a payoff-dominated Nash equilibrium at (*D, D*).

Harsanyi and Selten (1988) solved the problem by introducing a *payoff-dominance principle* as an axiom of rationality. According to this principle, it is simply an axiomatic feature of human rationality that if one Nash equilibrium payoff-dominates all others in a game, then players will choose the strategies associated with it. Harsanyi and Selten proposed this principle (together with a secondary risk-dominance principle that we need not discuss here) merely as a temporary workaround, acknowledging that it provides no *explanation* for the fact that payoff-dominant equilibria are easily chosen by players in practice, or for the powerful intuitive appeal of such solutions (see their comments on pp. 355–363). Janssen's (2001, 2006) principle of *individual team member rationality* is a weak variant of the same ad hoc axiom, applying to payoff-dominant outcomes whether or not they are Nash equilibria.

A number of other researchers have grasped the nettle and attempted to explain the payoff-dominance phenomenon. Following these introductory remarks, we critically review the most prominent explanations. We refer to these explanations as *theories,* restricting the term *strategies* to the game-theoretic sense of options that players can choose in any particular game. We leave aside theories that depend on essential alterations of the rules of the game—that it is played just once and that players choose their strategies independently—particularly those that require repetitions of the game (Anderlini & Sabourian, 1995; Aumann & Sorin, 1989) or costless preplay communication or "cheap talk" between the players (Anderlini, 1999; Ellingsen & Östling, 2010; Farrell, 1988; Rabin, 1994). Among the theories that we discuss is a theory of strong Stackelberg reasoning, designed to avoid the problems associated with other theories and presented here for the first time in its improved form. We then report the results of two experiments designed to compare the performance of the leading theories against one another, and we conclude with a discussion of the results.

Principle of Indifference and Theory of Rational Beliefs

According to the *principle of indifference* (Keynes, 1921, pp. 41–64), also called the *principle of insufficient reason*, we may assign equal probabilities to events whenever there is no reason to believe that one is more probable than another. A common fallacy (e.g., Gintis, 2003; Monterosso & Ainslie, 2003) involves an attempt to apply this principle to the payoff-dominance problem, using an argument along the following lines. In the Hi-Lo game shown in Figure 1, if Player 1 does not know which strategy Player 2 is likely to choose, then Player 1 may assign equal probabilities to Player 2's strategies. Player 1's expected payoff from choosing *H* is then $(1/2 \times 2) + (1/2 \times 0) = 1$, and because this is better than the expected payoff from choosing *L*, namely $(1/2 \times 0) + (1/2 \times 1) = 1/2$, it is rational for Player 1 to choose *H* and, by symmetry, it is also rational for Player 2 to choose *H*. This appears to establish a reason for choosing strategies associated with the payoff-dominant Nash equilibrium (*H*, *H*). Unfortunately, it is not valid to assign probabilities to a co-player's strategies in this way, as the following proof by *reductio ad absurdum* shows (Colman, 2003a). According to the standard common knowledge assumption (1) stated formally earlier in the article, Player 1's deduction that it is rational to choose *H*, if it were valid, would be common knowledge in the game, and according to the rationality assumption (2), Player 2 would therefore choose *H* with certainty, contradicting Player 1's initial assumption that this probability is 1/2. Therefore, Player 1 cannot apply the principle of indifference without contradiction.

Other versions of this argument, more subtle but equally invalid, are also based on subjective probabilities. For example, Hausman (2003) invoked the theory of *rational beliefs* as a basis for arguing as follows: "Player 1 can believe that the probability that Player 2 will play *H* is not less than one half, and also believe that Player 2 believes the same of Player 1. Player 1 can then reason that Player 2 will definitely play *H*, update his or her subject probability accordingly, and play *H*" (pp. 163–164).[1] But why should a player believe that the probability is not less than 1/2 that the co-player will choose *H*? Can Player 1 believe that the probability is not less than 3/4 and then apply the same reasoning? The absurdity becomes clear if we carry this to its logical conclusion and assume Player 1 believes that the probability is not less than unity that Player 2 will play *H*. This exposes the hidden logical fallacy of begging the question (*petitio principii*) or assuming what we are trying to prove. Solutions of this type, together with those based on the principle of indifference, are spurious, and we do not include them in our experimental tests, described later in this article.

Social Projection Theory

Social projection theory (Acevedo & Krueger, 2004, 2005; Krueger, 2007, 2008; Krueger & Acevedo, 2005; Krueger, DiDonato, & Freestone, 2012) rests on a fundamental assumption that people tend to project their preferences and intentions on to others and, in particular, that "most people have a strong expectation that members of their own groups will act as they themselves do" (Krueger, 2008, p. 399). In the Hi-Lo game (see Figure 1), a player therefore expects either strategy choice to be matched by the co-player. The payoff from choosing *H* is thus expected to be 2 and from choosing *L* it is expected to be 1, and a rational player has a reason to choose *H*, because $2 > 1$. This theory seeks to explain not only coordination in symmetric common interest games, but also cooperation in social dilemmas, from the assumption that players expect their strategy choices to be matched.

A mathematical foundation for social projection theory was first suggested by Dawes (1989) in his rationalization of the *false consensus effect*, later elaborated by Krueger (1998), Dawes (2000), and Krueger, DiDonato, and Freestone (2012), among others. Suppose that you have literally no basis for predicting whether your co-player will choose *H* or *L* in the Hi-Lo game. You apply the principle of indifference and, in the terminology of Bayesian decision theory, start off with a *uniform prior*—your initial subjective probability that your co-player will choose the *H* strategy is $P(H) = 1/2$. Suppose further that you decide to choose *H*. Your own choice provides data that you should use to update your probability estimate of the co-player's choice to $P(H) > 1/2$. Dawes argued that there is no good reason to ignore it, because it is empirical evidence of how people behave, even though the size of the sample is just $N = 1$ and its only member is you. We can model this with the process of judging the probability that a ball drawn from an urn will be red, assuming that the urn contains an unknown mixture of red and black balls, every possible ratio of red and black balls being equally likely. Initially, you have no way of knowing the composition of the urn, so it seems reasonable to start with a uniform prior. Suppose that you then draw a single ball, note that it is red, and replace it in the urn. Your task is to estimate the probability that a second draw will produce another red ball. Your estimate should be influenced by the color of the first ball, because whatever your prior was before drawing any balls, a red ball should nudge your posterior (the probability of a second red ball) higher. This appears to provide a reason for choosing H in the Hi-Lo game, because if you choose H and consider your own choice as evidence, you should expect that your co-player will probably choose H also.

---

[1] Hausman (2003) was referring to a slightly different coordination game, but this does not affect the point that he was making or our rebuttal.

Dawes (1989, 1990, 2000) claimed that, according to Bayesian theory, a person who starts out with a uniform prior, $P(H) = 1/2$, should update the probability to $P(H) = 2/3$ after drawing a single red ball. He offered several heuristic arguments and partial proofs, and other researchers have repeated his claim but, as far as we are aware, no complete proof has appeared. The result is, however, correct, and the phenomenon turns out to be a special case of Laplace's rule of succession (see Kendall, 1945, pp. 176–177). In Appendix A we provide a proof of it that is not difficult but (surprisingly) requires integral calculus. The relevance of this to social projection theory is that a player with a uniform prior who chooses either H or L in the Hi-Lo game shown in Figure 1 should assign a probability of 2/3 to the event that the co-player will choose the matching strategy. It is argued that a player's expected payoff from choosing H is therefore $(2/3 \times 2) + (1/3 \times 0) = 4/3$, and from choosing L is $(1/3 \times 0) + (2/3 \times 1) = 2/3$, therefore a player who is rational in the sense of the rationality assumption (2) stated formally earlier in the article will choose H. This is the mathematical foundation of the social projection theory of coordination in common interest games first suggested by Acevedo and Krueger (2004, 2005).

There are two main problems with this. The first is that the Bayesian reasoning underlying it applies only if players have no grounds whatsoever, apart from the evidence of their own choices, for judging the probabilities associated with the choices of others. This condition is never satisfied in real-life social interactions. Past experience invariably provides clues from similar interactions that are bound to influence a player's assessment of the probabilities. In fact, in the context of past experience, a player's own choice is hardly likely to provide a significant amount of additional evidence, still less the only evidence.

A more fundamental problem arises from the temporal structure of the hypothesized judgment and decision process. According to social projection theory as applied to the Hi-Lo game, (a) players expect others to choose the same strategies that they choose themselves; (b) therefore, whatever they choose in the Hi-Lo game, they expect their co-players to choose the matching strategy; and (c) this explains why they normally choose H and why this seems intuitively appealing. The problem is that players' expectations follow after and as a consequence or effect of their own choices and cannot therefore *explain* those choices, because an effect cannot precede its own cause. This problem can be evaded by assuming that players merely contemplate choosing each of the available strategies, assuming in each case that their choice would be matched if they were to make it, and then choose the strategy that promises the best payoff given that assumption (Krueger et al., 2012), but this seems acceptable only if choices that are merely contemplated are treated on an equal footing, for purposes of Bayesian updating, with choices that are actually made. A player's actual choices may possibly provide evidence that others are likely to choose the same strategies, but it could be argued that options that are merely contemplated and *not* chosen provide no such evidence. A consistent interpretation of social projection theory suggests that people should expect others *not* to choose what they themselves have not chosen, after contemplation.

