

The Psychology of Coordination and Common Knowledge

Kyle A. Thomas
Harvard University

Peter DeScioli
Harvard University and Stony Brook University

Omar Sultan Haque and Steven Pinker
Harvard University

Research on human cooperation has concentrated on the puzzle of altruism, in which 1 actor incurs a cost to benefit another, and the psychology of reciprocity, which evolved to solve this problem. We examine the complementary puzzle of mutualism, in which actors can benefit each other simultaneously, and the psychology of coordination, which ensures such benefits. Coordination is facilitated by *common knowledge*: the recursive belief state in which A knows X, B knows X, A knows that B knows X, B knows that A knows X, ad infinitum. We test whether people are sensitive to common knowledge when deciding whether to engage in risky coordination. Participants decided between working alone for a certain profit and working together for a potentially higher profit that they would receive only if their partner made the same choice. Results showed that more participants attempted risky coordination when they and their prospective partner had common knowledge of the payoffs (broadcast over a loudspeaker) than when they had only shared knowledge (conveyed to both by a messenger) or private knowledge (revealed to each partner separately). These results support the hypothesis that people represent common knowledge as a distinct cognitive category that licenses them to coordinate with others for mutual gain. We discuss how this hypothesis can provide a unified explanation for diverse phenomena in human social life, including recursive mentalizing, performative speech acts, public protests, hypocrisy, and self-conscious emotional expressions.

Keywords: common knowledge, coordination, theory of mind, cooperation, mutualism

A strange and ethereal protest took place in Belarus during the summer of 2011, consisting solely of protesters' phones ringing simultaneously. Police swarmed the event, recorded who was there, and made aggressive arrests (Barry, 2011). What were the protesters trying to accomplish? And why were the police concerned with such a seemingly trivial event?

People interact in a variety of situations in which they need to coordinate their actions to achieve common goals, such as opposing unfair governments, capturing gains in trade, agreeing on the use of standard symbols and protocols, and countless everyday activities, such as scheduling meetings, contributing complemen-

tary courses to potluck dinners, and carrying two ends of a heavy object. Because it may be costly to engage in a coordinated activity when no one else does so, attempts to coordinate can be risky when it is unclear what other people will do. In repressive regimes a single protester risks prosecution and violence, a risk that can be mitigated only by overwhelming numbers of people successfully coordinating their actions: If one protester shows up he gets shot, but if a million show up they may send the dictator packing. In these situations, even modest displays of synchrony, such as simultaneous phone rings, can set the stage for larger scale coordination. However, even when it is clear that other people want to work together, coordination can be a challenge. Exactly how, for instance, do thousands of would-be protesters converge on a single time and place to voice their concerns?

Coordination problems are a subtopic in the psychology of cooperation. Though cooperation has become a burgeoning area in psychology, economics, and evolutionary biology, research and theory have concentrated on the subtype of cooperation that is *altruistic* (in the biological sense): A cooperator confers a benefit on a partner at a cost to himself. Altruistic cooperation has received the lion's share of attention because it raises the evolutionary puzzle of how a behavior that harms the actor could be selected for. The paradox is often captured in the game-theoretic scenario of the prisoner's dilemma, and the challenge to the psychologist is in characterizing the cognitive abilities and emotional motives that allow humans to surmount it. These include the ability to recognize individuals and detect cheaters and a suite of emotions that police

This article was published Online First August 11, 2014.

Kyle A. Thomas, Department of Psychology, Harvard University; Peter DeScioli, Department of Psychology, Harvard University, and Department of Political Science, Stony Brook University; Omar Sultan Haque and Steven Pinker, Department of Psychology, Harvard University.

Omar Sultan Haque is now also at the Department of Psychiatry, Brown University Medical School.

This work was first presented at the 24th annual meeting of the Human Behavior and Evolution Society in Albuquerque, New Mexico, June 2012. We thank Moshe Hoffman for providing feedback on a draft of this article and Natalie Aharon, Cheng Li, Joel Martinez, Pooja Ami Patel, and Sara Paul for helping with data collection and analysis.

Correspondence concerning this article should be addressed to Kyle A. Thomas, Department of Psychology, Harvard University, William James Hall 964, 33 Kirkland Street, Cambridge, MA 02138. E-mail: kathomas@fas.harvard.edu

reciprocation, such as sympathy, anger, gratitude, forgiveness, guilt, and trust (Cosmides & Tooby, 1992, 2005; Trivers, 1971).

Coordination, in contrast, is *mutualistic*: Each cooperator confers a benefit on the other while simultaneously conferring a benefit on himself or herself. Despite this convergence of interests, coordination, too, poses an evolutionary challenge. The challenge is not motivational but epistemological: accurately representing the other actor's state of knowledge. The epistemological problem results from the difficulty of converging on a single solution when more than one is available. For instance, two friends both benefit if they meet at Starbucks or at Peet's, but for this to happen each friend has to know that the other knows which location they have agreed upon.

If this problem can be resolved, the incentives of the game pose no further obstacle, and can even help guide optimal behavior rather than hinder it (Lewis, 1969; Schelling, 1960; Skyrms, 2004). The paradigm game-theoretic model of a coordination problem is the stag hunt, introduced by Jean-Jacques Rousseau (1754/1984; Skyrms, 2004). Two hunters set out in the morning either to hunt stag together (a large payoff) or to hunt rabbit separately (a small payoff); a single hunter cannot fell a stag and will return empty-handed (a high opportunity cost). To attain the highest payoff, each hunter not only must know that stag offers higher payoffs but also must know that the other hunter knows the payoffs, know that the other hunter knows that he or she knows the payoffs, and so on.

Yet, despite this epistemological problem, humans are adept at achieving coordination. Protesters meet up in Tahrir Square at 5 p.m. on Friday, different suppliers produce the parts for a complex product, allied battalions converge on an enemy, diners use the bread plate to the left, coworkers in a building settle on an informal name for a meeting space. Given a long evolutionary history of group living, human cognition may have been shaped by natural selection to solve coordination problems (Tooby & Cosmides, 2010; Tooby, Cosmides, & Price, 2006). If game theorists are correct that common knowledge is needed for coordination, then humans might have cognitive mechanisms for recognizing it.

This article attempts to begin to redress the imbalance in the literature on the psychology of cooperation by exploring the epistemological challenges and the possible cognitive and motivational adaptations surrounding the problem of mutualistic coordination.¹ We focus on a special kind of representation called *common knowledge* (sometimes called *mutual knowledge* or *common ground*; Clark, 1996; Clark & Marshall, 1981; Lewis, 1969; Pinker, 2007; Rubinstein, 1989; Schelling, 1960; Smith, 1982). Common knowledge is defined as an infinite string of embedded levels of knowledge (i.e., Michael knows X; Lisa knows X; Michael knows that Lisa knows X; Lisa knows that Michael knows X; Michael knows that Lisa knows that Michael knows X; ad infinitum).

The infinite levels of knowledge required for common knowledge may seem to present a different kind of epistemological problem, namely, that a finite mind cannot represent an infinite set of nested propositions. However, people need not represent each level of knowledge explicitly but could simply represent a recursive formula that entails all levels of knowledge, such as $Y = \text{"Everyone knows } X, \text{ and everyone knows } Y,"$ or even just a single symbol that indicates the state of common knowledge itself (Clark, 1996; Pinker, 2007). This formula or symbol, moreover, can be

activated in people's minds by any salient public signal that reliably causes the knowledge, such as a message broadcast on a loudspeaker: Everyone who receives the signal knows that everyone else has received it and can deduce that everyone else can deduce that, ad infinitum (Aumann, 1976).

Nor is it necessary that the commonly entertained propositions be known with absolute certainty, which is often impossible in real-world environments. Coordination may be achieved with the weaker notion of *common belief*, in which two agents each believe that a proposition is likely to be true with probability at least p , each believes that the other believes it with probability at least p , and so on (Monderer & Samet, 1989). For any situation with a stag-hunt payoff structure, there is a minimum level of p , whose value depends on the relative advantage of coordination over acting alone, for which it is rational for agents with *common p -belief* to choose to coordinate (Dalkiran, Hoffman, Paturi, Ricketts, & Vattani, 2012). In the rest of this article, we will use the term *common knowledge* broadly, to include "sufficiently high common p -belief."

Common knowledge can be contrasted with what we will refer to as *shared knowledge*, any string of embedded levels of knowledge that falls short of infinity, and with *private knowledge*, knowledge that individuals possess without knowing whether anyone else possesses it. Shared knowledge can be further broken down into distinct levels, such as *second-order* or *secondary* knowledge, in which A knows that B knows X but nothing else, and *third-order* or *tertiary* knowledge, in which A knows that B knows that A knows X but nothing else. Common knowledge is intimately connected with the logical problem of coordination; in theory, coordination can be irrational without it. With the help of four experiments in which participants are given the opportunity to engage in a simple form of economic cooperation, we examine the extent to which people really do depend on common knowledge and other levels of knowledge to achieve coordination.

The Game Theory of Coordination and Common Knowledge

Research in game theory on coordination games shows why shared knowledge may be insufficient for coordination. Technically, coordination games are situations of interdependent decision making that have multiple equilibria. Conceptually, they are situations in which two or more people each make a decision, with the potential to achieve mutual benefits only if their decisions are consistent (Lewis, 1969; Schelling, 1960). The rendezvous example is a coordination game, because both friends benefit from choosing the same location, but that location could be either Starbucks or Peet's. To choose among multiple solutions an individual must take into account what she expects the other actor to do. However, what another actor is likely to do is in turn dependent upon his expectations of what she will do, leading to interdependent expectations that generate an infinite recursion of embedded beliefs.

¹ A PsycINFO search reveals that in the years 1992–2013, 1,936 papers listed *altruism* as a major subject heading or keyword, whereas only 71 listed *mutualism* (and most of these were for studies of nonhuman animals). There were 400 references to the prisoner's dilemma but only 4 to the stag hunt.

A classic paper demonstrated the importance of common knowledge for maximizing payoffs from a coordination game: [Rubinstein \(1989\)](#) developed a model that showed that under a specific, restrictive set of assumptions, any level of knowledge short of common knowledge is no better than no knowledge at all. Subsequent work has suggested that this conclusion was too strong, and that shared knowledge or less-than-certain beliefs can enable coordination better than private knowledge ([Binmore & Samuelson, 2001](#); [Dalkiran et al., 2012](#); [Monderer & Samet, 1989](#)). However, even in these models, common knowledge has a privileged role to play in facilitating coordination, in part because it avoids a second-order coordination problem presented by shared knowledge. With shared knowledge, people must decide how many levels of shared knowledge are enough to attempt coordination: How can individuals be certain that everyone requires the same number of levels of shared knowledge to attempt risky coordination? In short, all of these models demonstrate that common knowledge provides the most effective and reliable path to coordination.

The problem of coordination and common knowledge has been examined by many disciplines, including political science ([Ostrom, 1990](#)), philosophy ([Hume, 1739–1740/1969](#); [Lewis, 1969](#); [Rousseau, 1754/1984](#); [Skyrms, 2004](#)), economics ([Chwe, 2001](#); [Geanakoplos, 1992](#)), linguistics ([Clark, 1992, 1996](#); [Smith, 1982](#)), sociology ([Willer, Kuwabara, & Macy, 2009](#); [Zuckerman, 2010](#)), legal theory ([McAdams & Nadler, 2005](#)), and even computer science ([Alberucci & Jäger, 2005](#); [Halpern & Moses, 1990](#)). Yet, even though common knowledge is fundamentally a psychological phenomenon, little is known about the psychology of common knowledge (some notable exceptions include [Chaudhuri, Schotter, & Sopher, 2009](#); [Clark, 1996](#); [Lee & Pinker, 2010](#); [Nov & Rafaeli, 2009](#)). We briefly review two literatures, experimental economics and theory of mind, that are indirectly relevant to the phenomenon before outlining our own research questions.