Social projection theory relies on a form of reasoning, called *evidential decision theory*, that is rejected by many decision theorists but also has some distinguished advocates (e.g., Eells, 1984, 1989; Jeffrey, 1983, 2004; Nozick, 1993, pp. 43–59). The temporal structure objection set out in the previous paragraph is not universally accepted and, whether or not evidential decision theory is valid, there is experimental evidence that many people do, in fact, apply it in certain circumstances (Anand, 1990; Quattrone & Tversky, 1984). However, social projection theory is obviously limited to symmetric games, because decision makers have no reason to expect others to act as they themselves do in situations in which similar actions have different consequences, or in games in which similar actions cannot even be unambiguously specified. Cognitive hierarchy theory and other theories of coordination outlined in the following paragraphs apply to both symmetric and asymmetric games, and any empirical comparison of theories requires asymmetric experimental games, because different theories all predict the same choices in symmetric games. For this reason, we do not include social projection theory in our experimental tests, although we acknowledge that it may help, in part at least, to explain choices in symmetric games such as the Hi-Lo and Stag Hunt games shown in Figures 1 and 2.

Cognitive Hierarchy Theory

Cognitive hierarchy theory was first proposed by Camerer, Ho, and Chong (2004), following slightly more elaborate versions put forward by Stahl and Wilson (1994, 1995) and Bacharach and Stahl (2000). It is a theory of bounded rationality designed to model players who reason with varying levels of strategic depth. Level-0 players have no beliefs about their co-players and choose strategies randomly, with uniform probability; Level-1 players maximize their own payoffs given their belief that their co-players are Level-0 players; Level-2 players maximize their own payoffs given their belief that their co-players are Level-1 or Level-0 players; and so on. The integrative models of social value orientation proposed by Van Lange (1999) amount to Level-1 reasoning applied to transformed payoffs, after each player's payoffs have been incremented by a fixed amount *h* in outcomes where their co-players receive identical payoffs or decreased by *h* in outcomes where they are different, to reflect equality-seeking, and in some models, additional payoff transformations are applied (Colman, Pulford, & Rose, 2008b; Van Lange, 2008).

Experimental evidence reported by Camerer et al. (2004) suggested that Level 1 is most common, followed by Level 2, and that higher levels occur only very infrequently. This has been corroborated by Hedden and Zhang (2002), Bardsley, Mehta, Starmer, and Sugden (2010), and others, and it is also consistent with evidence of levels of recursive thinking typically found in other domains of cognition (Colman, 2003b).

The reasoning used by Level-1 players amounts to unweighted (or equal-weighted) expected payoff maximization, and it is nonstrategic in the sense that it makes no attempt to analyze the game from the co-player's point of view. In the Hi-Lo game shown in Figure 1, a Level-1 player believes that the co-player will choose *H* or *L* with equal probability, and the Level-1 player chooses the strategy that maximizes expected payoff relative to that belief about the co-player. That strategy is *H*, because $(1/2 \times 2) + (1/2 \times 0) > (1/2 \times 0) + (1/2 \times 1)$. A Level-2 player also chooses *H* because it yields the best payoff (2 rather than 0) against a Level-1 player who, as we have just shown, chooses *H*.

One major weakness of cognitive hierarchy theory as a theory of coordination is that, whereas it works for pure coordination games such as the Hi-Lo game, it fails in other common interest games. It will suffice to consider the Stag Hunt game shown in Figure 2. A Level-1 player chooses *D*, because $(1/2 \times 9) + (1/2 \times 0) < (1/2 \times 8) + (1/2 \times 7)$, a Level-2 player also chooses *D*, because $7 > 0$, and the same applies to higher-level players. Thus, cognitive hierarchy theory fails to predict the payoff-dominant (*C*, *C*) outcome in this frequently discussed version of an iconic common interest game, even with arbitrarily deep levels of strategic reasoning. However, (*C*, *C*) is intuitively appealing, and 79% of players in an experimental Stag Hunt game reported by Colman and Stirk (1998) chose *C*.

A second problem with the theory is its asymmetry: It relies on an assumption that players never credit their co-players with the same level of strategic sophistication as themselves. Although there is evidence for self-serving beliefs in other domains (Krueger & Wright, 2011), it is not clear why players should hold this patronizing belief, and it seems unlikely that they do. A world in which everyone is a deeper strategic thinker than everyone else is reminiscent of Lake Wobegon in Garrison Keillor's radio show *A Prairie Home Companion*, "where all the women are strong, all the men are good-looking, and all the children are above average." Nevertheless, the theory certainly provides a simple explanation for coordination in many common interest games, including asymmetric games, therefore we include it in our experimental tests.

Team Reasoning

According to theories of team reasoning (Bacharach, 1999, 2006; Gold & Sugden, 2007; Smerilli, 2012; Sugden, 1993, 2005), there are circumstances in which players are motivated to maximize not their *individual* expected payoffs, as specified by the standard rationality assumption (2) stated formally earlier in the article, but the *collective* payoff of the group of players involved in the game. Decision making based on such collective preferences is usually called *team reasoning*, and theories of team reasoning generally assume that circumstances dictate whether players attempt to maximize their collective payoffs (team reasoning) or their individual payoffs (orthodox individual reasoning). Common interest games provide typical circumstances in which collective payoff maximization and team reasoning might be expected to occur (Tan & Zizzo, 2008). Standard game theory is interpreted as a special case of team reasoning in which the group happens to be a singleton.

The key assumption, that team-reasoning players attempt to maximize collective rather than individual payoffs, is a radical departure from orthodox game theory and decision theory. In orthodox theory, players are assumed to ask themselves, in effect, *What do I want, and what should I do to achieve it?* Team-reasoning players are assumed to ask, *What do we want, and what should I do to play my part in achieving it?* If we make the natural assumption that the collective payoff is simply the sum (or equivalently, the average) of the individual payoffs in any outcome, then in the Hi-Lo game shown in Figure 1, the team-reasoning answer is obviously: *We want* (*H, H*), *because the collective payoff of 4 is greater than in any other outcome, and what I need to do to play my part in achieving that outcome is to choose H.* In the Stag Hunt game shown in Figure 2, a team-reasoning player reasons: *We want* (*C, C*), *because the collective payoff of 18 in that outcome is greater than in any other, and I should play my part by choosing C.* Team reasoning thus provides a compelling solution to both of these games, and to all other common interest games as well.

Team-reasoning players use a distinctive mode of reasoning to reach decisions on the basis of their own and their co-players' individual preferences. They begin by searching for a profile of strategies that maximizes the collective payoff of the pair or group of players. If such a profile is unique, they choose and play its component strategies. If the game has no unique strategy profile that is collectively rational, then the theory is indeterminate. Team reasoning provides a complete solution to the payoff-dominance problem. It solves all common interest games, because these games have payoff-dominant Nash equilibria that are necessarily collectively rational outcomes.

A player has no reason to adopt the team-reasoning mode in the absence of a belief that the other player(s) will do likewise, and all versions of the theory assume that team reasoning occurs only when a player expects the co-player(s) to adopt the team-reasoning mode. This should be distinguished from social projection theory, according to which players assume automatically that others will act as they do themselves. In Bacharach's (1999, 2006) stochastic version of the theory, players adopt the team-reasoning mode if the probability that the co-player(s) will do the same is high enough. The probability that a player will adopt the team-reasoning mode is represented by a parameter omega ($0 \leq \omega \leq 1$), the value of which is common knowledge in the game, and players are assumed to adopt the team-reasoning mode if and only if ω is high enough to ensure that the expected collective payoff is maximized by team reasoning; otherwise, the player is assumed to lapse into the individual payoff maximization characteristic of standard game-theoretic reasoning.

Team reasoning requires the abandonment of a fundamental assumption of *methodological individualism*, the bedrock of rational choice theory, incorporated in the standard rationality assumption (2) stated formally earlier in the article. More generally, methodological individualism implies that decision makers are rational in the sense of attempting to do the best for themselves as individuals in all circumstances that arise (Elster, 1982). This creates problems in interpreting team reasoning within the framework of orthodox game theory, where the payoffs are von Neumann–Morgenstern utilities in the sense of expected utility theory. These are supposed to represent the players' preferences, taking into account everything that affects these preferences, including other-regarding considerations of how co-players are affected by the outcomes. Theories of team reasoning assume that team-reasoning players' utilities are functions of their own and their co-players' individual payoffs, but individual payoffs, if they are von Neumann–Morgenstern utilities, already incorporate considerations of co-players' payoffs, and all that ought to remain is for players to maximize their own individual utilities. We cannot get around this problem by conceding that team-reasoning players do indeed attempt to do the best for themselves as individuals, but that they happen to be individually motivated to achieve collectively rational outcomes, because in that case their collective preferences should be fully reflected in their individual payoff functions, and conventional individual reasoning should suffice, so that we could simply replace each individual payoff by the sum of both players' payoffs in the corresponding outcome. For example, in Figure 1, we could replace the payoffs in (*H*, *H*) with (4, 4) and the payoffs in (*L*, *L*) with (2, 2), but this would not provide players with any reason to choose *H*: The strategic structure of the game would be unaltered and team reasoning would still be required.

However, the financial payoffs used to motivate players in experimental games do not and cannot incorporate other-regarding preferences, and there is persuasive experimental evidence that team reasoning applied to individual *objective payoffs*, rather than *subjective utilities*, does indeed occur in experimental games (Bardsley et al., 2010; Butler, 2012; Colman, Pulford, & Rose, 2008a). Although team reasoning violates the assumption of methodological individualism at a theoretical level, from a more practical point of view it offers a persuasive explanation for coordination in experimental games, where it applies to objective payoffs rather than utilities, and we therefore include it in our experiments.