Experimental Economics: Coordination Using Salient Focal Points

A few experiments have examined whether people are better at solving coordination problems than classical game theory suggests. They focused on [Schelling's \(1960\)](#) concept of a *focal point*, an option that stands out from a set of possible choices as uniquely salient, encouraging everyone to converge upon it as a single choice. Schelling suggested that in practice people may rely on focal points to solve coordination problems because they generate common knowledge of a single solution ([Schelling, 1960](#); [Sugden, 1995](#)). [Mehta, Starmer, and Sugden \(1994a, 1994b\)](#) examined people's play in coordination games and people's ability to converge on focal points (what they called *Schelling salience*). Participants responded to questions with many possible answers (e.g., "Write down any positive number" and "Name any flower"). In one group, participants were paid to answer any way they wanted. In another, they were paid based on how well their answers matched with those of another randomly chosen participant. Participants were far more successful at coordinating answers when they were trying to do so than when they answered as they wished. This suggests that people can meet the challenge of coordination by identifying it as a problem distinct from the primary demands of a task. Though the finding, by itself, cannot distinguish whether people used shared knowledge or common knowledge to improve

their coordination, recent unpublished studies suggest that people really do use common knowledge in these tasks ([Bardsley, Mehta, Starmer, & Sugden, 2010](#); [Chartier, Abele, Stasser, & Shriver, 2012](#)).

Theory of Mind Research: Representing Shared Knowledge

Most existing research on knowledge about other people's knowledge falls in the area known as theory of mind, intuitive psychology, mind reading, or mentalizing, all terms for the mental representation of other people's mental states ([Baron-Cohen, 1995](#); [Frith & Frith, 2003](#); [Wimmer & Perner, 1983](#); for recent reviews, see [Apperly & Butterfill, 2009](#); [Saxe & Young, in press](#)). Developmental psychologists have found that by 6–7 months children are able to use implicit representations of attention, desires, goals, and intentions to guide their behavior ([Hamlin, Hallinan, & Woodward, 2008](#)). By 15 months, children implicitly differentiate their own knowledge from another person's knowledge; for example, infants are surprised when someone seeks out an object in a spot where it was moved when the person was absent ([Onishi & Baillargeon, 2005](#)). By 3–5 years, children show an ability to explicitly represent others' mental states in the false-belief task ([Callaghan et al., 2005](#); [Wellman, Cross, & Watson, 2001](#)). By 6–7 years, children are able to represent two levels of shared knowledge, as evidenced by their ability to understand that someone else can have false beliefs ([Perner & Wimmer, 1985](#)). By adulthood, people can correctly answer questions about fourth-order levels of shared knowledge (e.g., Bob knows that Carol knows that Ted knows that Alice knows X). However, they tend to fail questions about fifth-order knowledge ([Kinderman, Dunbar, & Bentall, 1998](#)), possibly because this exceeds the capacity of short-term verbal memory ([Cowan, 2000](#)).

Although people are capable of representing other people's mental states, they do not always do so effectively. Both adults and children tend to assume that their knowledge is shared by other people. This shortcoming is evident in the well-documented failure of 3-year-olds to pass a false-belief task and is also seen in adults in work on the *curse of knowledge* ([Birch & Bloom, 2003, 2007](#); [Camerer, Lowenstein, & Weber, 1989](#); [Keysar, Lin, & Barr, 2003](#)).

Because coordination depends on the ability to anticipate other people's actions, and because people's actions depend on people's mental states, one would expect mentalizing ability to facilitate coordination. Indeed, [Curry and Chesters \(2012\)](#) found that people who are better at employing theory of mind are also better at coordinating their answers with other people on questions with many possible responses. Yet characterizations of theory of mind focus on shared knowledge as the paradigm case, and shared knowledge is in general insufficient to solve coordination problems. Researchers have shown that increasing the salience of shared knowledge in cooperative and competitive environments can lead to more competitive behavior ([Epley, Caruso, & Bazerman, 2006](#); [Pierce, Kilduff, Galinsky, & Sivanathan, 2013](#)), and they have begun to map the neural correlates of representing shared knowledge ([Coricelli & Nagel, 2009](#)). Despite such progress in understanding how people represent shared knowledge, far less is known about how people represent common knowledge and use these representations in coordination.

To appreciate the distinctive role of common knowledge in coordination, recall the example of two friends, Sally and Ann, who are trying to find each other downtown. They previously discussed meeting at Starbucks or Peet's but never came to an agreement. Where should Sally go to meet Ann? Sally can represent Ann's knowledge of the two locations, her desire to meet at the same location, and vice versa. Yet, even if Ann thought it would be best to meet at Starbucks and Sally knew that Ann thought so, but Ann worried that Sally thought it would be best to meet at Peet's, Ann might go to Peet's while Sally went to Starbucks. No matter how many nested levels of shared knowledge Sally represents, she will not know where to look for Ann. Coordination games thus pose a key problem for research on theory of mind: How does one read the mind of a mind reader?

The Present Research

In these experiments we examine the cognitive processes underlying coordination. Participants interact with partners in a role-playing scenario that involves a symmetric one-shot coordination game, with payoffs that instantiate a stag hunt. In the game, participants must decide either to work alone, which offers a small but certain profit, or to try to work with a partner. The latter decision offers the potential to make more money, but only if the partner makes the same choice: If a participant chooses to work with a partner but the partner does not, they receive nothing. We test whether people differentiate between shared and common knowledge in making this decision, whether shared knowledge and common knowledge have distinct cognitive representations, and whether people use workarounds for a lack of common knowledge.

The game involves two merchants, a butcher and a baker, who decide each day whether to work independently to sell chicken wings and dinner rolls, respectively, or to work together to sell complete hot dogs, for which they earn more (see Figure 1). No one will buy just the buns or just the hot dog meat, so they risk earning nothing if they fail to coordinate their actions. Moreover, participants are told that sometimes the hot dogs can earn both of them more money than working independently, but sometimes hot dogs earn less money, so the merchants need common knowledge of higher profits to coordinate. But, their only means of communication with each other is an unreliable messenger boy. (This is a stylized instantiation of the *coordinated attack problem*; see Halpern & Moses, 1990; Rubinstein, 1989.)

To appreciate the need for common knowledge in this scenario, consider what happens on a given day that the baker sends a

message to the butcher telling him to bring hot dogs. The butcher sends a confirmation to let the baker know he received the message. The baker receives the confirmation but realizes that the butcher cannot be sure whether the messenger delivered the confirmation. So, the baker sends a confirmation of the confirmation. Upon receipt of this message, the butcher realizes that yet another confirmation is required. In fact, no finite number of successful confirmations can help the hapless merchants. They can never be sure that the most recent confirmation message was delivered by the unreliable messenger boy, and neither knows how many messages might be sufficient for the other merchant to bring his ingredient for the hot dogs (embodying the second-order coordination problem presented by multiple shared knowledge solutions). Common knowledge is therefore needed to reliably solve the merchants' problem.

To test whether people tacitly appreciate this requirement, we manipulated what they knew about their partner's knowledge about the payoffs—whether knowledge of the payoffs was private, shared, or common. The game-theoretic analysis of coordination suggests the *common knowledge recognition hypothesis*: In coordination environments, people strategically differentiate between shared knowledge and common knowledge, working together more frequently when they have common knowledge of the payoffs than when they have shared or private knowledge.

Alternatively, people may not represent common knowledge as a distinct state. The only major distinction affecting their coordination decisions would be the difference between private and shared knowledge (as suggested by the theory of mind literature). We call this the *shared knowledge hypothesis*.

Finally, the literature on the curse of knowledge raises the possibility that people do not reliably use either shared or common knowledge to solve coordination problems. If people attribute their own knowledge to other people, distinctions among levels of knowledge would be irrelevant. This *curse of knowledge hypothesis* predicts that participants will try to work together with the same frequency across all knowledge levels.

Knowledge-Level Representation

If people do distinguish between common and shared knowledge, this raises the further question of how they represent the distinction. One possibility is that these knowledge states have a single cognitive format and the distinction between them is simply quantitative, with common knowledge represented as an upper limit of shared knowledge. Alternatively, shared and common knowledge may have distinct representations, which would make the distinction qualitative and categorical.

These possibilities can potentially be distinguished by the pattern of classification errors people make when reporting their level of knowledge. Research on theory of mind capabilities suggests that shared knowledge becomes more difficult to represent as the levels of knowledge increase. If people entertain a single kind of representation (the *single-representation hypothesis*), the most errors will be observed in the common knowledge condition (the maximum number of shared knowledge levels), with fewer errors made as the levels of shared knowledge decrease. If, in contrast, common knowledge has its own representation, it need not contain multiple levels of embedded knowledge; it could consist of a single mental symbol that means "We have common knowledge."

| | | Butcher's options | |
|-----------------|---------------------------|--------------------------|----------------------------|
| | | Work Together (Hot Dogs) | Work Alone (Chicken Wings) |
| Baker's options | Work Together (Hot Dogs) | \$1.10, \$1.10 | \$0, \$1.00 |
| | Work Alone (Dinner Rolls) | \$1.00, \$0 | \$1.00, \$1.00 |

Figure 1. An interaction between a butcher and a baker in Experiment 1. The baker chooses a row and the butcher chooses a column. The four cells show the payoffs (baker's payoff, butcher's payoff) for each combination of choices. These payoffs generate a coordination game, specifically a stag hunt game (Skyrms, 2004), in which one equilibrium is better for both players than another equilibrium. The other three payoff conditions substitute \$2.00, \$5.00, or \$10.00 for the \$1.00 payoff and \$2.10, \$5.10, or \$10.10 for the \$1.10 payoff.

Thus, errors will increase only with the number of levels of shared knowledge, whereas errors on common knowledge will be few (similar to the error rate for secondary knowledge, which also requires only a single level of representation). The *distinct-representation hypothesis* makes the further prediction that errors will be systematic, respecting the boundary between the two kinds of knowledge: Different levels of shared knowledge will be mistaken for each other, but not for common knowledge.

Sensitivity to Costs and Benefits

To further characterize the decision processes behind coordination, we vary the game payoffs to test people's sensitivity to costs and benefits. Rational choice theory (e.g., Becker, 1976) models human decisions as expected utility calculations, which in coordination situations involve comparing the payoffs of different decisions scaled by the probabilities of achieving each outcome. According to this approach, people should decide to work together if they determine that the expected value—the amount earned for successful coordination multiplied by the probability that they think their partner will do the same—is greater than the amount earned for working alone. This *rational actor hypothesis*, in which the assumptions of economic modeling are directly interpreted as a psychological theory, predicts that as the benefits of working together decrease relative to the payoff for working alone, people will be less likely to try to work together and their decisions will track this ratio in the interaction.

In contrast, a more explicitly psychological approach treats game-theoretic models not as literal theories of decision making but as task analyses to help identify possible evolved information-processing mechanisms (Gigerenzer, Todd, & ABC Research Group, 1999). The probabilities associated with another individual's behavioral choices are not directly accessible to a human perceiver but must be heuristically inferred in real time from available cues. Rather than basing their decisions on potentially indeterminate probabilistic calculations of expected utility, humans may instead use simpler heuristics, which categorically distinguish common, shared, and private knowledge using ecologically typical cues to infer whether another person will attempt to coordinate. This *knowledge-level heuristic hypothesis* predicts that people's decisions will be driven primarily by categorical perceptions of knowledge states and thus may be relatively insensitive to the costs and benefits of coordination.

Other Motivations for Coordination

Coordination decisions may be influenced by factors other than public knowledge. Because coordination requires elusive knowledge that another person has made the same choice as oneself, people may use their own decision-making processes to simulate how their partner will think and behave (Gallese & Goldman, 1998), particularly when they view themselves as similar to their partner (Mitchell, Macrae, & Banaji, 2006; Tamir & Mitchell, 2013). Indeed, perceived similarity has been shown to help people solve coordination problems in which coordination requires that both make the same decision, but not when it requires that they make different decisions (Abele, Stasser, & Chartier, 2014). According to this *perceived similarity hypothesis*, the more similar an actor perceives himself or herself to be to a potential coordination

partner, the more likely the actor will be to predict that their partner will choose as they do and thus attempt risky coordination.²

People may also be motivated to coordinate by altruistic or other-regarding preferences. Altruistic and reputational motives have been well documented in social psychology and experimental economics (e.g., Haley & Fessler, 2005; Messick & McClintock, 1968; Milinski, Semmann, & Krambeck, 2002; Van Lange, 1999). The Big Five personality trait of Agreeableness is associated with altruism, prosociality, friendliness, heightened self-presentation concerns, and generosity (Goldberg, 1992; Graziano & Tobin, 2002; Roccas, Sagiv, Schwartz, & Knafo, 2002; Sun & Wu, 2012) and has specifically been associated with altruistic motivations towards nonrelatives and strangers (Graziano, Habashi, Sheese, & Tobin, 2007). The *altruistic motives hypothesis* predicts that people who are higher on Agreeableness will be more likely to try to work together.