Strong Stackelberg Reasoning

The theory of strong Stackelberg reasoning is an improved version of an earlier theory of Stackelberg reasoning (Colman & Bacharach, 1997), named after a simple model of leader-follower games proposed by Heinrich von Stackelberg (1934). The new theory is designed to overcome a limitation of the previous version and to avoid problems associated with the other theories of coordination that we have outlined. It requires no modification of the one-shot, independent choice rules of the game, or of methodological individualism, and it retains the standard common knowledge and rationality assumptions of orthodox game theory. Its distinctive assumption is that players choose strategies as if they believed that their co-players could anticipate their choices and invariably choose best replies to them, and that they maximize their own payoffs accordingly. In other words, players choose strategies according to standard game-theoretic assumptions, but they behave as if choosing first in a sequential game with *perfect information*—a game in which the player moving second always knows the first player's move, as in chess, for example. Game theory places no limitations on players' beliefs, apart from technical assumptions of completeness and consistency, hence this theory would not require any nonstandard assumptions even if we stipulated that players using strong Stackelberg reasoning actually believed that their co-players could anticipate their choices. A related theory proposed by Weber, Camerer, and Knez (2004) applies to games in which players actually do choose strategies sequentially but the player moving second does not know the first player's previous move.

Players are assumed to generate strong Stackelberg strategies by identifying the strategies that maximize their own payoffs against invariably best-replying co-players, and they then play those strong Stackelberg strategies if they form Nash equilibria. Stackelberg reasoning thus involves a strong Stackelberg strategy generator (a mode of reasoning that generates Stackelberg strategies, provided that best replies are strong) followed by a Nash filter to check that the Stackelberg strategies are in Nash equilibrium.[2] Games in which strong Stackelberg strategies form Nash equilibria are *S-soluble*. Games in which strong Stackelberg strategies are not well defined or are out of equilibrium are *non-S-soluble*, and in such games the theory makes no specific predictions. The theory is presented formally in Appendix B.

The principal limitation of the earlier theory of Stackelberg reasoning (Colman & Bacharach, 1997) is its failure to deal with the problem that Stackelberg strategies are not necessarily well defined if the co-player has a best reply that is not unique—if two or more replies are best and yield payoffs that are jointly optimal. An example of such a game is given in Figure 3. Suppose that I am Player 1 and you are Player 2. Using Stackelberg reasoning, I expect that any strategy choice of mine will be met by your best reply. But you have two jointly best replies to my *A* strategy, because you receive the same payoff (2) whether you reply with *A* or *B*. Although I am motivated to choose a payoff-maximizing strategy of my own, I cannot anticipate your response to my *A* strategy, and hence Stackelberg reasoning does not enable me to work out which of my own strategies is best for me: *A* would be best for me if you were to reply with *A* (my payoff would be 3, rather than 2 from choosing *B*), otherwise *B* would be best for me (my payoff would be 1, rather than 0 from choosing *A*). The Stackelberg strategy generator simply breaks down in a game such as this, and the earlier theory of Stackelberg reasoning provides no solution to this problem.

The theory of strong Stackelberg reasoning overcomes the problem by exploiting the *strong best reply* function, introduced by Harsanyi and Selten (1988), in which best replies are unique by definition. Although strong best replies in this sense obviate the problem of ill-defined Stackelberg strategies, we show that a slightly weaker condition, inspired by the concept of mutual ordinality (Howard, 1971, pp. 147–151), suffices to ensure that the Stackelberg generator always yields a unique best reply. The key assumption (see Appendix B) is that when a player strictly prefers one outcome to another, the co-player is never indifferent between the same two outcomes. This ensures that strong Stackelberg strategies are invariably well defined. Our condition may appear, at first, to limit the applicability of the theory by arbitrarily excluding some (perhaps many) real-life strategic interactions from the scope of the theory, but that turns out not to be the case for two reasons. First, even in games in which our condition does not hold, strong Stackelberg strategies may nevertheless be well defined—all that is strictly necessary is that the maximum payoff to Player 1 in each column of the payoff matrix and the maximum payoff to Player 2 in each row is unique. Second and more important, our condition for strong best replies is violated only if players are completely indifferent to the payoffs of their co-players, and that is never the case in real-life strategic interactions. In the game depicted in Figure 3, Player 2 is indifferent between the (*A*, *A*) to the (*A*, *B*) outcomes because $2 = 2$, although Player 1 prefers (*A*, *A*) to the (*A*, *B*) because $3 > 0$. This could arise in real life only if Player 2 were utterly indifferent to Player 1's payoffs.

		Player 2	
		A	*B*
Player 1	*A*	3, 2	0, 2
	B	2, 2	1, 1

Figure 3. A game in which Player 2 does not have a strong best reply to Player 1's A strategy.

It is widely acknowledged that other-regarding social value orientations such as cooperativeness, competitiveness, and altruism (Colman, Körner, Musy, & Tazdaït, 2011; Rusbult & Van Lange, 2003; Van Lange, 2000), and other-regarding considerations of fairness and reciprocity (e.g., Arnsperger & Varoufakis, 2003; Bolton & Ockenfels, 2000; Brosnan & de Waal, 2002; Fehr & Schmidt, 1999; Trivers, 2005) influence the preferences of humans and even nonhuman primates. We are never entirely indifferent to the payoffs of others with whom we interact. It is also worth noting that the condition that turns out to be sufficient to establish that best replies are strong does not imply any abandonment of methodological individualism, because the generic other-regarding preferences specified in Equations and Inequalities 4 and 5 in Appendix B are fully integrated into the players' own payoff functions, and players are motivated to maximize their individual payoffs in accordance with standard game-theoretic assumptions.

---

[2] We are grateful to Werner Güth for suggesting this terminology.

Applying strong Stackelberg reasoning to the Hi-Lo game shown in Figure 1, Player 1 chooses a strategy as though believing that *H* would be met by Player 2's best reply *H*, and *L* would be met by Player 2's best reply *L*. Player 1 receives a payoff of 2 in the first case and 1 in the second and, therefore, prefers the payoff-maximizing strategy *H* and, because the game is symmetric, Player 2 arrives at the same conclusion. Because (*H*, *H*) is in equilibrium, it is the strong Stackelberg solution of the game, and both players therefore choose *H*. In the Stag Hunt game shown in Figure 2, a similar analysis shows that both players choose *C*.

It is not difficult to prove that every common interest game is S-soluble, and that, if a game with multiple Nash equilibria has one equilibrium that payoff-dominates the others, then its strong Stackelberg solution is its payoff-dominant Nash equilibrium (see Colman & Bacharach, 1997, for proofs that apply with only minor modifications to strong Stackelberg reasoning). The theory of strong Stackelberg reasoning therefore appears to provide a comprehensive explanation of coordination in common interest games between players who care about one another's payoffs, and it avoids the problems associated with the other theories that we have reviewed.

Other theories discussed in this article appear also to explain coordination, at least in experimental games, although it is conceivable that none of the existing theories is satisfactory. We therefore tested the theories of cognitive hierarchy theory (Level-1 and Level-2 reasoning), team reasoning, and strong Stackelberg reasoning in two experiments specifically designed to compare their performance against one another. We did not use common interest games because the theories under investigation would all predict the same strategy choices, and we did not use 2 × 2 games because they would not have been very useful for distinguishing between four potentially different theoretical predictions. Instead, we designed 3 × 3 and 4 × 4 experimental games in which the theories generate conflicting predictions.

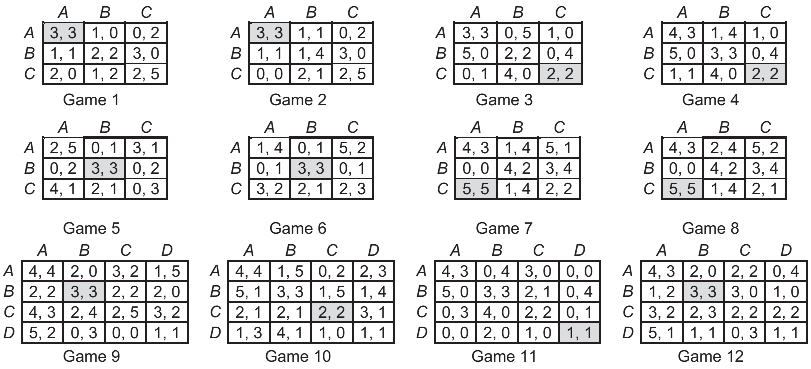
Experiment 1

Method

Participants. The participants were 68 students and employees at the University of Leicester (45 female, 23 male), aged 18–52 years ($M = 25.28$, $SD = 6.71$) recruited from the School of Psychology's participant panel, an approximate sample size of 70 having been determined in advance. They were remunerated according to the random lottery incentive system, a technique that avoids a number of problems associated with other payment schemes (Lee, 2008) and has been shown to elicit true preferences (Cubitt, Starmer, & Sugden, 1998; Starmer & Sugden, 1991). We paid each participant a show-up fee of £3.00 ($5.00) plus an additional amount, up to £5.00 ($8.00) according to their payoffs in a single game randomly preselected from the 12 games used in the experiment. To maximize the incentive value of the remuneration, we did not mention the show-up fee until the experiment was over: Before and during the experiment the participants knew only that they could win up to £8.00 ($13), depending on their choices in a randomly chosen game.

Materials. We devised eight S-soluble 3 × 3 games and four S-soluble 4 × 4 games capable of distinguishing between the various theories under investigation, namely cognitive hierarchy theory (Level-1 and Level-2 reasoning), team reasoning, and strong Stackelberg reasoning. The 12 games used in the experiment are displayed in Figure 4. Seven of the eight 3 × 3 games (all apart from Game 3) and all four of the 4 × 4 games are asymmetric, and best replies are strong in the sense defined in Appendix B. There are no strongly or weakly dominant strategies in any game. We randomized the order of the 12 games and presented them in reverse order to half the players to check for possible order effects. We did not drop any variables, conditions, or games from our analyses in this experiment or in Experiment 2.

The experimental participants were assigned to the role of Player 1 (row chooser). In each testing session, a single participant, whose responses were not included in our data analysis, served as Player 2 for all other players in the same testing session. Predictions derived for Player 1 choices are unique in every game, in



	A	B	C
A	3, 3	1, 0	0, 2
B	1, 1	2, 2	3, 0
C	2, 0	1, 2	2, 5

Game 1

	A	B	C
A	3, 3	1, 1	0, 2
B	1, 1	1, 4	3, 0
C	0, 0	2, 1	2, 5

Game 2

	A	B	C
A	3, 3	0, 5	1, 0
B	5, 0	2, 2	0, 4
C	0, 1	4, 0	2, 2

Game 3

	A	B	C
A	4, 3	1, 4	1, 0
B	5, 0	3, 3	0, 4
C	1, 1	4, 0	2, 2

Game 4

	A	B	C
A	2, 5	0, 1	3, 1
B	0, 2	3, 3	0, 2
C	4, 1	2, 1	0, 3

Game 5

	A	B	C
A	1, 4	0, 1	5, 2
B	0, 1	3, 3	0, 1
C	3, 2	2, 1	2, 3

Game 6

	A	B	C
A	4, 3	1, 4	5, 1
B	0, 0	4, 2	3, 4
C	5, 5	1, 4	2, 2

Game 7

	A	B	C
A	4, 3	2, 4	5, 2
B	0, 0	4, 2	3, 4
C	5, 5	1, 4	2, 1

Game 8

	A	B	C	D
A	4, 4	2, 0	3, 2	1, 5
B	2, 2	3, 3	2, 2	2, 0
C	4, 3	2, 4	2, 5	3, 2
D	5, 2	0, 3	0, 0	1, 1

Game 9

	A	B	C	D
A	4, 4	1, 5	0, 2	2, 3
B	5, 1	3, 3	1, 5	1, 4
C	2, 1	2, 1	2, 2	3, 1
D	1, 3	4, 1	1, 0	1, 1

Game 10

	A	B	C	D
A	4, 3	0, 4	3, 0	0, 0
B	5, 0	3, 3	2, 1	0, 4
C	0, 3	4, 0	2, 2	0, 1
D	0, 0	2, 0	1, 0	1, 1

Game 11

	A	B	C	D
A	4, 3	2, 0	2, 2	0, 4
B	1, 2	3, 3	3, 0	1, 0
C	3, 2	2, 3	2, 2	2, 2
D	5, 1	1, 1	0, 3	1, 1

Game 12

Figure 4. S-soluble experimental games, with shaded cells indicating strong Stackelberg solutions and hence also Nash equilibria. Player labels (1 for row chooser and 2 for column chooser) are suppressed to save space.

the sense that each theory predicts exactly one strategy choice. In Game 1, for example, the cognitive hierarchy Level-1 prediction for Player 1 is *B*, because a Player 1 who believes that Player 2 will choose randomly with uniform probability has the following expected payoffs: from choosing *A*, (3 + 1 + 0)/3; from choosing *B*, (1 + 2 + 3)/3; and from choosing *C*, (2 + 1 + 2)/3; and because the greatest expected payoff is from choosing *B*, a Level-1 cognitive hierarchy reasoner will choose *B*. The cognitive hierarchy Level-2 prediction for Player 1 is *B*, because a Level-2 Player 1 who believes that Player 2 will use Level-1 reasoning will expect Player 2 to choose *C*, and the best reply to *C* is *B*. The team reasoning prediction is that Player 1 will choose *C*, because the collective payoff in the (*C*, *C*) outcome (2 + 5 = 7) is greater than in any other outcome. The theory of strong Stackelberg reasoning predicts that Player 1 will choose *A* because a Player 1 who expects Player 2 to choose a best reply to any strategy expects *A* to elicit the reply *A*, paying 3 to Player 1, *B* to elicit the reply *B*, paying 2 to Player 1, and *C* to elicit the reply *C*, paying 2 to Player 1. Therefore, Player 1 will choose *A* (because 3 > 2), and a similar argument shows that Player 2 will also choose *A*. Finally, (*A*, *A*) is a Nash equilibrium because these *A* strategies are best replies to each other, hence the game is S-soluble, and Player 1 will therefore choose *A*.

It is clearly impossible to have four distinct predictions in a game with only three strategies per player, hence in Game 1, cognitive hierarchy Level-1 reasoning, team reasoning, and strong Stackelberg reasoning make distinct predictions for Player 1's choice, but cognitive hierarchy Level-2 reasoning predicts the same choice as cognitive hierarchy Level-1 reasoning. In the other 3 × 3 games, there was always one strategy choice that was predicted by two theories. In each of our 4 × 4 games, cognitive hierarchy Level-1 reasoning, cognitive hierarchy Level-2 reasoning, team reasoning, and strong Stackelberg reasoning are fully separated, each yielding a different prediction for Player 1's choice.

Procedure. The experiment was conducted over four 35-min testing sessions, with approximately 16–18 participants per session. The participants sat at computer monitors and logged on to the SurveyGizmo data-gathering Web site, where they were presented with the following on-screen instructions:

> You will be presented with a series of 12 grids. For some grids you will be asked to choose between A, B, and C, and for others you will be asked to choose between A, B, C, and D. You will be paired with another randomly selected participant in this room for each of your 12 decisions. In each case, the other participant will be presented with the identical grid and will also be choosing between A, B, and C, or A, B, C, and D. Your objective for each grid will be to maximize the number of points that you score. At the end of the experiment, one of the grids will be chosen randomly from the 12. The number of points that you and the other participant scored in that grid will be converted to pounds Sterling, and you will be paid that in cash at the end of today's session. When you are making your choices, you will not know who you are paired with or what choices they are making, and they will also not know what choices you are

> making. For each grid, please indicate your choice by selecting either A, B, C, or D.

The participants were given the opportunity to seek clarification of anything they did not understand, after which the payoff matrices were presented in succession on their monitors, with Player 1's labels and payoffs shown in blue and Player 2's in red. In each session, the participant in the role of Player 2 was presented with similar material, but written from the perspective of the red player. For the participants in the role of Player 1, the following text was displayed below each payoff matrix to help them interpret the game: "You are the Blue decision maker, choosing between the rows marked *A*, *B,* or *C* [or *D*]. The person you have been paired with is the Red decision maker, choosing between columns *A*, *B,* or *C* [or *D*]. Depending on what you and the other decision maker choose, you will get one of the blue payoffs, and the red decision maker will get one of the red payoffs." This was followed by a full textual summary of the information shown in the payoff matrix, as follows (this example relates to Game 1):

> If you choose A, then:
>
> If Red chooses A, you will get 3, and Red will get 3
>
> If Red chooses B, you will get 1, and Red will get 0
>
> If Red chooses C, you will get 0, and Red will get 2
>
> [and so on . . . .]

The participants then made one-off (unrepeated) strategy choices in each of the 12 games by clicking radio buttons marked *A*, *B*, *C*, or *D*. No feedback was provided. They were able to change their strategy choice at any time until they hit the *Next* button to move on to the following game (returning to previous games was not allowed).

After the participants had indicated their decisions for all 12 games, they were presented with a list, with the order randomized separately for each participant, of eight possible reasons that may have influenced their decisions in choosing between *A*, *B*, *C*, and *D*, and they were asked to indicate on a 7-point Likert scale to what extent they agreed or disagreed with the reason (*Strongly disagree*; *Moderately disagree*; *Slightly disagree*; *Neutral*; *Slightly agree*; *Moderately agree*; *Strongly agree*). The eight reasons were based on an extensive pilot study (N = 127), in which participants were asked to state in their own words their reasons for choosing strategies; the eight items in our list represent the reasons most frequently cited by the participants in the pilot study. Below the list of eight reasons, participants were asked: "If you often used one reason followed by another in the same grid, please indicate here the reasons in the order in which you usually used them. If you tended to use only one reason, please select N/A." Lastly, an open text box was provided for participants to type any additional reasons, not listed, that they might have used, but this did not elicit any new reasons, apart from a response from one participant who cited a purely altruistic reason for choice ("I chose the rows sometimes by maximizing the benefits of the other person").

Data were then downloaded from SurveyGizmo into a preprogrammed Microsoft Excel spreadsheet. In order to calculate the payoffs, data from the participant in the role of Player 2 were matched to each Player 1, and their payoffs were then automatically computed for a randomly preselected game. Participants were paid what they had earned and thanked before they left the laboratory. Participants in the role of Player 2 were remunerated according to their payoffs in their pairing with the first Player 1 who happened to log on.