Finally, some people may simply be willing to accept the potential cost of discoordination in the hope that they can earn more money through high-payoff coordination. The Big Five personality trait of Openness is associated with risk seeking (Nicholson, Soane, Fenton-O'Creevy, & Willman, 2005) and in particular with the seeking of chances for gains (Lauriola & Levin, 2001). The decision to work together is a social gamble where one can bet a certain payout to win an additional increment of profit. The *risk-seeking hypothesis* predicts that people who are higher on Openness will be more likely to take the bet by trying to work together.

We report four experiments designed to test these hypotheses. In Experiments 1 and 2, we test the effects of knowledge level and payoff structure on coordination decisions involving, respectively, one and three partners. In Experiment 3 we investigate how shared and common knowledge are cognitively represented and test the three social motivation hypotheses. In Experiment 4 we verify that the participants' coordination behavior generalizes to a different fictional context and thus is not an artifact of specific features of the first scenario.

Experiment 1

Experiment 1 implements the butcher-baker coordination game explained above. Each participant interacted with a partner, playing the role of either the butcher or the baker. The participants read that they could work either alone or with the partner; the amount they could earn for working alone was constant, but the amount earned for working together would vary from day to day and might be less than or greater than the amount they could make by working alone. They were then told that, on the day of the actual decision facing them, the payoff for working together was greater than the payoff for working alone. The participants were then given one of four kinds of information about what their partner

² This is related to the concept of superrationality, in which rational actors decide to cooperate in a prisoner's dilemma because each assumes that both he (or she) and the partner rationally see the wisdom of mutual cooperation and know that the other sees it, that both he and the partner know that the other knows that he knows that the other sees it, and so on (see Colman, 2003; Fischer, 2009; Hofstadter, 1985). The same logic can be applied to coordination games with symmetrical payoff structures.

knew about the payoff, which we varied in a between-subject design.

In the *private knowledge* condition, a participant was told he or she could earn 10 cents more for working with the partner but was not given information about what the partner knew. In the *secondary knowledge* condition, a participant was told that the partner also knew about this payoff. In the *tertiary knowledge* condition, a participant was told that the partner knew the payoff and knew that the participant himself or herself knew the payoff. In the *common knowledge* condition, the payoff was presented as public information, commonly known between the two participants. This information was given to participants in one of two boxes presented on the screen. Private, secondary, or tertiary knowledge information was presented in a private knowledge box that they were told only they could see, while common knowledge information was presented in a public knowledge box that both they and the partner could see.

To see whether coordination decisions were sensitive to costs and benefits, we manipulated, between participants, the amount the participants could earn by working alone and together, yielding four payoff structures (see Figure 1): \$1.00/\$1.10, \$2.00/\$2.10, \$5.00/\$5.10, and \$10.00/\$10.10 (hereafter referred to as the \$1, \$2, \$5, and \$10 payoff conditions). We note that these stakes are typical for games run on Mechanical Turk (Amir, Rand, & Gal, 2012) and that previous research has shown that, although real cash incentives are critical, the absolute size of the stakes tends not to alter the pattern of results (for a review, see Camerer, 2003 and Camerer & Hogarth, 1999). We chose small incremental payoffs for coordination (and thus relatively high opportunity costs for discoordination) to counter the typical demand characteristics of experimental games, which tend to encourage cooperative actions (Pedersen, Kurzban, & McCullough, 2013).

Method

Participants. We used Amazon Mechanical Turk to recruit 1,600 participants (100 per condition) from the United States to complete a short study for a small payment. Mechanical Turk presents special challenges to ensuring internal and external validity, because online participants may not understand or engage with a task. Data quality and validity can be enhanced with either ex ante prescreening tasks or ex post exclusion based on comprehension questions. Research shows that both methods can effectively reduce statistical noise without systematically biasing the results (Berinsky, Margolis, & Sances, & 2013; Goodman, Cryder, & Cheema, 2013; Horton, Rand, & Zeckhauser, 2011; Oppenheimer, Meyvis, & Davidenko, 2009; Rand, Greene, & Nowak, 2012; Suri & Watts, 2011). However, prescreening with comprehension checks can potentially alter participant behavior (Rand et al., 2012) and may increase self-presentation effects (Clifford & Jerit, 2012). Hence, we used ex post exclusion based on comprehension questions about the game's payoff structure (see Procedure).

The final sample consisted of 1,033 participants (58% female) with a mean age of 32.8 years ($SD = 15.0$). This 35% ex post exclusion rate is within the range of rates observed in previous research using multiple comprehension checks with Mechanical Turk samples, which can vary between 25% and 50% (Berinsky et al., 2013; Downs, Holbrook, Sheng, & Cranor, 2010; Goodman et

al., 2013; Horton et al., 2011; Rand et al., 2012). Furthermore, researchers that have compared similar procedures in the lab and on Mechanical Turk have observed equivalent exclusion rates (Oppenheimer et al., 2009; Paolacci, Chandler, & Ipeirotis, 2010; Rand et al., 2012).

Procedure. Participants read instructions explaining that they would earn a minimum of 50 cents, which they could augment based on their decisions in their interaction with another participant on Mechanical Turk.³ They were told that one of them would play a butcher and the other a baker. Each could either work alone for a sure profit (the butcher could make chicken wings, the baker dinner rolls) or attempt to work with the partner, the butcher making hot dogs, the baker the buns. By choosing to work together, they were told, the participant could earn a profit, but only if the partner also chose to work together; if either decided to collaborate but the partner did not, that person would not earn anything, because one cannot sell a bun without a hot dog or vice versa. Participants then read that they would earn a certain amount (\$1, \$2, \$5, or \$10) if they decided to work alone but that the hot dog price varied from day to day; thus, the earnings for working together might be more than or less than this sure profit. Finally, they read that the information about hot dog earnings might be conveyed to them by a messenger boy (displayed on their screen in a private box they were told only they could see) or by a loudspeaker (displayed on their screen in a public box they were told the other participant could see on his or her screen as well).

The participant then clicked a button to reveal the day's information about the price of hot dogs and hence the potential profit for collaborating; in each case it was 10 cents more than each would earn by working alone. In the second between-subjects manipulation, participants received one of the following pieces of information (presented here from the perspective of the baker):

1. *Private knowledge.* In the private box the participant read, "The Messenger Boy has not seen the Butcher today, so he cannot tell you anything about what the Butcher knows." The public box stated that the loudspeaker was silent.
2. *Secondary knowledge.* In the private box the participant read, "The Messenger Boy says he stopped by the butcher shop before coming to your bakery. He tells you that the Butcher knows what today's hot dog price is. However, he says that he forgot to mention to the Butcher that he was coming to see you, so the Butcher is not aware that you know today's hot dog price." The public box stated that the loudspeaker was silent.
3. *Tertiary knowledge.* In the private box the participant read, "The Messenger Boy mentions that he is heading over to the butcher shop, and will let the Butcher know today's price as well. The Messenger Boy will also tell the Butcher that he just came from your bakery and told you the price. However, the Messenger Boy will not inform the Butcher that he told you he would be heading over there. So, while the Butcher is aware that you know

³ A sample of the experimental stimuli can be viewed at http://www.pdescioli.com/CK_Materials/CK_Materials.html

today's price, he is not aware that you know that he knows that." The public box stated that the loudspeaker was silent.

4. *Common knowledge.* In the public box the participant read, "The loudspeaker broadcast the market price of [today's price] (of which you could earn [earnings for working together])." In the private box the participant read, "The messenger boy did not come by. Because the market price was broadcast on the loudspeaker, the Butcher knows [today's price], and he knows that you know this information as well."

The participants then made a decision to work alone or with the partner and indicated the choice with the keyboard. They were then asked to explain how they made the decision and were given two sets of comprehension questions. The first set was used to (ex post) exclude participants who did not understand the game's payoff structure. It contained three questions about the profits under various combinations of decisions:

1. "If you and the Butcher both chose to make hot dogs, then how much money would you earn for the interaction?"
2. "If you chose to make hot dogs but the Butcher did not, then how much money would you earn for the interaction?"
3. "If you chose not to make hot dogs, regardless of what the Butcher decided, then how much money would you earn for the interaction?"

The second set was used as a manipulation check. It contained four questions about what the participants and partners knew:

1. "Do you know the price of hot dogs today?"
2. "Does the Butcher know the price of hot dogs today?"
3. "Does the Butcher know that you know the price of hot dogs today?"
4. "Does the Butcher know that you know that he knows the price of hot dogs today?"

Finally, participants filled out a brief demographic questionnaire, submitted the task, and received the base rate payment for completion. We randomly paired participants offline to implement the conditions described in the scenario. To calculate each individual's payoffs, we matched private-level participants with private-level participants and matched common-level participants with common-level participants; secondary-level participants were matched with private-level participants, and tertiary-level participants were matched with secondary-level participants. (Participants were paid based on their pairing with lower level knowledge partners and not on pairings with higher level knowledge partners.) This matching procedure is necessary to avoid deception, because the shared knowledge instructions specify a partner at a lower knowledge level (which is required to establish secondary and

tertiary knowledge). Participants were paid their game earnings via Mechanical Turk's "Bonus Payment" feature.

Results and Discussion

Figure 2 shows that, with all four payoffs, the percentage of participants who tried to work together was significantly affected by their state of knowledge (see also the first row of Table 1). Planned comparisons across adjacent knowledge conditions (i.e., private–secondary, secondary–tertiary, and tertiary–common) are shown in Table 1 for all payoff conditions. In all four payoff conditions, more participants tried to work together with common knowledge than with tertiary knowledge. In three out of four payoff conditions, more participants tried to work together with secondary knowledge than with private knowledge (the difference was only marginally significant in the \$5 payoff condition). Coordination rates were the same with secondary and tertiary knowledge, except with the \$5 payoff, for which the rate with secondary knowledge was anomalously low.

These results are consistent with the common knowledge recognition hypothesis: Participants were more likely to try to work together with common knowledge than with any other state of knowledge. The results were inconsistent with a strong curse of knowledge hypothesis, because the likelihood of working together differed across knowledge conditions. In line with the shared knowledge hypothesis, few of the participants tried to work together with private knowledge and more tried to work together with secondary knowledge. However, only slightly more participants tried to work together with tertiary knowledge, and far more participants tried to work together with common knowledge (statistics presented in Table 1). The pattern is consistent with the hypothesis that people maintain a dual representation in which shared and common knowledge are thought of as qualitatively distinct.

These results were inconsistent with a strict rational actor hypothesis because the proportion of participants who decided to try to work together in each knowledge condition varied little across the payoff conditions, even as the ratio between the cost of the forgone profit from working alone to the additional benefit from working together increased tenfold: private, $\chi^2(3, N = 261) = 3.13, p = .373, \phi = .11$; secondary, $\chi^2(3, N = 259) = 7.66, p = .054, \phi = .17$; tertiary, $\chi^2(3, N = 260) = 6.81, p = .078, \phi = .16$; common, $\chi^2(3, N = 253) = 0.76, p = .859, \phi = .05$. A rational actor would expect that as this ratio increases, the other rational actor would take the increased opportunity cost into account and the probability that they would try to work together should correspondingly decrease, creating a positive feedback loop that would drive each of them to work alone. The fact that the proportion of people who tried to coordinate with common knowledge was invariant across payoffs contradicts the idea that coordination decisions were based on maximizing the expected payoff, and is instead consistent with the knowledge-level heuristic hypothesis.