Results[3] and Discussion

Strategy choices. Modal choices of players in all 12 S-soluble experimental games are shown in Table 1, together with unique predictions of Player 1's strategy choices for each of the theories under investigation. In the eight 3 × 3 games, the modal choice was predicted by cognitive hierarchy Level-1 reasoning in five games, by cognitive hierarchy Level-2 reasoning and strong Stackelberg reasoning in three games, and by team reasoning in two games. In the four 4 × 4 games, the modal choice was predicted by cognitive hierarchy Level-1 rea-

---

[3] Raw data are presented in the supplemental materials. A free online strategy generator for cognitive hierarchy Level-1 and Level-2 reasoning, strong Stackelberg reasoning, and team reasoning can be found at http://hdl.handle.net/2381/27886

Table 1
Experiment 1: Modal Choices of Players in 12 S-Soluble Experimental Games, and Unique Strategy Choice Predictions for Player 1 of Cognitive Hierarchy (CH) Theory for Level-1 and Level-2 Reasoning, Strong Stackelberg Reasoning, and Team Reasoning

Games	Modal choice	CH Level-1	CH Level-2	S. Stack.	Team R.
3 × 3 games					
1	B	B	B	A	C
2	A	B	B	A	C
3	B	B	C	C	A
4	B	B	C	C	A
5	C	C	C	B	A
6	C	C	C	B	A
7	C	A	B	C	C
8	C	A	B	C	C
4 × 4 games					
9	C	C	D	B	A
10	A/B	B	D	C	A
11	B	B	C	D	A
12	C	C	D	B	A

soning in all four games, by team reasoning in one game, and by cognitive hierarchy Level-2 reasoning and strong Stackelberg reasoning in none of the games. On this crude criterion, cognitive hierarchy Level-1 reasoning appears to have outperformed all of the other theories, especially in the 4 × 4 games.

Table 2 shows the choice data in more detail, with frequencies of Player 1 strategy choices for each experimental game and associated chi-square values indicating departures from a null hypothesis of equally likely expected frequencies, significance levels, and values of Cohen's (1988, 1992) effect size index $w = \sqrt{(\chi^2/N)}$. For the 3 × 3 games, all but one of the {A, B, C} strategy choice distributions deviate significantly from chance. The only exception is Game 1, in which the frequencies do not differ

Table 2
*Experiment 1: Frequencies of Player 1 Strategy Choices in 12 Experimental Games, With Chi-Square Values, Significance Levels, and Effect Sizes (*N *= 68)*

Games	CH-L-1	CH-L-2	S. Stack.	Team R.	χ^2	*df*	*p*	*Effect size w*
3 × 3 games								
1	28 (21)	28 (7)	25	15	4.09 (10.82)	2 (3)	.129 (.013)	0.25 (0.40)
2	27 (20)	27 (7)	33	8	15.03 (26.24)	2 (3)	.001 (.000)	0.47 (0.62)
3	31	13 (4)	13 (9)	24	7.27 (28.12)	2 (3)	.026 (.000)	0.33 (0.64)
4	34	21 (6)	21 (15)	13	9.91 (25.29)	2 (3)	.007 (.000)	0.38 (0.61)
5	34 (25)	34 (9)	19	15	8.85 (8.00)	2 (3)	.012 (.046)	0.36 (0.34)
6	39 (29)	39 (10)	21	8	21.38 (17.06)	2 (3)	.000 (.001)	0.56 (0.50)
7	20	7	41 (26)	41 (15)	25.97 (11.41)	2 (3)	.000 (.010)	0.62 (0.41)
8	22	12	34 (22)	34 (12)	10.71 (5.88)	2 (3)	.005 (.117)	0.40 (0.29)
Mean	29.37 (25.25)	22.62 (7.75)	25.88 (21.25)	19.75 (13.75)				
4 × 4 games								
9	25	9	10	24	13.29	3	.004	0.44
10	23	3	19	23	16.00	3	.001	0.48
11	42	8	4	14	52.00	3	.001	0.87
12	35	4	17	12	30.47	3	.001	0.67
Mean	31.25	6.00	12.50	18.25				

Note. For 3 × 3 games, adjusted frequencies and chi-square analyses for theories making overlapping predictions are shown in parentheses. For the effect size index, $w \geq 0.50$ large, $w \geq 0.30$ medium, $w \geq 0.10$ small.

significantly and the effect size is small; in the other 3 × 3 games, the deviations are significant, and the effect sizes are medium or large. For the 3 × 3 games only, adjusted frequencies and associated statistics as explained below are shown in parentheses. For the 4 × 4 games, all strategy choice distributions deviate significantly from chance, and effect sizes are medium (Games 9 and 10) or large (Games 11 and 12).

Order effects. To check for possible order effects, the games were presented to half the players in one randomized order and to the other half in the reverse order. A game-by-game analysis revealed that the distributions of choices do not differ significantly between presentation orders in any of the 12 experimental games, none of the chi-square values even approaching significance.

Parametric analysis of strategy choices. Every player made 12 strategy choices, and each choice corresponded to a prediction of one of the theories under consideration or in some cases, in 3 × 3 games, to a prediction of two of the theories. In the 4 × 4 games, each Player 1 strategy was uniquely predicted by a different theory. These games provided us with a means of computing the number of times that each of the theoretical predictions was corroborated by players' strategy choices.

For 3 × 3 games, the mean percentages of Player 1 strategy choices predicted by each of the theories, in descending order, are as follows (based on the unadjusted raw frequencies in Table 2): cognitive hierarchy Level-1 reasoning, 43%; strong Stackelberg reasoning, 38%; cognitive hierarchy Level-2 reasoning, 33%; and team reasoning, 29%. These percentages sum to more than 100% because, in 3 × 3 games, two of the four theories invariably predict the same strategy choice. For example, in Game 1, both cognitive hierarchy Level-1 reasoning and cognitive hierarchy Level-2 reasoning predict the strategy choice *B*, and 28 players (41%) chose *B*; therefore, all we can infer is that these 28 strategy choices (41% of the choices made) confirm at least one of the two theories that predict *B*. It is unlikely that all 28 of the *B* choices in Game 1 arose from players using cognitive hierarchy Level-2 reasoning, because that method of reasoning appears to have been used much less frequently than cognitive hierarchy Level-1 reasoning in games in which these theories were alone in predicting a particular strategy choice. This suggests that the unadjusted percentages, in addition to summing to more than 100%, may be systematically biased.

To obtain unbiased estimates of the true relative frequencies across all eight 3 × 3 games, we began by examining strategy choices in those games in which one theory alone predicts a particular choice, ignoring games in which another theory makes the same prediction. For example, cognitive hierarchy Level-1 reasoning makes unique predictions in Games 3, 4, 7, and 8, and in those games only, the predicted strategies were chosen by 26.75 players, on average, whereas cognitive hierarchy Level-2 reasoning makes unique predictions in Games 7 and 8, and in those games the predicted strategies were chosen by 9.50 players, on average. These means, and analogous means for the other theories, were used as weights to obtain adjusted relative frequencies for each theory wherever two theories predicted the same strategy choice in a 3 × 3 game. Using these weights in Game 1, for example, the 28 choices confirming either cognitive hierarchy Level-1 reasoning or cognitive hierarchy Level-2 reasoning were adjusted to $28 \times 26.75/(26.75 + 9.50) = 20.66$ for cognitive hierarchy Level-1 reasoning and $28 \times 9.50/(26.75 + 9.50) = 7.34$ for cognitive hierarchy Level-2 reasoning. Adjusted figures relate to the relative frequencies with which the four theories or reasoning processes were chosen, whereas unadjusted data relate to the game strategies chosen from the set $\{A, B, C\}$.

The adjusted relative frequencies and associated statistics are shown in parentheses in Table 2. Converting to percentages for ease of interpretation, the adjusted mean Player 1 strategy choices predicted by each of the theories in 3 × 3 games, in descending order, are: cognitive hierarchy Level-1 reasoning, 37%; strong Stackelberg reasoning, 31%; team reasoning, 20%; and cognitive hierarchy Level-2 reasoning, 11%. Analysis of variance performed on the adjusted frequencies reveals that these four means differ significantly: $F(3, 28) = 17.03$, $p < .001$, $\eta_p^2 = .65$ (large). Post hoc pairwise comparisons using the least significant difference (LSD) test reveal that players chose cognitive hierarchy Level-1 strategies significantly more frequently than team reasoning strategies ($p < .001$) and than cognitive hierarchy Level-2 strategies ($p < .001$), but not significantly more frequently than strong Stackelberg strategies.

Players also chose strong Stackelberg strategies significantly more frequently than team reasoning strategies ($p < .009$) and than cognitive hierarchy Level-2 strategies ($p < .001$), and they chose team reasoning strategies significantly more frequently than cognitive hierarchy Level-2 strategies ($p < .032$).

Relative frequencies for 4 × 4 games, for which no adjustments are required, are also shown in Table 2. For 4 × 4 games, converting again to percentages, the means for Player 1 strategy choices predicted by each of the theories, in descending order, were: cognitive hierarchy Level-1 reasoning, 46%; team reasoning, 27%; strong Stackelberg reasoning, 18%; and cognitive hierarchy Level-2, 9%. Analysis of variance reveals that the four means differ significantly: $F(3, 12) = 10.71, p < .001, \eta_p^2 = .73$ (large). Post hoc pairwise comparisons using the LSD test reveal that players chose cognitive hierarchy Level-1 strategies significantly more frequently than team reasoning strategies ($p = .016$), than strong Stackelberg strategies ($p = .002$), and than cognitive hierarchy Level-2 strategies ($p < .001$). Players also chose team reasoning strategies significantly more frequently than cognitive hierarchy Level-2 strategies ($p = .022$).

Reasons for choices. Results of the analysis of reasons for choices are presented in detail in online supplemental materials. In brief, the findings corroborate the analysis of choice data in revealing that, among the theories under investigation, players rated reasons associated with strong Stackelberg reasoning, team reasoning, and cognitive hierarchy Level-1 reasoning as most influential on their strategy choices, and that avoiding the worst payoff, equality-seeking, and cognitive hierarchy Level-2 reasoning also appear to have been influential.