Another test of the rational actor hypothesis may be obtained by examining the actual payouts that the participants would earn given their collective pattern of choices. Inspection of the frequency of coordination attempts in the secondary and tertiary knowledge conditions reveals that this payoff is likely to be low: Participants who decided to try to work together generally failed to coordinate with their partners (and thus relinquished their sure

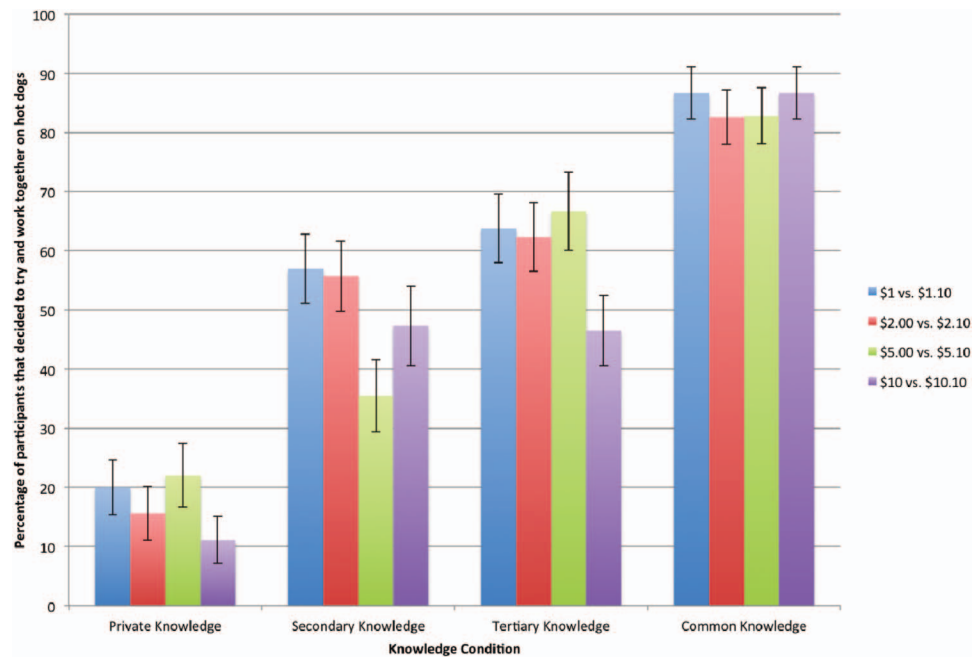


Figure 2. Percentage of participants who tried to work together in Experiment 1, organized by knowledge condition and payoff condition. Error bars represent standard error. See the online article for the color version of this figure.

profit from working alone). To assess the overall rationality of these choices, we calculated *expected earnings* based on all possible matchups with the other participants (rather than the actual earnings from the matchups we arbitrarily arranged in order to calculate their payments). This consists of the sum of the proportion of participants who chose to work alone multiplied by the smaller certain payoff, and the proportion of participants who would, on average, successfully work with a partner (when both they and the partner chose to coordinate, which is the product of the proportion of participants who tried to work together and the proportion of potential partners who made the same choice), multiplied by the higher, risky payoff. The discoordination payoff, from cases in which they would choose to cooperate but their partner would not, was zero, thus eliminating this term from the calculation. Figure 3a shows that for all payoffs, efficiency was higher with private and common knowledge than with either level of shared knowledge.

In sum, Experiment 1 shows that when people make coordination decisions, they differentiate between private, shared, and

common knowledge (though apparently not among different levels of shared knowledge). Moreover, the level of knowledge and the special appeal of common knowledge are far more salient to them than the expected value of the options: Increasing the cost–benefit ratio tenfold had no observable impact on their choices.

Experiment 2

How general are the phenomena of sensitivity to knowledge and insensitivity to payoffs observed in Experiment 1? Presumably, if achieving coordination is difficult enough and the stakes are high enough, then even with common knowledge people would opt to work alone; as an extreme example, imagine risking a sure payoff of \$1,000 for working alone for a chance at earning \$1,001 by coordinating with a million partners. To test the limits of common knowledge as a qualitative coordination heuristic, we designed Experiment 2 as a four-person coordination game in which all four partners had to decide to work together to achieve the benefits of coordination. Coordination on the higher paying option of working

Table 1
Comparison of Knowledge Levels in Each Payoff Condition, Experiment 1

| Knowledge level | \$1 payoff | | | \$2 payoff | | | \$5 payoff | | | \$10 payoff | | |
|-------------------------------|------------|----------|--------|------------|----------|--------|------------|----------|--------|-------------|----------|--------|
| | χ^2 | <i>n</i> | ϕ | χ^2 | <i>n</i> | ϕ | χ^2 | <i>n</i> | ϕ | χ^2 | <i>n</i> | ϕ |
| All levels | 63.70*** | 276 | .48 | 62.79*** | 272 | .48 | 56.71*** | 236 | .49 | 70.41*** | 249 | .53 |
| Private vs. secondary | 21.26*** | 147 | .38 | 23.27*** | 134 | .42 | 2.66 | 121 | .15 | 19.06*** | 118 | .40 |
| Secondary vs. tertiary | 0.69 | 141 | .07 | 0.63 | 139 | .07 | 10.88*** | 113 | .31 | 0.01 | 126 | .01 |
| Tertiary vs. common knowledge | 8.84** | 129 | .26 | 7.12** | 138 | .26 | 4.02* | 115 | .19 | 23.05*** | 131 | .42 |

* $p < .05$. ** $p < .01$. *** $p < .001$.

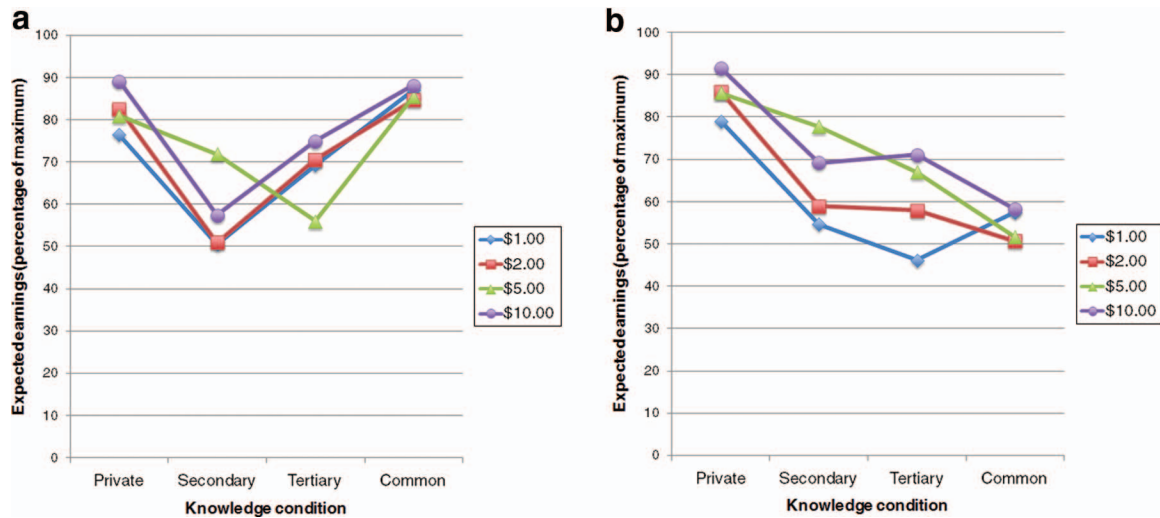


Figure 3. Average expected earnings as a percentage of maximum possible earnings for Experiment 1 (a) and Experiment 2 (b) by knowledge condition. We calculated *expected earnings* as the average amount a participant would earn across all possible pairings with the other participants. See the online article for the color version of this figure.

together is far more difficult with four people, because the probability of success is equal to the probability that any one partner decides to work together cubed.

In fact, the perceived probability of successful coordination may fall even faster than that. In addition to requiring common knowledge of the payoffs, coordination requires individuals to be confident in their partners' rationality. An irrational partner could prefer lower payoffs, choose blindly, or make some other unpredictable choice. With only two players, the chance of an irrational partner might be negligible, but this risk can be greater in larger groups. Because even a single irrational partner can be enough to torpedo coordination in a group, as the number of players goes up, the likelihood of discoordination increases rapidly (everyone must be both knowledgeable and rational and believe everyone else is as well). For this reason the cost-benefit structure may become more salient to a participant as the number of other partners increases. Recall that a rational actor may choose to coordinate with less-than-perfect common knowledge (i.e., with common *p*-belief) as long as the probability of the other's belief exceeds a critical value that depends on the relative payoffs: The higher the opportunity cost, the higher that probability must be. Thus we may see a greater sensitivity to payoffs in a coordination game involving more people.

Method

Participants. As in Experiment 1, 1,600 participants were recruited from Amazon Mechanical Turk, evenly distributed across the 16 combinations of four payoff and four knowledge conditions. After we excluded participants who did not understand the payoffs, the sample consisted of 1,150 participants (48% female, $M_{\text{age}} = 31.9$, $SD_{\text{age}} = 11.1$).

Design and procedure. Participants were told they could work together to make "superburgers," which require a burger, a bun, cheese, and toppings from, respectively, a butcher, a baker, a

cheese maker, and a produce vender. One participant was assigned to each of these four roles. As in Experiment 1, each participant also had the option to make a food item on his or her own for a sure profit. Participants were told that they would receive a profit for contributing to superburgers only if all three of the other merchants made the same choice and that they would receive nothing otherwise. In each of the four knowledge conditions the participant's three partners were said to have the same level of knowledge. This was conveyed with instructions identical to those of Experiment 1, except that "the Butcher" or "the Baker" was replaced with "the other merchants" (so all other merchants were said to have the same level of knowledge). All other aspects of the procedure were the same as in Experiment 1.

Results and Discussion

As in Experiment 1, participants' state of knowledge affected participants' decision to work together (see Figure 4). Table 2 shows that in all payoff conditions, significantly more participants tried to work together with common knowledge than with tertiary knowledge, and in three of the four payoff conditions, significantly more tried to work together with secondary than with private knowledge. In none of the payoff conditions was there a significant difference between secondary and tertiary knowledge. This consistent lack of significant differences between the shared knowledge conditions suggests that people treat secondary and tertiary knowledge similarly.

Unlike Experiment 1, people showed some sensitivity to the payoff structure. In all knowledge conditions, increasing the relative costs of coordination failures brought down coordination rates. This is consistent with the observation that the minimum level of confidence in common knowledge (i.e., the minimum common *p*-belief) required for rational coordination rises more steeply with opportunity costs when the number of players (and hence the

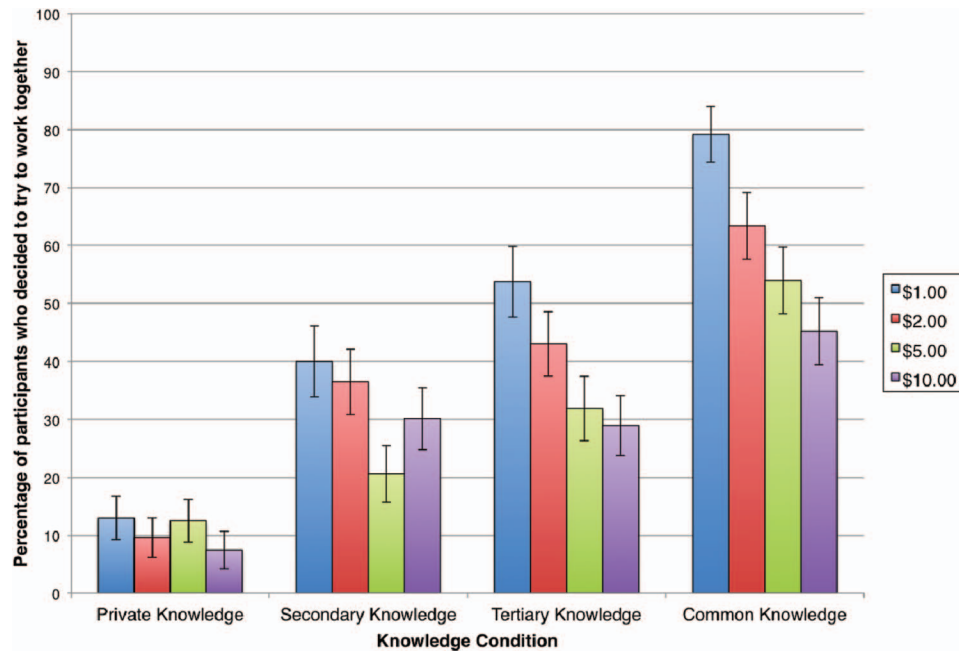


Figure 4. Percentage of participants who tried to work together in Experiment 2, organized by knowledge condition and payoff condition. Error bars represent standard error. See the online article for the color version of this figure.

chance that at least one will be ignorant or irrational or both) increases.

Though participants' sensitivity to payoffs was more consistent with the rational actor hypothesis than in Experiment 1, another aspect of the results was not. Unlike what we obtained in Experiment 1, the most profitable knowledge condition with all four payoffs was private knowledge (see Figure 3b); common knowledge was less profitable than shared knowledge except with the least costly forgone payoff of \$1. This shows an important limit to the advantages that people can obtain from common knowledge. When either the knowledge state or the rationality of all the necessary potential partners is less than perfect, coordination is difficult to achieve and hence poses a high risk of failure. In those cases even high rates of decisions to coordinate may not be enough to consummate successful coordination, and the temptation to coordinate presented by common knowledge can actually reduce the coordinators' payoff. Yet, more than half of the participants provided with common knowledge still opted for the risky higher payoff.

Experiment 3

Experiment 3 is a replication of one of the payoff conditions from Experiment 1 with additional components that allowed us to test how shared and common knowledge are represented and why people sometimes make what appears to be an irrational decision to cooperate with just shared knowledge.

At least since Miller and Nicely (1955), cognitive psychologists have used confusion matrices to test hypotheses about underlying mental representations, based on the assumption that confusable stimuli are likely to be represented similarly. A similar logic underlies the memory confusion paradigms commonly used in social psychology to reveal the dimensions of social categorization, such as the "Who said what?" paradigm (e.g., Klauer & Wegener, 1998; Lieberman, Oum, & Kurzban, 2008; Taylor, Fiske, Etcoff, & Ruderman, 1978). In our case, we used errors in responses to our questions about participants' comprehension about the level of knowledge as evidence of whether shared and common knowledge are represented in the same or in qualitatively

Table 2
Comparison of Knowledge Levels in Each Payoff Condition, Experiment 2

| Knowledge level | \$1 payoff | | | \$2 payoff | | | \$5 payoff | | | \$10 payoff | | |
|-------------------------------|------------|----------|--------|------------|----------|--------|------------|----------|--------|-------------|----------|--------|
| | χ^2 | <i>n</i> | ϕ | χ^2 | <i>n</i> | ϕ | χ^2 | <i>n</i> | ϕ | χ^2 | <i>n</i> | ϕ |
| All levels | 68.24*** | 281 | .49 | 45.32*** | 297 | .39 | 46.36*** | 283 | .41 | 24.72*** | 289 | .29 |
| Private vs. secondary | 13.59*** | 142 | .31 | 14.95*** | 147 | .32 | 1.77 | 148 | .11 | 11.54*** | 140 | .29 |
| Secondary vs. tertiary | 2.50 | 132 | .11 | 0.69 | 153 | .07 | 2.26 | 137 | .13 | 0.03 | 149 | .01 |
| Tertiary vs. common knowledge | 10.14** | 139 | .27 | 6.21* | 150 | .20 | 12.39*** | 135 | .30 | 4.23* | 149 | .17 |

* $p < .05$. ** $p < .01$. *** $p < .001$.

distinct ways. Unfortunately, in the first two experiments these questions were so easy that almost all participants got them all correct. In this experiment, we made the questions more difficult in three ways: by putting them at the end of the survey, by adding a task before participants answered them, and by concealing the relevant information while they answered the questions (in the first two experiments, this information was visible on the screen).

This experiment tests several additional hypotheses. Recall from the introduction that there are several reasons that people may choose to coordinate even in the absence of common knowledge. One is that an actor may ascertain that she has similar values and biases to a potential partner and thus that the partner is likely to assess the situation in the same way that she does, including an assessment of whether she herself is likely to choose to coordinate with the partner. We thus manipulated whether the participants thought they were interacting with a partner who was similar or dissimilar to themselves in age, political orientation, tastes in music, and decision-making style. The other social motivations for coordination consist of personality traits that make the choice inherently appealing, including Agreeableness, which impels people to act in a prosocial manner, and Openness, whose risk-seeking component may impel people to gamble for a big payoff rather than accept a smaller but surer payoff.

Method

Participants. We recruited 800 participants from Mechanical Turk, evenly distributed across similarity and knowledge conditions. After eliminating people who failed the comprehension questions, we were left with 550 participants in the final analyses (approximately 46% male, $M_{\text{age}} = 31.6$, $SD_{\text{age}} = 11.3$).

Design and procedure. The design added three components to the \$2 condition from Experiment 1.

Similarity manipulation. At the beginning of the experiment, participants answered four questions:

- “Do you prefer more intense kinds of music (e.g., rock or rap) or more mellow kinds of music (e.g., classical or jazz)?”
- “If you had to pick, would you say you are more liberal or more conservative?”
- “How old are you?” [available answers: “I’m 35 years old or older” and “I’m younger than 35 years old”]
- “When making decisions do you tend to rely more on intuition or more on reason?”

In the Similar condition, participants were told that they would be matched with a partner who gave the same answers to three or more of these questions. In the Dissimilar condition, participants are told that they would be matched with a partner who gave the same answers to two or fewer of these questions. Participants were then asked to report how similar they perceived their partner to be to them, on a scale from 0% to 100%.

Big Five personality questionnaire. After participants read the role-playing scenario and made their decision, they were asked to fill out a standard 50-question survey that measured the Big Five personality traits (Goldberg, 1999).

Knowledge-level comprehension questions. Finally, the knowledge-level comprehension questions were administered on a separate page; when answering them, participants were unable to refer back to the initial instructions.

Results and Discussion

Figure 5 shows that the similarity manipulation made no systematic difference. Although ratings of perceived similarity were higher in the Similar condition, $t(548) = 13.87$, $p < .001$, these perceptions of similarity had no effect on the participants’ decisions, Wald $\chi^2(1, N = 550) = 0.01$, $p = .944$. We thus collapse across similarity in all other analyses.

Knowledge level had the same effect as in the first two experiments: More people tried to work together with common knowledge than with tertiary knowledge, $\chi^2(1, N = 270) = 22.28$, $p < .001$, $\phi = .29$, and more tried to work together with secondary than with private knowledge, $\chi^2(1, N = 280) = 16.87$, $p < .001$, $\phi = .25$, but there was no difference between secondary and tertiary knowledge, $\chi^2(1, N = 284) = 0.72$, $p = .397$, $\phi = .05$.

Representations of shared and common knowledge. The confusion matrix for the questions about levels of knowledge is shown in Table 3. Participants made significantly more errors with tertiary knowledge than with any of other level of knowledge: planned comparisons, private–tertiary, $\chi^2(1, N = 266) = 33.76$, $p < .001$, $\phi = .36$; secondary–tertiary, $\chi^2(1, N = 284) = 27.91$, $p < .001$, $\phi = .31$; common–tertiary, $\chi^2(1, N = 270) = 13.93$, $p < .001$, $\phi = .23$. These errors consisted overwhelmingly of misremembering it as secondary knowledge (an error made by 23% of the participants in this condition). Error rates with common knowledge and with secondary knowledge were not significantly different, $\chi^2(1, N = 284) = 2.43$, $p = .119$, $\phi = .09$. None of the other off-diagonal confusions was as high as the one for mistaking tertiary for secondary knowledge. The next highest was 4% (mistaking tertiary knowledge for common knowledge), which was significantly different from the 23% rate for mistaking tertiary for secondary knowledge ($p < .001$).

These results show that higher levels of knowledge are increasingly difficult to represent, as suggested by the theory of mind literature, but only when the knowledge is merely shared; the highest level of all, common knowledge, is almost as easy to represent as the lowest level of shared knowledge. The confusion matrix thus suggests that shared and common knowledge have distinct cognitive representations but that quantitatively different levels of shared knowledge do not.

Altruistic motives. Figure 6 shows that participants in the shared knowledge conditions who tried to work together scored higher in Agreeableness than those who decided to work alone, a difference not observed in the private or common knowledge conditions. Logistic regression, controlling for the main effect of knowledge condition, revealed a significant Knowledge \times Agreeableness interaction in coordination attempts, Wald $\chi^2(3, N = 550) = 8.18$, $p = .043$. Post hoc t tests with Agreeableness as the dependent variable and decision (work alone vs. together) as an independent variable confirmed that the people who decided to work together with secondary and tertiary knowledge were significantly more agreeable, $t(147) = 2.25$, $p = .013$, and $t(133) = 1.89$, $p = .030$, respectively), but people who decided to work together with private or common knowledge were not ($p > .60$). These results are consistent with the hypothesis that with shared knowledge, people may choose to coordinate with others out of a sense of altruism (perhaps as a signal to encourage coordination in possible future opportunities). In contrast, Agreeableness did not affect behavior in the common knowledge condition, suggesting

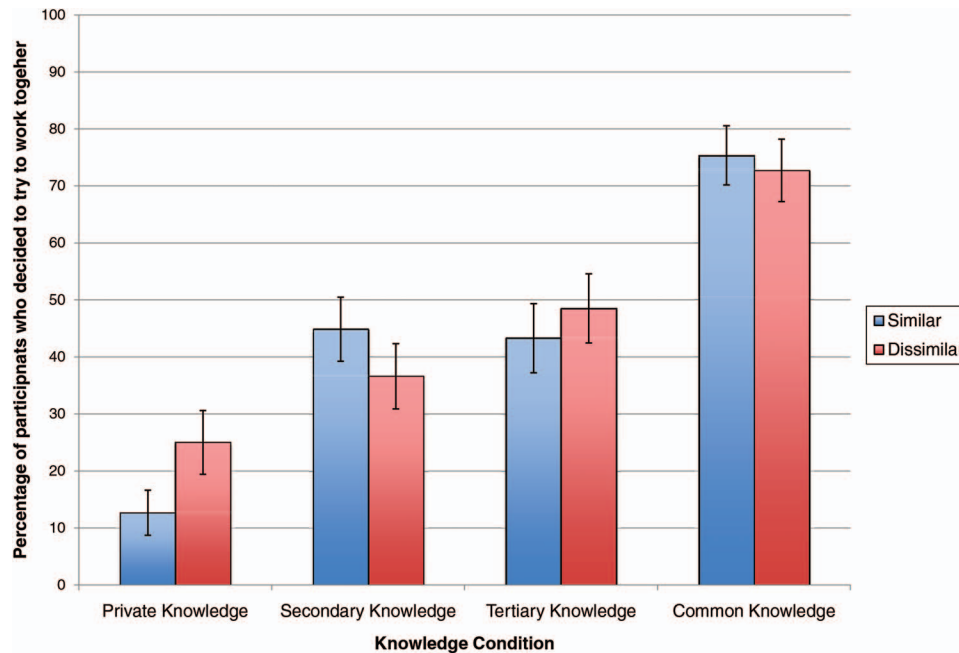


Figure 5. Percentage of participants who tried to work together in Experiment 3, organized by knowledge condition and similarity/dissimilarity condition. Error bars represent standard error. See the online article for the color version of this figure.

that the influence of common knowledge is not driven by reputational concerns.

Risk-seeking. Figure 7 shows a similar pattern for the trait of Openness to Experience, Wald $\chi^2(3, N = 550) = 13.64, p = .003$. Coordinators were more open than noncoordinators when they made their decision with secondary knowledge, $t(147) = 3.30, p < .001$, and when they made it with tertiary knowledge, $t(133) = 1.82, p = .036$, but not when they made it with private or common knowledge ($p > .80$). These results are consistent with the hypothesis that people recognize that (as game theory predicts) attempting to coordinate with shared knowledge is risky but attempting to

coordinate with common knowledge is not; those who seek risks for gains may thus gamble in conditions of shared knowledge.

No differences were found for the other three personality factors (Extraversion, Conscientiousness, or Neuroticism; all $ps > .2$).

Experiment 4

This experiment was designed to explore how the context in which a coordination problem is framed might affect rates of coordination, and to assess participants' expectations of the partner's decision. Participants engaged in a two-person interaction

Table 3
Proportion of Participants Reporting Different Levels of Knowledge in Each Condition in Experiment 3

| Condition | Reported level of knowledge | | | | | Proportion correct |
|-----------|-----------------------------|-------------------|----------|--------|-----------------------------|--------------------|
| | Private | Secondary | Tertiary | Common | Unclassifiable ^a | |
| Private | .931 | .008 | .008 | .008 | .046 | .931 |
| Secondary | .020 | .899 | .013 | .007 | .060 | .899 |
| Tertiary | 0 | .230 ^b | .637 | .044 | .089 | .637 |
| Common | .007 | 0 | .015 | .837 | .141 | .837 |

Note. Participants' perceived knowledge level by (actual) knowledge condition. Participants' perceived knowledge level was assessed with comprehension questions. Accurate judgments are those on the diagonal and are given again in the last column.

^a Unclassifiable errors correspond to patterns of errors that were logically inconsistent (e.g., reporting that they had tertiary knowledge but not private knowledge) or incomplete or those in which participants reported the correct level of knowledge but chose "can't tell" rather than "yes" for some level of knowledge that they did have. ^b According to a sign test, tertiary knowledge was mistaken for secondary knowledge more frequently than for common knowledge ($p < .001$).

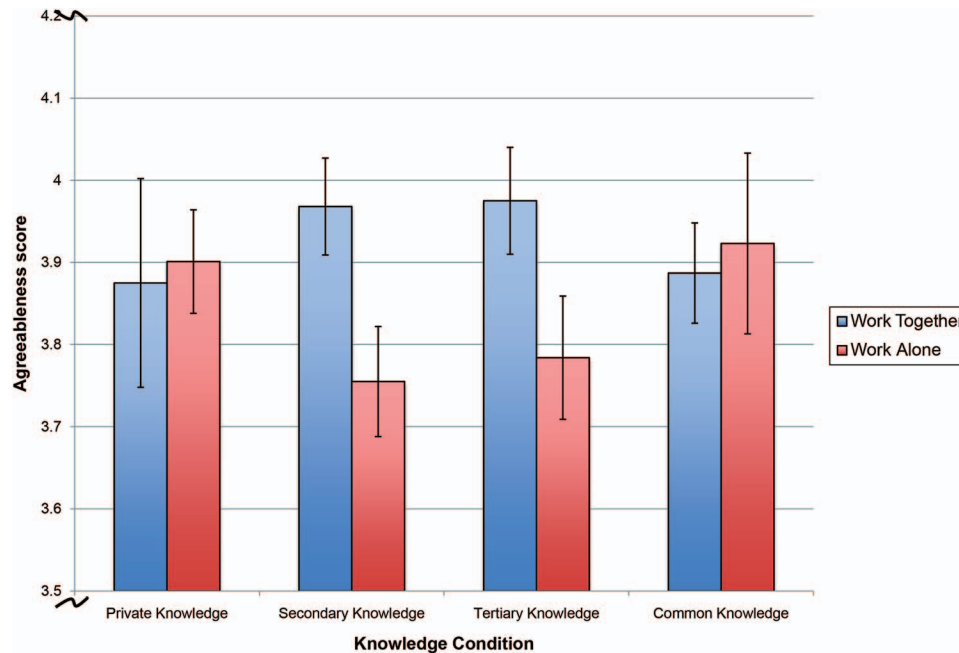


Figure 6. Average Agreeableness scale score for participants who tried to work together versus participants who decided to work alone by knowledge condition in Experiment 3. The figure shows the abbreviated range of 3.5–4.2 (full range is 1–5). Error bars represent standard error. See the online article for the color version of this figure.

that was isomorphic to the \$1 payoff condition from Experiment 1, but the scenario literally described a stag hunt involving coordination between hunters.

In this scenario one person was assigned to be an archer and the other a tracker. They each read that they could always hunt rabbits on their own for a certain payoff of \$1 or sometimes hunt deer together to earn \$1.10. As in the butcher-baker scenario, they were told they could receive private information from messengers or a public signal, any of which indicated when they could earn more by working together. Replacing the loudspeaker we used in the butcher-baker scenario was a signal fire that could be lit on a hill they could both see and that indicated the presence of deer. Information about participants' level of knowledge was given to them as part of the story rather than in the private and public knowledge boxes that were used in the previous three experiments.

After they had made their decision, participants were asked to report what they believed their partner decided to do, in order to test whether participants' decisions were contingent on such guesses, as the logic of a stag hunt game suggests is rational. All other aspects of the experiment were the same as in the \$1 condition of Experiment 1.

Method

Participants. Four hundred participants from the United States were recruited from Amazon Mechanical Turk (100 per knowledge condition) to complete a short study for a small payment. After exclusion of participants who missed comprehension questions about the game's payoff structure, the final sample consisted of 348 participants (40% female, $M_{\text{age}} = 34.8$, $SD_{\text{age}} = 11.8$).

Design and procedure. One participant was assigned to the role of archer and the other to the role of tracker. Participants were told they could either hunt rabbits on their own for a certain payoff of \$1 or try to hunt deer together for a payoff of \$1.10, which they would receive only if their partner made the same choice (these are the same payoffs presented in Figure 1). In the private, secondary, and tertiary knowledge conditions, a messenger delivered the information about the deer being present in the neighboring valley. In the common knowledge condition, participants read (from the archer's perspective), "You see smoke from the fire pit on the hill that both you and the Tracker can see, which signals to both of you that there are deer in the neighboring valley (so you could potentially earn \$1.10 if both you and the Tracker hunt deer). So, the Tracker knows that there are deer today, and knows that you know that there are deer today. The Tracker also knows that you know that he knows that there are deer today, and vice versa."

All subsequent procedures were the same as Experiment 1 (with appropriate rewording), with the addition of a question asking them what they thought their partner decided to do, presented after they made their decision.

Results and Discussion

As in the previous three experiments, more participants decided to try to work together in the common knowledge condition (60%) than in the secondary (33%) or tertiary (29%) knowledge conditions, $\chi^2(1, N = 173) = 12.72$, $p < .001$, $\phi = .27$; $\chi^2(1, N = 170) = 16.09$, $p < .001$, $\phi = .31$, respectively. Fewer participants decided to try to work together in the private knowledge condition (16%) than in the secondary knowledge condition, $\chi^2(1, N = 178) = 7.35$, $p = .007$, $\phi = .20$, and there was no significant

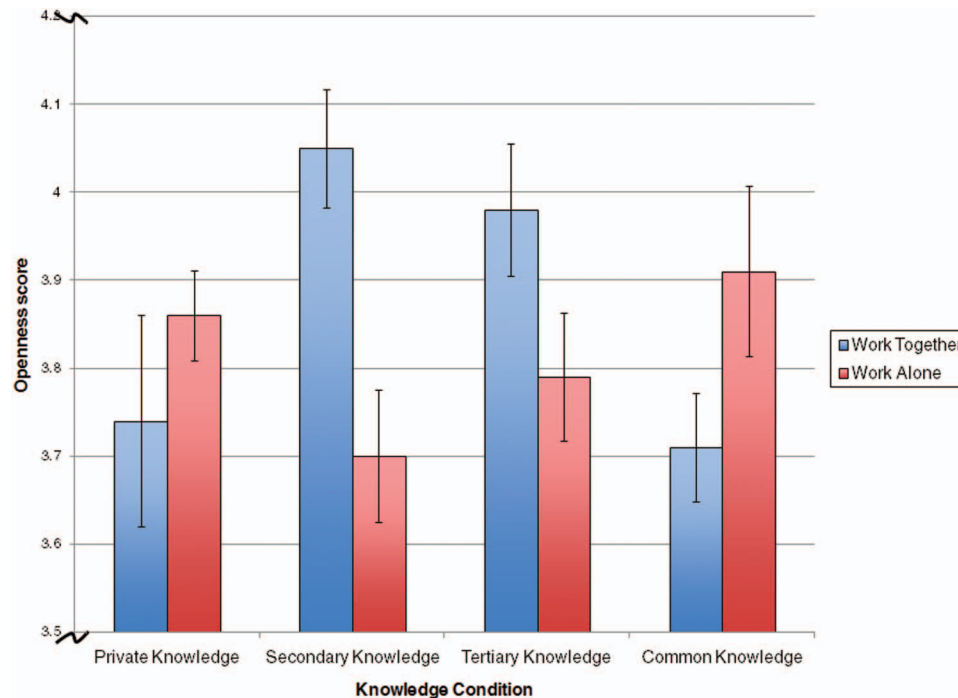


Figure 7. Average Openness scale score for participants who tried to work together versus participants who decided to work alone by knowledge condition in Experiment 3. The figure shows the abbreviated range of 3.5–4.2 (full range is 1–5). Error bars represent standard error. See the online article for the color version of this figure.

difference between the number of participants who tried to work together with secondary knowledge than with tertiary knowledge, $\chi^2(1, N = 173) = 0.253, p = .615, \phi = .04$.

The logic of a stag hunt dictates that one should do what one expects one's partner to do (a rule that we assumed participants obeyed in Experiments 1–3 but without direct evidence). Indeed, the correlation between participants' decisions and what they reported they thought their partner would do was $r(346) = .80, p < .001$. This is perhaps unsurprising, given that a number of previous studies have shown that Mechanical Turk participants believe that they are interacting with a partner as much as participants believe this in the lab, and many classic experiments that critically depend on participant interaction have been replicated on Mechanical Turk (see Amir et al., 2012; Horton et al., 2011; Rand et al., 2012; Summerville & Chartier, 2013; Suri & Watts, 2011).

In sum, the effect of knowledge condition observed in Experiments 1–3 generalized to a different fictional context and, fittingly, the one in which the logic of coordination was historically first elucidated. Moreover, we now have evidence that the key factor predicted to mediate between the knowledge level cued by the environment and people's decision to coordinate—namely, the guess that their coordination partner intends to coordinate as well—indeed is critical to that decision.

General Discussion

Humans have lived in large groups throughout their evolutionary history, providing many opportunities for mutually beneficial coordination. Game-theoretic models show that common knowl-

edge has a privileged role in helping individuals solve coordination problems. Taken together, these observations suggest that humans evolved cognitive mechanisms for recognizing common knowledge and distinguishing it from shared knowledge. Our results support this hypothesis. In all four experiments and with every combination of payoffs, participants were more likely to attempt risky coordination with common knowledge than with shared knowledge. Moreover, coordination attempts with common knowledge did not closely track the cost–benefit ratio across payoff conditions, indicating that behavior was driven less by estimated probabilities of coordination than by a categorical recognition of a state of common knowledge.

In contrast to the marked distinctions participants made between common, shared, and private knowledge, they made little distinction between different levels of shared knowledge. This is notable because the levels of shared knowledge tested here, secondary and tertiary, span almost the entire range of shared knowledge that people can readily represent, falling just one level short of the four-level maximum observed under favorable conditions by Kinderman et al. (1998). The similar rates of coordination attempts with different levels of shared knowledge are echoed by the pattern of confusions in knowing which state of knowledge was present: People confused tertiary with secondary knowledge, but they rarely confused common knowledge with shared knowledge or with private knowledge, nor shared knowledge with private knowledge. This is further reinforced by the finding that people who were higher on the personality traits of Agreeableness or Openness were more likely to attempt risky coordination in both shared

knowledge conditions and only in these shared knowledge conditions.

The cognitive difference between private and shared knowledge has long been established in the literature on theory of mind (Apperly & Butterfill, 2009; Saxe & Young, *in press*). The present results suggest that there is also a fundamental difference between shared knowledge and common knowledge. In contrast to their sensitivity to their partner's knowledge state, participants were largely insensitive to the game's payoff structure, responding to changes in expected returns only in an extreme case when successful coordination required three other people to make the same choice.

A puzzle in the results was why a fair number of participants chose to coordinate with shared knowledge. Recall that a participant with secondary knowledge was always matched with a partner with only private knowledge (they were told that their partner knew the profit for working together, but that their partner was unaware that they also knew this profit). Hence, to coordinate with their partner, a participant should act in the same way that he or she would act with private knowledge. But overall, our participants defied this logic, with more of them opting to work together with secondary knowledge than with private knowledge, providing further support for the claim that participants used the different knowledge states heuristically and did not rationally calculate their best move.

If they were not playing their best move in cost–benefit terms, why would people risk a certain payout for an uncertain small gain? The personality results indicate that participants may have chosen to try to work together to signal their cooperative character (Agreeableness) or because of a risk-seeking disposition (Openness). Participants did not leverage similarity in making their decisions, but the possibility remains that this manipulation was too cursory to be effective. The rewards for working alone and working together were abstract monetary payoffs, which differ only in magnitude and not in semantic content. However, in many real-world situations the nature of the similarity between people is highly relevant to the kind of coordination they are considering, and perceived similarity has been shown to facilitate coordination when coordinating requires matching on some kind of semantic content (Abele et al., 2014). Two roommates, for example, who will have to endure or enjoy each other's music, ought to find similarity in musical tastes more relevant to the decision to live together than similarity in personality or politics.

The pattern of participant payouts provides clues as to why certain aspects of participants' behavior in these experiments were suboptimal. In an arbitrary social situation (such as the artificial scenario of interacting with strangers on the Internet in a contrived game), it is extraordinarily hard to predict whether coordination will be profitable, because it depends critically on small and unpredictable differences in the decisions of the other participants. In the case of the four-person game above (Experiment 3), coordination turned out to be the least profitable strategy, even when a majority of participants chose to coordinate, because even a relatively small number of noncoordinators was enough to scuttle coordination and its rewards. With people unable to predict at exactly which combination of probabilities and payoffs in a given situation this tipping point lies, they may focus predominantly on information that indicates other people's state of knowledge. Actors coordinate when they have evidence for common knowledge

and refrain from coordinating when they do not. For this heuristic to be advantageous in real life, people must have high-quality information about common knowledge in ecologically typical environments and about other people's sensitivity to the same information.

Broader Implications: Common Knowledge in Social Life

The finding that people use common knowledge in their decisions to coordinate their behavior, the evidence that common knowledge is a distinct cognitive category, and the suggestion that everyday social life provides reliable cues to common knowledge in opportunities for coordination all imply that common knowledge has a strong presence in human life and in the phenomena studied by social psychology. This makes it surprising that the psychology of common knowledge has apparently had so little visibility in psychology and raises the question of whether it is similarly invisible in everyday life, which would be puzzling. If coordination is as important to social life as altruism, and if common knowledge is as indispensable to coordination as reciprocity is to altruism, shouldn't we expect our language and our lives to be permeated with ideas of common knowledge?

We suggest that this is indeed the case, even if it has not been fully appreciated. Just as the logic of reciprocity makes us obsessed with concerns such as debt, favor, bargain, and obligation, we suggest that the logic of common knowledge makes us obsessed with concerns such as publicity, privacy, confidentiality, conventional wisdom, fame, celebrity, hypocrisy, taboo, tact, euphemism, piety, mock outrage, political correctness, and "Washington gaffes" (when a politician says something that is true). In other words, both psychology and everyday social life have been concerned with the manifestations of common knowledge, even if psychologists have not hitherto treated them as exemplars of a single principle.

We conclude by suggesting that an acknowledgment of the role of common knowledge in enabling coordination can unify and explain a variety of seemingly unrelated and puzzling phenomena. In particular, much of social life is affected by common-knowledge generators, and much of language and cognition is sensitive to the state of common knowledge.

The most obvious common-knowledge generator is *direct speech*. When one person says something to another in "plain language" or "in so many words," the content of the proposition is common knowledge. Lee and Pinker (2010) showed that when an experimental participant read a vignette in which one person issues an overt threat, bribe, or sexual come-on to another, the participant assumed that each party knows that the other knows that he knows (etc.) the relevant intention, whereas when the same proposition is proffered in an innuendo, even an obvious one, the participant assumes only that the parties know the content of the proposition as private knowledge (e.g., "Michael offered a bribe") and not at higher order levels (e.g., "Michael knows that the officer knows that he offered a bribe").

The generation of common knowledge may be the function of other deliberate and salient communicative acts. One such example is *joint attention* (Scaife & Bruner, 1975; Tomasello, 1995), in which two people look back and forth at an object and at each other. Joint attention is thought to facilitate the acquisition of

words, a classic example of a coordination equilibrium (see Lewis, 1969), and children seem to assume that the meaning of a novel word is common knowledge, but they do not make this assumption about novel facts (Diesendruck & Markson, 2001).

Another class of common-knowledge generators consists of *performatives* (Austin, 1962; Searle, 1989) and the associated phenomenon of *public ceremonies*, in which the public utterance of a proposition (e.g., “I now pronounce you man and wife”) ratifies a new coordination equilibrium such as a marriage, law, or court decision. Much of our moral psychology, including moral debate and condemnation, generates common knowledge of prohibited actions, which allows people to coordinate aggression toward wrongdoers (DeScioli & Kurzban, 2013) and may be the logic underlying the well-documented omission bias in moral psychology (DeScioli, Bruening, & Kurzban, 2011). And, as we mentioned in the introduction, challenges to power that require coordination among many actors are often effected by *public protests* and, increasingly, their electronic equivalents.

Common knowledge can also be conveyed nonverbally. Indeed, we propose that the nonverbal signals that accompany self-conscious emotions evolved with their peculiar anatomy and physiological configurations (Tracy & Matsumoto, 2008) precisely because those configurations are simultaneously salient to the expresser and the perceiver (see Provine, 1996, 2012, for a discussion). The perceiver not only knows the intended mental state of the expresser but knows that the expresser knows it, that the expresser knows that the perceiver knows it, and so on. Among these nonverbal common-knowledge generators may be the following.

- *Eye contact* is a potent social signal of threats and sexual come-ons precisely because both parties commonly know that they are acknowledging each other’s acknowledgment.
- *Blushing* is felt as a somatosensory sensation by the blusher at the same time as it is displayed as a change in skin color to the perceiver. The acute discomfort in blushing resides largely in the knowledge that the blusher knows he or she is blushing, knows that an onlooker knows it, that the onlooker knows that the blusher knows that the onlooker knows, and so on.
- *Crying* has the same inside-outside salience: A distraught person looking at onlookers through tears cannot avoid the knowledge that others know his tearful state, know that he knows, and so on.
- *Laughter*, with its disruption of the respiration rhythms necessary for speech and its unignorable noise, is also mutually salient to expresser and perceiver.

If this analysis is correct, it predicts that the common knowledge generated by each of these displays is necessary to attain a mutually beneficial equilibrium in a coordination game. (Although such knowledge may be less certain in real-world environments than in the experiments presented here, recall from the introduction that coordination can be effected by the weaker concept of common p-belief, which generalizes perfectly certain common knowledge to more realistic environments with less certainty.) Pinker (2007); Pinker, Nowak, and Lee (2008); and Lee and Pinker (2010) suggest that one such relevant game is the joint adoption of a relational model that consensually governs their interactions, such as communal sharing, authority ranking, equality matching, or market pricing (Fiske 1992, 2004). For example, two people can prosper if they agree to be friends and share things indiscriminately, or if

they agree to transact business and one sells something to the other, but not if one believes they are friends and helps himself to a possession that the other is in the business of selling. In the case of expressions of self-conscious emotions, the game may consist of two parties agreeing that one of them has committed an unintended or regretted harmful act or is in a vulnerable state and, thus, that the second one need not punish or ostracize him. This equilibrium leaves both of them better off than they would be if the second incurred the cost of punishing or ostracizing the first for a harm he would never repeat anyway (McCullough, 2008). If this theory of nonverbal communication is correct, expressions that are less likely to generate common knowledge (e.g., facial expressions that a person can express with little awareness he is expressing it) should not be yoked to an identifiable coordination game.

Because coordination and common knowledge by definition involve multiple parties, we should expect that they manifest themselves not just in small-scale, two- or three-person interactions but also within larger groups. That is, the role of common knowledge in solving coordination problems should manifest itself in a number of social-psychological phenomena. Here we can only list the rich possibilities for uniting diverse large-scale societal phenomena as manifestations of coordination problems that involve common knowledge:

- transactive memory (Wegner, 1995; Wegner, Erber, & Raymond, 1991), which may have interesting parallels with coordinating distributed storage in networked databases (Alberucci & Jäger, 2005; Halpern & Moses, 1990);
- creating and popping market bubbles (Dalkiran et al., 2012; Zuckerman, 2010);
- inaction brought on by diffusion of responsibility (Buchan, Croson, & Dawes, 2002);
- creating and maintaining social norms (Cronk & Leech, 2013), including the suboptimal norms that result from pluralistic ignorance (Centola, Willer, & Macy, 2005; Willer et al., 2009);
- the need to define hierarchical roles of leaders and followers when common knowledge is not attainable (Van Vugt, 2006);
- negotiation and bargaining (Ayres & Nalebuff, 1996; Schelling, 1960);
- international relations and diplomacy and so-called red lines (Byman, 2013; Hoffman & Yoeli, 2013);
- the subjective perception of currency valuation (Friedman, 1991);
- arbitrary groups, as studied by social identity theorists (Ockenfels & Werner, 2013; Yamagishi, Mifune, Liu, & Pauling, 2008);
- arbitrary rules of etiquette (Schelling, 1960);
- identifiable signals of conspicuous consumption (Veblen, 1899/2007);
- and many other kinds of seemingly arbitrary social constructions (Searle, 1995).

Finally, if common knowledge is a pervasive concern of social life, then it should leave a mark on language in the form of a *conceptual metaphor* (Lakoff & Johnson, 1980; Pinker, 2007): a family of idioms organized around a central image, such as ARGUMENT IS WAR or LOVE IS A JOURNEY. In the case of common knowledge, the central image alludes to the quintessential common-knowledge generator: COMMON KNOWLEDGE IS A CONSPICUOUS OBJECT OR SOUND. Thus we have a family of expressions that invoke a salient object or event to assert that some proposition or

speech act is common knowledge (and hence compels acknowledgment and action by two or more parties) or the converse, that even if some proposition is known by everyone it should strategically be kept out of common knowledge:

The emperor's new clothes.

The elephant in the room.

It's out there; you can't take it back.

It's on the record; to go on record.

The bell can't be unrung (also Some things once said cannot be unsaid).

That's a pretty big matzo ball hanging out there [when one person says "I love you" and the other doesn't reciprocate; from the television show *Seinfeld*]

A bald lie; a barefaced lie [cf. a veiled threat; a fig leaf]

To save face; to lose face.

That insult was in his face; he couldn't ignore it.

It's as plain as the nose on your face.

In recent decades, psychologists have recognized that cooperation is one of the hallmarks of the human species and that its game-theoretic demands have shaped our emotions, our morality, our social relationships, and our language. Much has been learned about these domains of psychology from a focus on the problem of altruistic cooperation and the mechanisms of reciprocity. We hope that comparable insights are waiting to be discovered by psychologists as they investigate the problem of mutualistic cooperation and as the mechanisms of common knowledge are—as we say—put out there.

References

- Abele, S., Stasser, G., & Chartier, C. (2014). Use of social knowledge in tacit coordination: Social focal points. *Organizational Behavior and Human Decision Processes*, 123, 23–33. doi:10.1016/j.obhdp.2013.10.005
- Alberucci, L., & Jäger, G. (2005). About cut elimination for logics of common knowledge. *Annals of Pure and Applied Logic*, 133, 73–99. doi:10.1016/j.apal.2004.10.004
- Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the Internet: The effect of \$1 stakes. *PLoS ONE*, 7, e31461. doi:10.1371/journal.pone.0031461
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–970. doi:10.1037/a0016923
- Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, 4, 1236–1239. doi:10.1214/aos/1176343654
- Austin, J. L. (1962). *How to do things with words* (2nd ed.). Cambridge, MA: Harvard University Press.
- Ayres, I., & Nalebuff, B. J. (1996). Common knowledge as a barrier to negotiation. *UCLA Law Review*, 44, 1631–1659.
- Bardsley, N., Mehta, J., Starmar, 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
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Barry, E. (2011, July 14). Sound of post-Soviet protest: Claps and beeps. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Becker, G. S. (1976). *The economic approach to human behavior*. Chicago, IL: University of Chicago Press.
- Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2013). Separating the shirkers from the workers? Making sure respondents pay attention to self-administered surveys. *American Journal of Political Science*. Advance online publication. doi:10.1111/ajps.12081
- Binmore, K., & Samuelson, L. (2001). Coordinated action in the electronic mail game. *Games and Economic Behavior*, 35, 6–30. doi:10.1006/game.1998.0749
- Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, 14, 283–286. doi:10.1111/1467-9280.03436
- Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18, 382–386. doi:10.1111/j.1467-9280.2007.01909.x
- Buchan, N. R., Croson, R. T. A., & Dawes, R. M. (2002). Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. *American Journal of Sociology*, 108, 168–206. doi:10.1086/344546
- Byman, D. (2013, May 4). Mr. Obama, don't draw that line. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., . . . Singh, S. (2005). Synchrony in the onset of mental-state reasoning. *Psychological Science*, 16, 378–384. doi:10.1111/j.0956-7976.2005.01544.x
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19, 7–42. doi:10.1023/A:1007850605129
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97, 1232–1254. doi:10.1086/261651
- Centola, D., Willer, R., & Macy, M. (2005). The emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, 110, 1009–1040. doi:10.1086/427321
- Chartier, C., Abele, S., Stasser, G., & Shriver, E. (2012, January). *Secondary and Schelling salience in matching and mismatching coordination*. Poster session presented at the meeting of the Society for Personality and Social Psychology, San Diego, CA.
- Chaudhuri, A., Schotter, A., & Sopher, B. (2009). Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice. *Economic Journal*, 119, 91–122. doi:10.1111/j.1468-0297.2008.02207.x
- Chwe, M. S. (2001). *Rational ritual: Culture, coordination, and common knowledge*. Princeton, NJ: Princeton University Press.
- Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.
- Clark, H. H. (1996). *Using language*. New York, NY: Cambridge University Press.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). New York, NY: Cambridge University Press.
- Clifford, S., & Jerit, J. (2012). *Do attempts to improve respondent attention increase social desirability bias?* Retrieved from the American Political Association website: <https://www.apsanet.org>
- Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26, 139–153. doi:10.1017/S0140525X03000050

- Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences, USA*, 106, 9163–9168. doi:10.1073/pnas.0807721106
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford, United Kingdom: Oxford University Press.
- Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 584–627). Hoboken, NJ: Wiley.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185. doi:10.1017/S0140525X01003922
- Cronk, L., & Leech, B. L. (2013). *Meeting at Grand Central: Understanding the social and evolutionary roots of cooperation*. Princeton, NJ: Princeton University Press.
- Curry, O., & Chesters, M. J. (2012). “Putting ourselves in the other fellow’s shoes”: The role of “theory of mind” in solving coordination problems. *Journal of Cognition and Culture*, 12, 147–159. doi:10.1163/156853712X633974
- Dalkiran, N. A., Hoffman, M., Paturi, R., Ricketts, D., & Vattani, A. (2012). Common knowledge and state-dependent equilibria. In *Proceedings of Symposium on Algorithmic Game Theory* (pp. 84–95). doi:10.1007/978-3-642-33996-7_8
- DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior*, 32, 204–215. doi:10.1016/j.evolhumbehav.2011.01.003
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477–496. doi:10.1037/a0029065
- Diesendruck, G., & Markson, L. (2001). Children’s avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, 37, 630–641. doi:10.1037/0012-1649.37.5.630
- Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system? Screening Mechanical Turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2399–2402). doi:10.1145/1753326.1753688
- Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: Reactive egoism in social interaction. *Journal of Personality and Social Psychology*, 91, 872–889. doi:10.1037/0022-3514.91.5.872
- Fischer, I. (2009). Friend or foe: Subjective expected relative similarity as a determinant of cooperation. *Journal of Experimental Psychology: General*, 138, 341–350. doi:10.1037/a0016073
- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99, 689–723. doi:10.1037/0033-295X.99.4.689
- Fiske, A. P. (2004). Four modes of constituting relationships: Consubstantial assimilation; space, magnitude, time, and force; concrete procedures; abstract symbolism. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 61–146). Mahwah, NJ: Erlbaum.
- Friedman, M. (1991). *The island of stone money* (Working Paper No. E-91-3). Retrieved from <http://hoohila.stanford.edu/workingpapers/getWorkingPaper.php?filename=E-91-3.pdf>
- Frith, U., & Frith, C. D. (2003). Developmental and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B*, 358, 459–473. doi:10.1098/rstb.2002.1218
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493–501. doi:10.1016/S1364-6613(98)01262-5
- Geanakoplos, J. (1992). Common knowledge. *Journal of Economic Perspectives*, 6, 53–82. doi:10.1257/jep.6.4.53
- Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford, England: Oxford University Press.
- Goldberg, L. R. (1992). The development of markers of the Big Five factor structure. *Psychological Assessment*, 4, 26–42. doi:10.1037/1040-3590.4.1.26
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Dreary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, the Netherlands: Tilburg University Press.
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26, 213–224. doi:10.1002/bdm.1753
- Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person \times situation perspective. *Journal of Personality and Social Psychology*, 93, 583–599. doi:10.1037/0022-3514.93.4.583
- Graziano, W. G., & Tobin, R. M. (2002). Agreeableness: Dimension of personality or social desirability artifact? *Journal of Personality*, 70, 695–728. doi:10.1111/1467-6494.05021
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody’s watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245–256. doi:10.1016/j.evolhumbehav.2005.01.002
- Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37, 549–587. doi:10.1145/79147.79161
- Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others’ goals. *Developmental Science*, 11, 487–494. doi:10.1111/j.1467-7687.2008.00694.x
- Hoffman, M., & Yoeli, E. (2013, September 14). Effective red lines. *The Economist*. Retrieved from <http://www.economist.com>
- Hofstadter, D. (1985). *Metamagical thems: Questing for the essence of mind and pattern*. New York, NY: Basic Books.
- Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14, 399–425. doi:10.1007/s10683-011-9273-9
- Hume, D. (1969). *A treatise of human nature*. London, United Kingdom: Penguin. (Original work published 1739–1740)
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25–41. doi:10.1016/S0010-0277(03)00064-7
- Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89, 191–204. doi:10.1111/j.2044-8295.1998.tb02680.x
- Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the “who said what?” paradigm. *Journal of Personality and Social Psychology*, 75, 1155–1178. doi:10.1037/0022-3514.75.5.1155
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.
- Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, 31, 215–226. doi:10.1016/S0191-8869(00)00130-6
- Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, 117, 785–807. doi:10.1037/a0019688
- Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Lieberman, D., Oum, R., & Kurzban, R. (2008). The family of fundamental social categories includes kinship: Evidence from the memory confusion paradigm. *European Journal of Social Psychology*, 38, 998–1012. doi:10.1002/ejsp.528

- McAdams, R. H., & Nadler, J. (2005). Testing the focal point theory of legal compliance: The effect of third-party expression in an experimental hawk/dove game. *Journal of Empirical Legal Studies*, 2, 87–123. doi:10.1111/j.1740-1461.2005.00032.x
- McCullough, M. E. (2008). *Beyond revenge: The evolution of the forgiveness instinct*. San Francisco, CA: Jossey-Bass.
- Mehta, J., Starmer, C., & Sugden, R. (1994a). Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, 36, 163–185. doi:10.1007/BF01079211
- Mehta, J., Starmer, C., & Sugden, R. (1994b). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review*, 84, 658–673. doi:10.1007/BF01079211
- Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, 4, 1–25. doi:10.1016/0022-1031(68)90046-2
- Milinski, M., Semmann, D., & Krambeck, H.-J. (2002, January 24). Reputation helps solve the “tragedy of the commons.” *Nature*, 415, 424–426. doi:10.1038/415424a
- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352. doi:10.1121/1.1907526
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50, 655–663. doi:10.1016/j.neuron.2006.03.040
- Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1, 170–190. doi:10.1016/0899-8256(89)90017-1
- Nicholson, N., Soane, E., Fenton-O’Creedy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, 8, 157–176. doi:10.1080/1366987032000123856
- Nov, O., & Rafaeli, S. (2009). Measuring the premium on common knowledge in computer-mediated coordination problems. *Computers in Human Behavior*, 25, 171–174. doi:10.1016/j.chb.2008.08.005
- Ockenfels, A., & Werner, P. (2013). Beliefs and ingroup favoritism. *Journal of Economic Behavior & Organization*. Advance online publication. doi:10.1016/j.jebo.2013.12.003
- Onishi, K. H., & Baillargeon, R. (2005, April 8). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258. doi:10.1126/science.1107621
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45, 867–872. doi:10.1016/j.jesp.2009.03.009
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411–419.
- Pedersen, E. J., Kurzban, R., & McCullough, M. E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B: Biological Sciences*, 280, Article 20122723. doi:10.1098/rspb.2012.2723
- Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that . . .”: Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39, 437–471. doi:10.1016/0022-0965(85)90051-7
- Pierce, J. R., Kilduff, G. J., Galinsky, A. D., & Sivanathan, N. (2013). From glue to gasoline: How competition turns perspective takers unethical. *Psychological Science*, 24, 1986–1994. doi:10.1177/0956797613482144
- Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York, NY: Viking.
- Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences, USA*, 105, 833–838. doi:10.1073/pnas.0707192105
- Provine, R. R. (1996). Laughter. *American Scientist*, 84, 38–45. Retrieved from <http://www.jstor.org/stable/29775596>
- Provine, R. R. (2012). *Curious behavior: Yawning, laughing, hiccupping, and beyond*. doi:10.4159/harvard.9780674067226
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012, September 20). Spontaneous giving and calculated greed. *Nature*, 489, 427–430. doi:10.1038/nature11467
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin*, 28, 789–801. doi:10.1177/0146167202289008
- Rousseau, J. (1984). *A discourse on inequality* (M. Cranston, Trans.). New York, NY: Penguin Books. (Original work published 1754)
- Rubinstein, A. (1989). The electronic mail game: Strategic behavior under “almost common knowledge.” *American Economic Review*, 79, 385–391. Retrieved from <http://www.jstor.org/stable/1806851>
- Saxe, R., & Young, L. (in press). Theory of mind: How brains think about thoughts. In K. Ochsner & S. Kosslyn (Eds.), *Handbook of cognitive neuroscience*. New York, NY: Oxford University Press.
- Scaife, M., & Bruner, J. S. (1975, January 24). The capacity for joint visual attention in the infant. *Nature*, 253, 265–266. doi:10.1038/253265a0
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Searle, J. (1989). How performatives work. *Linguistics and Philosophy*, 12, 535–558. doi:10.1007/BF00627773
- Searle, J. (1995). *The construction of social reality*. New York, NY: Free Press.
- Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. New York, NY: Cambridge University Press.
- Smith, N. (1982). *Mutual knowledge*. Orlando, FL: Academic Press.
- Sugden, R. (1995). A theory of focal points. *Economic Journal*, 105, 533–550. doi:10.2307/2235016
- Summerville, A., & Charrier, C. R. (2013). Pseudo-dyadic “interaction” on Amazon’s Mechanical Turk. *Behavior Research Methods*, 45, 116–124. doi:10.3758/s13428-012-0250-9
- Sun, T., & Wu, G. (2012). Traits, predictors, and consequences of Facebook self-presentation. *Social Science Computer Review*, 30, 419–433. doi:10.1177/0894439311425978
- Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE*, 6, e16836. doi:10.1371/journal.pone.0016836
- Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, 142, 151–162. doi:10.1037/a0028232
- Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 36, 778–793. doi:10.1037/0022-3514.36.7.778
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.
- Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human morality & sociality: Evolutionary and comparative perspectives* (pp. 91–234). New York, NY: Palgrave Macmillan.
- Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics*, 27, 103–129. doi:10.1002/mde.1287
- Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Pro-*

- ceedings of the National Academy of Sciences, USA*, 105, 11655–11660. doi:10.1073/pnas.0802686105
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57. doi:10.1086/406755
- 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 Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review*, 10, 354–371. doi:10.1207/s15327957pspr1004_5
- Veblen, T. (2007). *The theory of the leisure class*. New York, NY: Oxford University Press. (Original work published 1899)
- Wegner, D. M. (1995). A computer network model of human transactive memory. *Social Cognition*, 13, 319–339. doi:10.1521/soco.1995.13.3.319
- Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61, 923–929. doi:10.1037/0022-3514.61.6.923
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684. doi:10.1111/1467-8624.00304
- Willer, R., Kuwabara, K., & Macy, M. (2009). The false enforcement of unpopular norms. *American Journal of Sociology*, 115, 451–490. doi:10.1086/599250
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128. doi:10.1016/0010-0277(83)90004-5
- Yamagishi, T., Mifune, N., Liu, J. H., & Pauling, J. (2008). Exchanges of group-based favours: Ingroup bias in the prisoner's dilemma game with minimal groups in Japan and New Zealand. *Asian Journal of Social Psychology*, 11, 196–207. doi:10.1111/j.1467-839X.2008.00258.x
- Zuckerman, E. W. (2010). What if we had been in charge? The sociologist as builder of rational institutions. In M. Lounsbury & P. M. Hirsch (Eds.), *Markets on trial: The economic sociology of the U.S. financial crisis: Part B* (pp. 359–378). doi:10.1108/S0733-558X(2010)000030B016

Received July 1, 2013

Revision received April 1, 2014

Accepted April 8, 2014 ■