Avoiding the worst payoff was the most frequently chosen reason of all, and it is not immediately obvious why this is so. Cognitive hierarchy Level-1 reasoning was the most successful theory in explaining actual strategy choices, according to our choice data, but players rated the reason most closely matching that form of reasoning (*I chose rows by working out or estimating the average payoff that I could expect if the other person was equally likely to choose any column, and then choosing the best rows for me on that basis*) as less influential than several other reasons. However, it turns out that "Avoid the Worst" (ATW), a "fast and frugal" heuristic introduced by Gigerenzer and Goldstein (1996), approximates cognitive hierarchy Level-1 choices in many of our experimental games (see also Ert & Erev, 2008). If players avoid any strategy that could yield a zero payoff to them, picking strategies randomly from those that are playable when more than one avoids a possible zero payoff and when none avoids a possible zero payoff, then almost half (47%) of their strategy choices correspond to the predictions of cognitive hierarchy Level-1 reasoning in our games. We infer from this that some, at least, of the cognitive hierarchy Level-1 choices may have been generated by the ATW heuristic.

Our analysis of reasons for choices also revealed that 79% of players used two or more strategies in the same game. Reasons associated with avoiding the worst payoff, strong Stackelberg reasoning, and team reasoning were used most frequently in conjunction with each other. Players sometimes began by reasoning strategically, using strong Stackelberg reasoning, but then switched to team reasoning or cognitive hierarchy Level-1 reasoning, presumably because these alternatives were easier and less demanding than strong Stackelberg reasoning.

Experiment 2

The principal aim of Experiment 2 was to check the robustness and replicability of the results of Experiment 1 with a fresh sample of 59 players. A secondary aim was to investigate the hypothesis that players use the simplest and easiest form of the ATW heuristic by merely avoiding any strategy that risks a possible zero payoff. We tested this hypothesis by means of a within-subjects comparison of strategy choices in 12 games identical to the games used in Experiment 1 with choices in versions of the same 12 games with three units added to every payoff, thus eliminating zero payoffs altogether.

Full details of Experiment 2 are presented in online supplemental materials. Modal choices in the original versions of the games were almost identical to those observed in Experiment 1 (see Table 1), fully replicating our main findings. Comparing original and plus-3 versions of the games in Experiment 2, the modal choices were the same in eight of the 12 games, and in those in which they differed, the discrepancies were very small, suggesting that although players may have avoided the worst payoffs, they did not merely avoid zero payoffs. Self-reported reasons for choices were strikingly similar to those of Experiment 1, although the reason associated with Stackelberg reasoning was rated as fractionally more influential than avoiding the worst payoff, reversing the marginal lead of avoiding the worst payoff in Experiment 1. Most players (88%) reported considering more than one reason in the same game, and it is noteworthy that 28% of players who first used a reason associated with strong Stackelberg reasoning reported switching from it to either avoiding the worst payoff or a reason associated with cognitive hierarchy Level-1 reasoning.

General Discussion

Results of the experiments reported here are highly consistent with each other, and they suggest that cognitive hierarchy Level-1 reasoning, strong Stackelberg reasoning, and team reasoning each played a part in explaining players' strategy choices in 3 × 3 and 4 × 4 experimental games. Cognitive hierarchy Level-1 reasoning was most influential, especially in 4 × 4 games, but strong Stackelberg reasoning was also influential in 3 × 3 games, and team reasoning in both 3 × 3 and 4 × 4 games. It seems reasonable to infer that coordination is partly explained by all three theories. Most players reported that they considered two or more reasoning processes before making their choices. Furthermore, it is possible that social projection theory may also provide part of the explanation in symmetric games. Although it might have been satisfying, from the point of view of simplicity and clarity, if one theory had emerged as a single and sovereign explanation in all of the games that we studied, the picture turns out to be more complex.

Our results also suggest that the Avoid the Worst (ATW) heuristic may have steered players toward cognitive hierarchy Level-1 choices in some cases, because ATW turns out to yield cognitive hierarchy Level-1 choices in a substantial proportion (almost half) of our games. Part of the motivation for Experiment 2 was to test the hypothesis that players used the most elementary version of ATW, simply avoiding strategies that entail a risk of a zero payoff. The results provided no clear evidence to support the "avoid zero" hypothesis, but that does not rule out ATW as a possible explanation for some cognitive hierarchy Level-1 strategy choices. ATW can be implemented by avoiding strategies that entail the risk of the lowest payoff in a game, whatever it might be, and it is only marginally more difficult when the lowest payoff is 3 rather than 0, as it was in all the plus-3 games in Experiment 2. In the players' self-reported reasons for choices, ATW emerged as one of the most popular reasons of all, and it seems reasonable to infer that it probably accounted for some cognitive hierarchy Level-1 strategy choices.

Our results are consistent with the findings of Camerer et al. (2004) and others who reported evidence for cognitive hierarchy Level-1 and Level-2 reasoning, and also with those of Colman, Pulford, and Rose (2008a) and Butler (2012), who reported evidence for team reasoning, using very different types of games and experimental methods. Our results and our interpretation of them are also consistent with the findings of Bardsley et al. (2010), who reported the results of two separate experiments designed to compare cognitive hierarchy and team reasoning theories against each other, using experimental coordination games and research methodologies radically different from our own. The first of their experiments appeared to support team reasoning and the second cognitive hierarchy theory. Bardsley et al. discussed various ways in which this difference might be due to differences in the way the players approached the coordination tasks, but an examination of their test materials reveals that the games used in their first experiment all had five pure strategies, whereas those used in their second experiment averaged slightly over four ($M = 4.36$). Team reasoning is easier than cognitive hierarchy reasoning (Level-1 and above) in larger and more complicated games, and it is therefore possible that the difference is explained by Bardsley et al.'s players tending to use team reasoning more frequently in games that were slightly larger and more complicated.

Our findings are also consistent with those of Colman and Stirk (1998), who reported evidence for Stackelberg reasoning in 2 × 2 games. In the experiments reported in this article, strong Stackelberg reasoning was abundantly evident in the strategy choice data in 3 × 3 games but much less so in 4 × 4 games. The players' self-reported reasons for choices confirm that they were influenced by strong Stackelberg reasoning, but the choice data suggest that this influence was felt chiefly in the 3 × 3 games and less strongly in the 4 × 4 games. Taken together with the evidence of Colman and Stirk, this suggests that strong Stackelberg reasoning may be used quite frequently in relatively simple 2 × 2 and 3 × 3 games but tends to be abandoned in favor of team reasoning and especially cognitive hierarchy Level-1 reasoning in more complicated 4 × 4 games. This is not difficult to understand because strong Stackelberg reasoning involves more complex computation and imposes greater demands than cognitive hierarchy Level-1 reasoning or team reasoning. From Player 1's perspective, cognitive hierarchy Level-1 reasoning can be implemented by simply glancing along the rows of the payoff matrix and estimating which row contains the highest payoffs to Player 1, and it can be approximated in many games by simply avoiding the worst payoff. Team reasoning can be accomplished even more easily, especially in larger games, by scanning the payoff matrix for the cell with the greatest payoff sum, and then checking that it is unique. In contrast, strong Stackelberg reasoning by Player 1 involves noting and remembering Player 2's best replies to every available strategy, earmarking the most profitable one for Player 1, then doing the same from Player 2's perspective, and finally checking that neither player can benefit by deviating unilaterally from the resulting strategy pair or outcome. This requires sequential strategic thinking and is obviously more demanding in terms of mental computation and working memory capacity than cognitive hierarchy Level-1 reasoning or team reasoning.

Orthodox game theory is incapable of explaining coordination, and several theories that have been proffered either fail to provide an adequate explanation or apply to symmetric games only, but cognitive hierarchy theory, strong Stackelberg reasoning, and theories of team reasoning offer potential explanations, even in asymmetric games. The major problems with cognitive hierarchy theory as an explanation of coordination are that it fails to explain coordination in some common interest games, such as the Stag Hunt game in Figure 2, and that it is vitiated by an implausible asymmetry—an assumption that players never credit their co-players with the same depth of strategic reasoning as they enjoy themselves. The main problem with theories of team reasoning is their abandonment of methodological individualism and consequently also of von Neumann–Morgenstern utilities. The theory of strong Stackelberg reasoning avoids these and other theoretical problems, and in that sense it seems preferable, but the results of our experiment suggest that all three theories provide part of the explanation, at least in experimental games.

Taking into account Experiments 1 and 2, our experimental findings suggest that cognitive hierarchy Level-1 reasoning was used in 37%–46% of choices in 3 × 3 games and 46%–50% of choices in 4 × 4 games, whereas strong Stackelberg reasoning was used in 31%–36% in 3 × 3 games and 17%–19% in 4 × 4 games. Team reasoning was used in 13%–20% of choices in 3 × 3 games and 26%–27% in 4 × 4 games. These differences suggest that cognitive hierarchy Level-1 reasoning, perhaps facilitated by the ATW heuristic, is influential in games of both sizes, that strong Stackelberg reasoning is influential especially in the smaller and simpler games, and that team reasoning becomes relatively more influential in the larger, more complex games. However, it is worth noting that people facing potentially life-changing strategic decisions may be willing and able to apply strong Stackelberg reasoning even in 4 × 4 and larger games.

Are there features of interactive decisions that determine which theory will predominate in any particular case? In general, this is a question for later research, but a consideration of the relative ease or difficulty of the various psychological processes implied by the theories offers some clues. Strong Stackelberg reasoning requires relatively slow, deliberate, and effortful *System 2* thinking (Kahneman, 2011) and tends to become prohibitively difficult and demanding in games that are larger and more complex than simple two-strategy and three-strategy games. Cognitive hierarchy Level-1 reasoning, especially if approximated by the fast and frugal Avoid the Worst (ATW) heuristic, provides a method of reasoning typical of automatic, relatively effortless *System 1* thinking. For collectively motivated players, team reasoning provides a method of play that is also easy to perform, even in large games.

It is generally acknowledged that strategy choices, like individual risky choices, are not strictly deterministic. If we assume that they are not merely error-prone but inherently probabilistic, such that decision processes map probabilistically into strategy sets, then it is conceivable that an intrinsically stochastic overarching theory may eventually emerge that fits the data well.

References

Acevedo, M., & Krueger, J. I. (2004). Two egocentric sources of the decision to vote: The voter's illusion and the belief in personal relevance. *Political Psychology, 25,* 115–134. doi:10.1111/j.1467-9221.2004.00359.x

Acevedo, M., & Krueger, J. I. (2005). Evidential reasoning in the Prisoner's Dilemma. *The American Journal of Psychology, 118,* 431–457.

Anand, P. (1990). Two types of utility: An experimental investigation into the prevalence of causal and evidential utility maximisation. *Greek Economic Review, 12,* 58–74.

Anderlini, L. (1999). Communication, computability, and common interest games. *Games and Economic Behavior, 27,* 1–37. doi:10.1006/game.1998.0652

Anderlini, L., & Sabourian, H. (1995). Cooperation and effective computability. *Econometrica, 63,* 1337–1369. doi:10.2307/2171773

Arnsperger, C., & Varoufakis, Y. (2003). Toward a theory of solidarity. *Erkenntnis, 59,* 157–188. doi:10.1023/A:1024630228818

Aumann, R. J., & Sorin, S. (1989). Cooperation and bounded recall. *Games and Economic Behavior, 1,* 5–39. doi:10.1016/0899-8256(89)90003-1

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics, 53,* 117–147. doi:10.1006/reec.1999.0188

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (N. Gold & R. Sugden, Eds.). Princeton, NJ: Princeton University Press.

Bacharach, M., & Stahl, D. O. (2000). Variable-frame level-*n* theory. *Games and Economic Behavior, 32,* 220–246. doi:10.1006/game.2000.0796

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal, 120,* 40–79. doi:10.1111/j.1468-0297.2009.02304.x

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90,* 166–193. doi:10.1257/aer.90.1.166

Brosnan, S. F., & de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature, 13,* 129–152. doi:10.1007/s12110-002-1017-2

Butler, D. J. (2012). A choice for "me" or for "us"? Using we-reasoning to predict cooperation and coordination in games. *Theory and Decision, 73,* 53–76. doi:10.1007/s11238-011-9270-7

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics, 119,* 861–898. doi:10.1162/0033553041502225

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Colman, A. M. (2003a). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences, 26,* 139–153. doi:10.1017/S0140525X03000050

Colman, A. M. (2003b). Depth of strategic reasoning in games. *Trends in Cognitive Sciences, 7,* 2–4. doi:10.1016/S1364-6613(02)00006-2

Colman, A. M., & Bacharach, M. (1997). Payoff dominance and the Stackelberg heuristic. *Theory and Decision, 43,* 1–19. doi:10.1023/A:1004911723951

Colman, A. M., Körner, T. W., Musy, O., & Tazdaït, T. (2011). Mutual support in games: Some properties of Berge equilibria. *Journal of Mathematical Psychology, 55,* 166–175. doi:10.1016/j.jmp.2011.02.001

Colman, A. M., Pulford, B. D., & Rose, J. (2008a). Collective rationality in interactive decisions: Evidence for team reasoning. *Acta Psychologica, 128,* 387–397. doi:10.1016/j.actpsy.2007.08.003

Colman, A. M., Pulford, B. D., & Rose, J. (2008b). Team reasoning and collective rationality: Piercing the veil of obviousness. *Acta Psychologica, 128,* 409–412. doi:10.1016/j.actpsy.2008.04.001

Colman, A. M., & Stirk, J. A. (1998). Stackelberg reasoning in mixed-motive games: An experimental investigation. *Journal of Economic Psychology, 19,* 279–293. doi:10.1016/S0167-4870(98)00008-7

Cooper, R. W., DeJong, D. V., Forsythe, R., & Ross, T. W. (1990). Selection criteria in coordination games: Some experimental results. *American Economic Review, 80,* 218–233.

Crawford, V. P., & Haller, H. (1990). Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica, 58,* 571–595. doi:10.2307/2938191

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics, 1,* 115–131. doi:10.1007/BF01669298

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25,* 1–17. doi:10.1016/0022-1031(89)90036-X

Dawes, R. M. (1990). The potential non-falsity of the false consensus effect. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 179–199). Chicago, IL: University of Chicago Press.

Dawes, R. M. (2000). A theory of irrationality as a "reasonable" response to an incomplete specification. *Synthese, 122,* 133–163. doi:10.1023/A:1005224211316

Eells, E. (1984). Newcomb's many solutions. *Theory and Decision, 16,* 59–105. doi:10.1007/BF00141675

Eells, E. (1989). The popcorn problem: Sobel on evidential decision theory and deliberation-probability dynamics. *Synthese, 81,* 9–20. doi:10.1007/BF00869342

Ellingsen, T., & Östling, R. (2010). When does communication improve coordination? *American Economic Review, 100,* 1695–1724. doi:10.1257/aer.100.4.1695

Elster, J. (1982). The case for methodological individualism. *Theory and Society, 11,* 453–482.

Ert, E., & Erev, I. (2008). The rejection of attractive gambles, loss aversion, and the lemon avoidance heuristic. *Journal of Economic Psychology, 29,* 715–723. doi:10.1016/j.joep.2007.06.003

Farrell, J. (1988). Communication, coordination and Nash equilibrium. *Economics Letters, 27,* 209–214. doi:10.1016/0165-1765(88)90172-3

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114,* 817–868. doi:10.1162/003355399556151

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669. doi:10.1037/0033-295X.103.4.650

Gilbert, M. (1989). Rationality and salience. *Philosophical Studies, 57,* 61–77. doi:10.1007/BF00355662

Gintis, H. (2003). A critique of team and Stackelberg reasoning. *Behavioral and Brain Sciences, 26,* 160–161. doi:10.1017/S0140525X03300056

Gold, N., & Sugden, R. (2007). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment* (pp. 280–312). Oxford, UK: Oxford University Press.

Harsanyi, J. C., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Hausman, D. M. (2003). Rational belief and social interaction. *Behavioral and Brain Sciences, 26,* 163–164. doi:10.1017/S0140525X03330055

Hedden, T., & Zhang, J. (2002). What do you think I think you think? Strategic reasoning in matrix games. *Cognition, 85,* 1–36. doi:10.1016/S0010-0277(02)00054-9

Howard, N. (1971). *Paradoxes of rationality: Theory of metagames and political behavior*. Cambridge, MA: MIT Press.

Janssen, M. C. W. (2001). Rationalising focal points. *Theory and Decision, 50,* 119–148. doi:10.1023/A:1010349014718

Janssen, M. C. W. (2006). On the strategic use of focal points in bargaining situations. *Journal of Economic Psychology, 27,* 622–634. doi:10.1016/j.joep.2006.04.006

Jeffrey, R. (1983). *The logic of decision* (2nd ed.). Chicago, IL: University of Chicago Press.

Jeffrey, R. (2004). *Subjective probability: The real thing*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511816161

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus & Giroux.

Kendall, M. G. (1945). *The advanced theory of statistics* (2nd ed.). London, UK: Charles Griffin.

Keynes, J. M. (1921). *A treatise on probability*. London, UK: Macmillan.

Krueger, J. I. (1998). On the perception of social consensus. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 163–240). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(08)60384-6

Krueger, J. I. (2007). From social projection to social behavior. *European Review of Social Psychology, 18,* 1–35. doi:10.1080/10463280701284645

Krueger, J. I. (2008). Methodological individualism in experimental games: Not so easily dismissed. *Acta Psychologica, 128,* 398–401. doi:10.1016/j.actpsy.2007.12.011

Krueger, J. I., & Acevedo, M. (2005). Social projection and the psychology of choice. In M. D. Alicke, D. Dunning, & J. I. Krueger (Eds.), *The self in social perception* (pp. 17–41). New York, NY: Psychology Press.

Krueger, J. I., DiDonato, T. E., & Freestone, D. (2012). Social projection can solve social dilemmas. *Psychological Inquiry, 23,* 1–27. doi:10.1080/1047840X.2012.641167

Krueger, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 472–494). New York, NY: Guilford Press.

Lee, J. (2008). The effect of the background risk in a simple chance improving decision model. *Journal of Risk and Uncertainty, 36,* 19–41. doi:10.1007/s11166-007-9028-3

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York, NY: Wiley.

Monterosso, J., & Ainslie, G. (2003). Game theory need not abandon individual maximization. *Behavioral and Brain Sciences, 26,* 171. doi:10.1017/S0140525X03400058

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the USA, 36,* 48–49. doi:10.1073/pnas.36.1.48

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics, 54,* 286–295. doi:10.2307/1969529

Nozick, R. (1993). *The nature of rationality*. Princeton, NJ: Princeton University Press.

Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and the voter's illusion. *Journal of Personality and Social Psychology, 46,* 237–248. doi:10.1037/0022-3514.46.2.237

Rabin, M. (1994). A model of pre-game communication. *Journal of Economic Theory, 63,* 370–391. doi:10.1006/jeth.1994.1047

Rousseau, J.-J. (1755). *Discours sur l'origine et les fondements de l'inégalité parmi les hommes* [Discourse on the origin and the foundations of inequality among men]. In J.-J. Rousseau (Ed.), *Oeuvres complètes* [Complete works] (Vol. 3). Paris, France: Edition Pléiade.

Rusbult, C. E., & Van Lange, P. A. M. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology, 54,* 351–375. doi:10.1146/annurev.psych.54.101601.145059

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Selten, R. (1975). Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory, 4,* 25–55. doi:10.1007/BF01766400

Smerilli, A. (2012). We-thinking and vacillation between frames: Filling a gap in Bacharach's theory. *Theory and Decision*. Advance online publication. doi10.1007/s11238-012-9294-7

Stackelberg, H. von. (1934). *Marktform und Gleichgewicht* [Market Structure and Equilibrium]. Berlin, Germany: Verlag Julius Springer.

Stahl, D. O., & Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization, 25,* 309–327. doi:10.1016/0167-2681(94)90103-1

Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior, 10,* 218–254. doi:10.1006/game.1995.1031

Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review, 81,* 971–978.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behaviour. *Social Philosophy and Policy, 10,* 69–89. doi:10.1017/S0265052500004027

Sugden, R. (2005). The logic of team reasoning. In N. Gold (Ed.), *Teamwork: Multi-disciplinary perspectives* (pp. 181–199). Basingstoke, UK: Palgrave Macmillan.

Tan, J. H. W., & Zizzo, D. J. (2008). Groups, cooperation and conflict in games. *The Journal of Socio-Economics, 37,* 1–17. doi:10.1016/j.socec.2006.12.023

Trivers, R. (2005). Reciprocal altruism: 30 years later. In C. P. van Schaik & P. M. Kappeler (Eds.), *Cooperation in primates and humans: Mechanisms and evolution* (pp. 67–83). Berlin, Germany: Springer-Verlag.

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77,* 337–349. doi:10.1037/0022-3514.77.2.337

Van Lange, P. A. M. (2000). Beyond self-interest: A set of propositions relevant to interpersonal orientations. *European Review of Social Psychology, 11,* 297–331. doi:10.1080/14792772043000068

Van Lange, P. A. M. (2008). Collective rationality: The integrative model explains it (as) well. *Acta Psychologica, 128,* 405–408. doi:10.1016/j.actpsy.2008.01.005

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Weber, R. A., Camerer, C. F., & Knez, M. (2004). Timing and virtual observability in ultimatum bargaining and "weak link" coordination games. *Experimental Economics, 7,* 25–48. doi:10.1023/A:1026257921046

(*Appendices follow*)

Appendix A

A Simple Proof of Laplace's Rule of Succession

Theorem

An urn containing r red and N – r black balls is chosen at random from a collection of N + 1 urns containing r = 0, 1 . . ., N red balls respectively. From the chosen urn, a ball is drawn at random and replaced, and the process is repeated m times. If the m draws produce m red balls, then the probability that the next ball to be drawn will also be red is approximately (*m* + *1*)/(*m* + *2*).

Proof

The probability of choosing a red ball on the first draw is r/N, and the joint probability of choosing a red ball on m successive draws is $(r/N)^m$. Call the event of drawing m successive red balls A. Then

$$P(A) = \frac{1^m + 2^m + \cdots + N^m}{N^m(N+1)}.$$

If we call the event of drawing a red ball on the $m + 1$st draw B, then the probability of B is simply the probability of drawing $m + 1$ balls in succession, hence

$$P(B) = \frac{1^{m+1} + 2^{m+1} + \cdots + N^{m+1}}{N^{m+1}(N+1)}.$$

The probability of drawing a red ball on the $m + 1$st draw is the probability of the event (B and A). According to a well-known rule of conditional probability, $P(B \mid A) = P(B \text{ and } A)/P(A)$. However, event B of drawing $m + 1$ red balls is equivalent to the event (B and A), hence $P(B \text{ and } A) = P(B)$ and $P(B \mid A) = P(B)/P(A)$. For large N, we have

$$P(A) \approx \frac{1}{N^m(N+1)} \int_0^N x^m dx = \frac{N}{N+1} \times \frac{1}{m+1} \approx \frac{1}{m+1}.$$

Similarly, $P(B) \approx 1/(m + 2)$. Therefore,

$$P(B \mid A) = \frac{P(B)}{P(A)} \approx \frac{m+1}{m+2}.$$

Remark

If just one ball is drawn from the urn and is found to be red, then $m = 1$, and the probability that a second ball drawn from the same urn would also be red is approximately 2/3.

(*Appendices continue*)

Appendix B

Strong Stackelberg Reasoning: Formal Development

We begin by defining a two-player game

$$G \ := \ \langle S_1, S_2; \pi_1, \pi_2 \rangle,$$

where S_i is Player i's strategy set and π_i is Player i's payoff function, $i \in \{1, 2\}$. We assume that S_i is finite and contains at least two strategies. Player i's payoff function is denoted by $\pi_i : S \rightarrow \mathbb{R}$. The possible outcomes of the game are members of the set $S = S_1 \times S_2$, and $\mathbf{s} = (s_1, s_2) \in S$ is a particular strategy profile determining an outcome. A strategy profile $\mathbf{s}' = (s_1', s_2')$ is a Nash equilibrium if both

$$\pi_1(s_1', s_2') \geq \pi_1(s_1, s_2'), \tag{1}$$

$$\pi_2(s_1', s_2') \geq \pi_2(s_1', s_2), \tag{2}$$

hold for all $s_1 \in S_1$ and all $s_2 \in S_2$. If $\mathbf{s}$ and $\mathbf{t}$ are two distinct outcomes of a game, then $\mathbf{s}$ strongly payoff-dominates $\mathbf{t}$ if $\pi_i(\mathbf{s}) > \pi_i(\mathbf{t})$ for $i \in \{1, 2\}$.

We use the notation $C_i s_i$ to denote the event of Player i choosing strategy i. In a game involving Players i and j, the distinctive assumption of strong Stackelberg reasoning is then:

$$C_i s_i \Rightarrow i \text{ believes that } j \text{ knows that } C_i s_i$$

for all $s_i \in S_i$ and for $i, j \in \{1, 2\}$. Both players act as though they believe that whatever strategy they choose will be anticipated by their co-player. The standard common knowledge and rationality assumptions of game theory apply, therefore each player chooses the strategy that yields the best payoff given the belief that the co-player will choose a best reply to any strategy.

Given Player i's belief that Player j will choose a best reply to any strategy that Player i might choose, Player i acts as though believing that Player j's strategy choice depends on Player i's. Thus, from Player i's perspective,

$$s_j = \beta(s_i), \tag{3}$$

where $\beta(s_i)$ is a correspondence assigning Player j's best replies to s_i. If $\beta(s_i)$ assigns a unique element s_j to a particular strategy s_i, then $\beta(s_i)$ is a *strong best reply* to s_i in the sense of Harsanyi and Selten (1988, p. 39), and if this holds for all i, then β is a function. We note the slightly weaker criterion of strong best replies defined when both of the following hold for every strategy pair $\mathbf{s} = (s_1, s_2)$, $\mathbf{t} = (t_1, t_2)$:

$$\pi_1(\mathbf{s}) > \pi_1(\mathbf{t}) \Rightarrow \pi_2(\mathbf{s}) \neq \pi_2(\mathbf{t}), \tag{4}$$

$$\pi_2(\mathbf{s}) > \pi_2(\mathbf{t}) \Rightarrow \pi_1(\mathbf{s}) \neq \pi_1(\mathbf{t}). \tag{5}$$

The condition defined by (4) and (5) means that if a player strictly prefers one outcome to another, then the co-player is not indifferent between those two outcomes, and it ensures that all best replies are strong. We define Player 1's *strong Stackelberg strategy* as the strategy $s_1 = s_1^*$ that maximizes the payoff

$$\pi_1(s_1, \beta(s_1)).$$

Similarly, we define a strong Stackelberg strategy for Player 2 as the strategy $s_2 = s_2^*$ that maximizes the payoff

$$\pi_2(s_2, \beta(s_2)).$$

In any game in which the condition defined by (4) and (5) for strong best replies is met, strong Stackelberg strategies s_i^* are well defined. If the condition for strong best replies does not hold, it is still possible for strong Stackelberg strategies s_i^* to be well defined, provided only that the maximum payoff to Player 1 in each column and the maximum payoff to Player 2 in each row are unique.

(*Appendices continue*)

Strong Stackelberg strategies may or may not form Nash equilibria. In any game in which strong Stackelberg strategies s_i^* are well defined, if $\mathbf{s}^* = (s_1^*, s_2^*)$ is a Nash equilibrium, we describe the game as being *S-soluble*. In every game that is S-soluble, the theory predicts that Stackelberg-reasoning players will choose and play their strong Stackelberg strategies, and we call $\mathbf{s}^*$ the *strong Stackelberg solution* of the game. Games in which strong Stackelberg strategies are not well defined or are not in equilibrium are *non-S-soluble*, and in such games the theory is indeterminate.

Received March 29, 2013
Revision received April 14, 2013
Accepted April 16, 2013 ■

Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx.