Common knowledge, coordination, and strategic mentalizing in human social life

Julian De Freitas[a], Kyle Thomas[b], Peter DeScioli[c], and Steven Pinker[a,1]

[a]Department of Psychology, Harvard University, Cambridge, MA 02138; [b]MotiveMetrics Inc., Palo Alto, CA 94306; and [c]Department of Political Science, Stony Brook University, Stony Brook, NY 11794

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2016.

Contributed by Steven Pinker, May 1, 2019 (sent for review April 1, 2019; reviewed by Oliver Scott Curry and Michael Macy)

People often coordinate for mutual gain, such as keeping to opposite sides of a stairway, dubbing an object or place with a name, or assembling en masse to protest a regime. Because successful coordination requires complementary choices, these opportunities raise the puzzle of how people attain the common knowledge that facilitates coordination, in which a person knows X, knows that the other knows X, knows that the other knows that he knows, ad infinitum. We show that people are highly sensitive to the distinction between common knowledge and mere private or shared knowledge, and that they deploy this distinction strategically in diverse social situations that have the structure of coordination games, including market cooperation, innuendo, bystander intervention, attributions of charitability, self-conscious emotions, and moral condemnation.

coordination | common knowledge | theory of mind | cooperation | bystander effect

In many situations, people coordinate with others for mutual gain. We agree on a time and place to meet, bring complementary fare to a potluck dinner, or divide responsibilities on a research project. Although sometimes effortless, coordination can fail if people are not on the same page, even when they want the same thing. Schedules clash, misunderstandings proliferate, and shared goals fall through the cracks or are spoiled by too many cooks. Coordination dilemmas are found not just in everyday situations but in the conventions and norms that make complex societies possible. The challenge of coordination is coming to be recognized as one of the deepest puzzles in the human sciences (1–4), and it requires a special kind of reasoning about mental states that goes beyond representing others' beliefs.

Imagine that Tom and Leyla are separated in a crowd and need to find each other. Tom knows Leyla's favorite place is the café and so he suspects she'll go there. But then he realizes that Leyla might go to his favorite place, the bar. Then again, Tom thinks, Leyla might reason that he will go to her favorite, the café. Tom continues to reason about Leyla's reasoning (about his reasoning, and so on), but this brings him no closer to a solution. Tom then notices a tower in the middle of the square. He infers that Leyla would see the tower too and recognize it as the obvious place to meet. Sure enough, she is waiting for him at the tower when he arrives.

Tom didn't coordinate with Leyla by explicitly thinking through her beliefs about his beliefs, and so on. Rather, he intuitively recognized the tower as a location ideally suited for the coordination. How does this kind of reasoning work? How does one read the mind of a mind reader?

Thomas Schelling (5) explained that to coordinate without communication, intelligent agents technically need infinitely nested knowledge of a single solution, which the philosopher David Lewis (6) called "common knowledge." The difference between common knowledge and private and shared knowledge is illustrated in Fig. 1.

In scenarios like the incommunicado rendezvous conundrum, the players need common knowledge not just of the problem and its potential solutions, but of which of these solutions to opt for among the multiple equilibria (situations in which no player can benefit by changing his choice without another player also having to change her choice). This problem is unsolvable within the framework of standard game theory, because rational players could choose any of these equilibria (5, 7, 8). Schelling (5) proposed that humans have effective psychological workarounds. In particular, they tend to settle on a focal point—an option that stands out so conspicuously that one can infer that others notice it and sense that everyone else notices it—in effect making this solution common knowledge among the observers, given their shared psychology.* For example, Tom and Leyla not only know about the tower, but can infer that Tom knows that Leyla knows that Tom knows, and so on.

The possibility that salient focal points, a psychological phenomenon, allow people to attain the common knowledge necessary to solve coordination dilemmas suggests there is a rich area of intersection between game theory on the one hand and cognitive and social psychology on the other. Surprisingly, this territory has barely been explored by either side. The game-theoretic role of common knowledge has been studied by mathematicians (10, 12), philosophers (6, 13), economists (14, 15), linguists (16–19), sociologists (20), political scientists (21), and computer scientists (22, 23),

Significance

Humans are an unusually cooperative species, and our cooperation is of 2 kinds: altruistic, when actors benefit others at a cost to themselves, and mutualistic, when actors benefit themselves and others simultaneously. One major form of mutualism is coordination, in which actors align their choices for mutual benefit. Formal examples include meetings, division of labor, and legal and technological standards; informal examples include friendships, authority hierarchies, alliances, and exchange partnerships. Successful coordination is enabled by common knowledge: knowledge of others' knowledge, knowledge of their knowledge of one's knowledge, ad infinitum. Uncovering how people acquire and represent the common knowledge needed for coordination is thus essential to understanding human sociality, from large-scale institutions to everyday experiences of civility, hypocrisy, outrage, and taboo.

Author contributions: J.D.F., K.T., P.D., and S.P. designed research; J.D.F., K.T., P.D., and S.P. performed research; J.D.F., K.T., and P.D. analyzed data; and J.D.F., K.T., P.D., and S.P. wrote the paper.

Reviewers: O.S.C., University of Oxford; and M.M., Cornell University.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: pinker@wjh.harvard.edu.

Published online June 28, 2019.

*Because the information generated falls short of certainty, technically it is "common belief" rather than "common knowledge," but the points we make about common knowledge in this report also hold for common belief with a sufficiently high degree of credence (9–11), and in this report we treat them interchangeably.

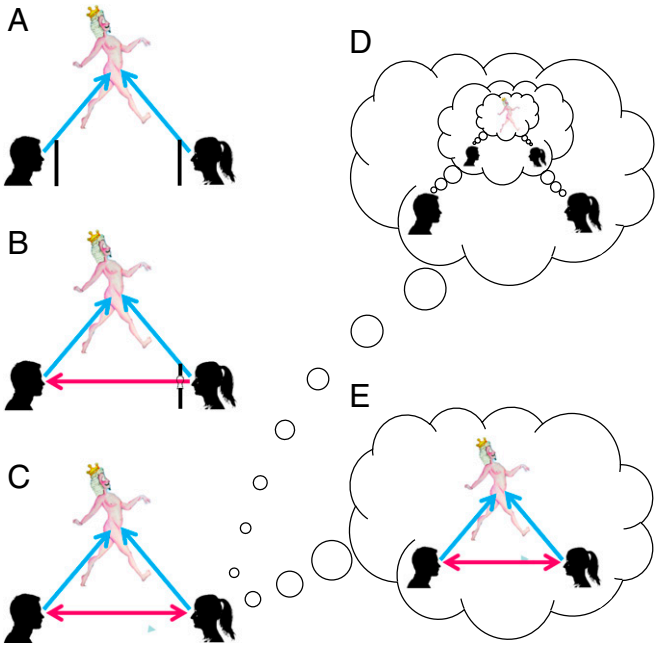

Fig. 1. Levels of knowledge. (*A*) Private knowledge, where each person knows something, but knows nothing about what anyone else knows. (*B*) Shared knowledge, where each person knows something, and also knows that other people know it. (*C*) Common knowledge, where everybody knows that everybody else knows it. (*D*) A hypothetical cognitive representation of shared knowledge; common knowledge would consist of the number of nested thoughts being infinite. (*E*) A plausible cognitive representation of common knowledge, in which the knower senses that something is publicly observable.

but this research treats the human recognition and representation of common knowledge as a black box. Cognitive scientists, for their part, have developed a large literature on mentalizing or "theory of mind," but they have focused almost entirely on how people represent others' beliefs about some state of affairs in the world, not how they represent their beliefs about beliefs, including common knowledge (for reviews, see refs. 24–26).

Here we review experiments designed to test the hypothesis that the human mind is acutely sensitive to common knowledge as an adaptation to successfully coordinate with others. Common knowledge is a kind of recursive metalizing: holding beliefs about other people's beliefs about other people's beliefs, and so on. At first glance, recursive mentalizing seems to be an unlikely accomplishment, because people have trouble entertaining multiply nested propositions, such as "She thinks that I think that she thinks that I think that...," which quickly overload the limited recursive capacity of human cognition (27).

In some circumstances, though, people can think about multiple levels of others' thoughts (28, 29). Young school-age children, for example, can reason that John mistakenly thinks that Mary is unaware that an ice-cream truck has moved (30). People are most likely to succeed at such reasoning when the number of embeddings is small, when the nested mental states are not identical so that the representation is not truly recursive, and when particular combinations of mental states are familiar and can be represented as single chunks. For example, everyday attributions that someone is "judgmental," "perceptive," "tactless," "compassionate," or "sadistic" require thinking about that person's beliefs about the mental states of others. In close-knit social circles, people routinely make convoluted attributions, such as Alice thinking that Bob is mistakenly worrying that Carol is offended by misunderstanding something Dave had said (31, 32).

Yet, the state of common knowledge itself cannot be represented as a set of propositions nested within other propositions, because no matter how many are embedded, it would fall short of the infinite number that distinguishes common knowledge from mere shared knowledge. Technically, common knowledge can be captured by a recursive formula, such as "Y: Everyone knows X, and everyone knows Y." We suggest that people represent common knowledge even more simply: as a distinctive cognitive state corresponding to the sense that something is public, unignorable, or "out there."

Common knowledge, moreover, need not be ascertained by reasoning through other people's beliefs about other people's beliefs and generalizing inductively to an infinity of levels. Rather, people can infer it by using a variety of perceptual or conceptual cues, including focal points, a broadcasted message, public rituals, eye contact, or a blurted-out statement. A prototype may be found in *The Emperor's New Clothes*, in which every onlooker privately knew that the emperor was naked, but could not be positive that everyone else noticed it until the public exclamation by the little boy made the fact common knowledge.

The story also has another moral. The common knowledge created by the boy's exclamation changed not just the onlookers' cognitive awareness but also their social relationship with the emperor and with each other, emboldening them to laugh at him. The reason, we suggest, is that navigating between different kinds of relationships is a coordination game, and people use common knowledge to ratify which kind of relationship holds between them. The game theory of coordination games thereby predicts that people's sensitivity to common knowledge is ubiquitous in social life, appearing in a diverse array of cooperative opportunities.[†]

Here we review experiments that place participants in real or imagined social situations which have the structure of coordination games. Sometimes the coordination payoffs are explicit, costed out in dollars and cents, but sometimes they are less obvious because the costs and benefits are social and emotional. The experiments manipulate whether participants receive information that generates private, shared, or common knowledge. The experiments reveal that people are sensitive to the difference, and respond strategically in the ways predicted by the game-theoretic analysis of coordination.

Economic Cooperation for Mutual Profit

Mehta et al. (34, 35) reported the first formal experiments on coordination and focal points. Participants answered underspecified questions, such as, "Write down any positive number," and "Name any flower." In the coordination condition, participants could earn bonus money if they chose the same answer as another randomly selected participant. Mehta et al. found that participants could coordinate their answers well above chance and well above a control condition where they were not incentivized to coordinate. Participants didn't pick their favorite choice or what they thought was their partner's favorite choice, but picked focal points: salient options they inferred stood out for everyone, such as "1" for a number or "rose" for a flower.

In these and subsequent studies (1, 36–39), an experimenter explicitly instructed the participants to coordinate their guesses, and the coordination required only a single nested belief (what

[†]Not in all of them, however: Thomas et al. (11) and Baumard et al. (33) note that most research in social and evolutionary psychology has focused on altruistic cooperation, in which an organism benefits another at a cost to himself, in which interacting choices may be modeled as a Prisoner's Dilemma, and in which successful cooperation hinges on reciprocity and the moral emotions supporting it, such as gratitude, anger, and guilt. Less research has examined mutualistic cooperation, in which an organism benefits himself and another simultaneously, interacting choices have the structure of a coordination game, and successful cooperation hinges on common knowledge. Even the Prisoner's Dilemma becomes a form of coordination game when it is repeated indefinitely and the players use conditional strategies, like tit-for-tat.

others would guess). The studies do not show that people spontaneously figure out that in social situations they need common knowledge as a means to the end of coordinating to their mutual benefit. So we examined how people use different levels of knowledge in a coordination game that was implicit in a fictitious scenario (11). Participants played the role of merchants (a butcher or baker) who had 2 choices: work alone by making either dinner rolls (if they were the baker) or chicken wings (if the butcher), or try to work together to sell hot dogs, which requires one merchant to bring the meat and another to bring the bun. Crucially, the market price of hot dogs varied, and on different days could be more or less than the revenue each merchant would earn working alone. When hot dogs were more profitable, participants faced a classic coordination dilemma (often called a "stag hunt" in game theory, after a scenario in which 2 hunters can hunt alone for rabbits, a small but sure catch, or coordinate their efforts to fell a stag, which is larger but riskier). The payoffs defining the dilemma, with its 2 profitable equilibria, are shown in Table 1.

The point was to see how participants reacted once they learned that they were in a coordination dilemma, in which working together would be more profitable due to a high market price for hot dogs. Across conditions, this information was revealed in different ways which, crucially, affected their knowledge of their counterpart's knowledge. In the "private knowledge" condition, a messenger told the participant that hot dogs were selling at a higher price but did not say whether his counterpart knew this. (In all conditions, the states of knowledge were introduced in a plausible context, such as the messenger having visited and chatted—or not—with the other merchant.) With "secondary knowledge," the messenger told the participant the price and added that the partner also knew it (but no more). With "tertiary knowledge," the messenger told the participant the price, said that the partner knew it, and also said that the partner knew that the participant knew it. In the "common knowledge" condition, the price was announced on a loudspeaker: an example of public, broadcasted, or mutually salient information, which we hypothesize people treat as common knowledge.

As predicted, participants chose to work together more often with secondary or tertiary shared knowledge than with private knowledge, and most often with common knowledge (Fig. 2), allowing them to profit in a classic coordination dilemma.

Participants also answered a series of questions about their partner's knowledge. To test the hypothesis that common knowledge is a distinctive cognitive category rather than a high point on a continuum of knowledge embedding, we examined the confusion matrix of these responses (the proportions of times that participants identified each knowledge state as each of the other knowledge states), under the time-honored assumption that if 2 stimuli are mentally encoded in the same way, they will be easily confused with each other (40). The confusion matrix is shown in Table 2, which confirms that participants often confused the 2 conditions of shared knowledge with each other (e.g., 23% of participants mistook tertiary for secondary knowledge), but rarely confused common knowledge with either shared or private knowledge. The pattern of errors is consistent with the hypothesis that common knowledge is a psychologically distinctive state.

Table 1. Payoffs for the coordination game in the market experiment

	Butcher's options	
Baker's options	Work together (hot dogs)	Work alone (chicken wings)
Work together (hot dogs)	\$1.10, \$1.10	\$0, \$1.00
Work alone (dinner rolls)	\$1.00, \$0	\$1.00, \$1.00

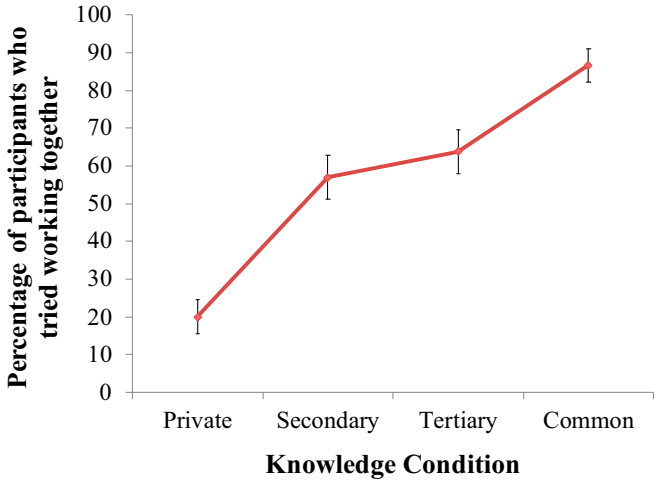

Fig. 2. Percentage of participants who tried to work together, depending on their knowledge and their partner's knowledge. Reprinted with permission from ref. 11.

This study presented participants with a stag hunt scenario in an imaginable context, but stipulated the structure of the game with numerical payoffs. In 4 other studies, we have shown that people recognize coordination dilemmas and ascertain common knowledge in diverse social predicaments in which the payoffs are implicit and abstract.

The Bystander Effect

In the classic bystander effect, the more people who could intervene to resolve a problem, the less likely any one of them will do so (41–43). For example, a victim in physical jeopardy is more likely to be rescued by a single bystander than by any member of a group of them. The standard explanation is that responsibility diffuses across the multiple bystanders, diluting the responsibility of each (41–45).

But the metaphors of diffusion and dilution do not capture the game-theoretic structure of bystander dilemmas, and thus miss the possibility that intervention may be a strategic decision that depends on the bystanders' knowledge. Diekmann (46, 47) analyzed bystanders' choices with a game he called the volunteer's dilemma: if someone intervenes, then each bystander enjoys a benefit (say, in the reduction of psychological distress at the thought of a person in danger), but the one who intervenes incurs a cost in risk or time. The best outcome, then, is for someone else to intervene, and the worst is for no one to intervene, with oneself intervening falling in between. The volunteer's dilemma has a mixed strategy equilibrium: each player should randomly help with a certain probability, determined by the payoffs.[‡] Crucially, when there are more players, that probability drops. The bystander effect is thus consistent with the strategic choices of rational players in the volunteer's dilemma.

Even more crucially for the present discussion, if the bystander effect is a rational strategy, it should depend on what each bystander knows about what the others know. When only one person knows of a problem requiring a volunteer, and knows he is the only one who knows, his best option is to intervene, because otherwise the problem is guaranteed to be unsolved. If, however, the player knows about the problem, knows his partner knows, but also knows that the partner is unaware that he knows, he can safely leave the job to the partner. In other words,

[‡]Mixed strategies are optimal in "outguessing standoffs," such as a rock-paper-scissors game or a soccer player facing a goalkeeper in a penalty kick.

Table 2. Participants' judgments of knowledge level (%) by condition

	Reported level of knowledge				
Condition	Private	Secondary	Tertiary	Common	Unclassifiable
Private	**0.931**	0.008	0.008	0.008	0.046
Secondary	0.020	**0.899**	0.013	0.007	0.060
Tertiary	0	0.230*	**0.637**	0.044	0.089
Common	0.007	0	0.015	**0.837**	0.141

Participants' judgments of knowledge level in each of the knowledge conditions. Accurate judgments are in bold.
*Participants were more likely to mistake tertiary and secondary knowledge, compared with tertiary and common knowledge (sign test, $P < 0.001$). From ref. 11, experiment 3.

asymmetric knowledge should impel one player to help and the others to shirk. But when players have common knowledge of the problem, they all have the same information, occupy symmetric positions in an outguessing standoff, and should help with the same probability (as if hoping that someone else will intervene but hedging their bets in case no one does). In this case, common knowledge can lead to an outcome that is worse overall than with private knowledge: namely, everyone shirking.

To test this theory, we placed participants in a fictional volunteer's dilemma (48), playing merchants who were on call to help their landlord, Smith. When Smith required help with a chore, at least one merchant had to step in, forgoing part of his daily earnings. If no one volunteered, all merchants had to pay a larger fine (Table 3).

Across conditions, we manipulated whether the participants were in groups of 2 or 5, and whether they had private, secondary, tertiary, quaternary, or common knowledge about the job, provided either by a messenger or over a loudspeaker (as in the experiments described in *Economic Cooperation for Mutual Profit*).

Replicating the classic bystander effect, participants were less likely to help when there were 5 rather than 2 bystanders (Fig. 3). However, this difference was absent or diminished when they had asymmetric knowledge. Specifically, we found a zigzag quartic pattern of high helping for private and tertiary knowledge and low helping for secondary and quaternary knowledge. This is precisely what the game-theoretic analysis of the volunteer's dilemma predicts: with private knowledge, a person might be the only one who knows help is needed, and so should help; with secondary knowledge, a person should not help, since their counterpart (who has private knowledge) can be relied upon to help; with tertiary knowledge, the person should help, since their counterpart (who has secondary knowledge) should not help; and so on, for increasing levels of knowledge. In line with this zigzagging logic, participants strategically tracked the knowledge states of their counterparts in order to help when they thought the others were unlikely to help. Finally, as predicted by the mixed strategy equilibrium, we found that the classic bystander effect (the more bystanders, the less helping) reliably occurred only when the need for help was common knowledge. Thus, we can better explain a classic finding in social psychology by considering how people's strategic decisions and common knowledge can foster, or impede, cooperation.

Innuendo and Other Forms of Indirect Speech

Why do speakers so often veil their intentions with innuendo and euphemism, rather than blurting out what they mean? Indirect speech is particularly common in socially fraught situations, such as sexual come-ons, illicit financial transactions, and threats. The logic of coordination and common knowledge explains this puzzle.

Pinker and his collaborators (19, 49, 50) proposed that different social relationships, such as those found between intimates, an authority and a subordinate, or trading partners, are alternative equilibria in a coordination game. Drawing on Fiske's Relational Models theory (51, 52), they proposed that both parties in a given relationship benefit by converging on a single understanding of how to allocate certain resources between them, such as by unmeasured sharing, a license to confiscate, tit-for-tat reciprocity, or rule-governed pricing. When expectations are mismatched (such as a citizen bribing a police officer, a supervisor soliciting a sexual favor, or a friend selling a car to another friend), the clash could be damaging for everyone (see refs. 9 and 49 for game-theoretic analyses). The choice of a Relational Model, like other coordination games, is ratified by common knowledge: each of 2 partners knows that the other considers herself a friend (or a lover, or a boss, or a customer), knows that the other knows she knows this, and so on.

This sets up the problem of how people can deviate from the expectations of their relationship, whether as a one-time exception (such as one friend needing a big unreciprocated favor from another) or as the first move in a transition to a different relationship (such as from friendship to romance). The problem is that the very act of entertaining a deviation calls into question the relational model currently in force. Indirect propositions solve this problem by allowing a speaker to proffer a relationship-threatening message while keeping the message out of common knowledge, which would destabilize the extant equilibrium. For example, if Michael wants to ascertain whether Lisa, a coworker, is willing to have sex with him, with the risk that revealing his interest could irrevocably change their professional relationship, he could ask, "Would you like to come up to see my etchings?" or "Wanna come over to Netflix and chill?" Both may know that this is a sexual overture, but since the proposition was never overtly stated, if she is not interested he could plausibly deny the invitation was sexual. Critically, even if the denial of the intended meaning is not genuinely plausible to a sophisticated adult, denial of common knowledge of the intent may be. While both parties privately realize a sexual overture was tendered and rebuffed, Lisa could think, "Maybe he thinks I'm naïve," while Michael could think, "Maybe she thinks I'm dense," allowing them to move on with a common understanding of a platonic relationship.

Lee and Pinker (50) asked participants to consider scenarios in which a person issues a threat, a sexual come-on, or a bribe, either overtly, presumably generating common knowledge (as in, "I'm very sorry, officer. If I give you a fifty, will you just let me go?") or euphemistically, presumably avoiding common knowledge (as in, "Maybe the best thing would be to take care of this here without going to court or doing any paperwork."). To test participants' understanding of the characters' recursive mental states without overloading them with multiply nested sentences, we asked them to put themselves in the shoes of the speaker or hearer and answer questions about how they or their interlocutor interpreted the intentions or understanding of their counterpart. In the Second-Order Hearer condition, for example, they were asked, "Put yourself in the officer's situation. He knows that Kyle was really trying to bribe him, and he has turned down the offer. Which of the following is the most likely thing that the officer is thinking at this point?" They were then asked to rate which of 7

Table 3. Volunteer's dilemma game

	Other merchant(s)	
Merchant	Help	Shirk
Help	\$0.50, \$0.50	\$0.50, \$1.00
Shirk	\$1.00, \$0.50	\$0.00, \$0.00

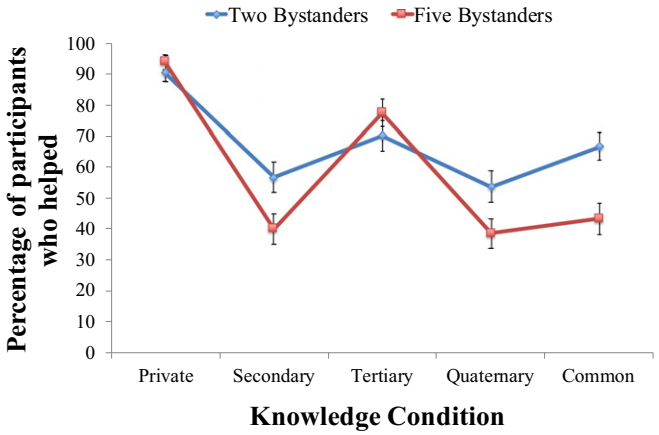

Fig. 3. Percentage of participants who decided to help, depending on their knowledge and on the group size. Reprinted with permission from ref. 48.

thoughts the officer was most likely thinking, ranging from ignorance ("This guy thinks that I didn't understand he was offering me a bribe") through slight suspicion ("This guy probably doesn't think I understood that he was offering me a bribe") to strong suspicion ("This guy probably knows that I understood he was offering me a bribe") to certainty ("This guy knows that I understood he was offering me a bribe").

As predicted, with indirect utterances participants were less and less certain that the hearer attributed the fraught intent to the speaker, and that the speaker realized this, the more deeply embedded each one's thoughts about the other were. But with the bald proposition, they were certain at every level of embedding (for example, attributing to the officer the thought: "This guy understands that I turned down his bribe. And he realizes that I know he understands that") (Fig. 4). This confirms that a direct proposition generates common knowledge, whereas a veiled proposition, even one whose underlying intent is obvious, does not. That in turn suggests that the function of innuendo is to protect a social relationship against the threat posed by common knowledge of a compromising proposition.

Self-Conscious Emotions

When people commit transgressions, they feel the self-conscious emotions of embarrassment, guilt, or shame (53–56). These emotions are widely understood to repair relationships after a breach of their expected norms. We add a corollary: Since relationships require coordination, the intensity of the emotions should depend on whether the transgression is common knowledge.

For example, someone who turns red after audibly flatulating, or mocking a friend without realizing she was within earshot, has honestly signaled that he recognizes that he has violated an expectation, regrets the action, and cares about what others think of him (as opposed to being a social scofflaw or psychopath with no respect for the expectations in the first place). But if the offender could credibly believe that the faux pas had escaped notice, or even that the onlookers could not know that he knew they had noticed, a blushing acknowledgment would not be as pressing, because his lack of acknowledgment of the transgression is not a lack of acknowledgment of the expectation. Similarly, if the onlookers value the relationship, they may choose to ignore the transgression (or to act as if the transgressor did not know that they knew), since that would allow them to avoid ratifying a breach of the norm.

Thomas et al. (57) tested the hypothesis that people feel self-conscious emotions more intensely when their transgressions are common knowledge with observers. Participants rated how embarrassed, guilty, and ashamed they would feel after flatulating during a lecture, mocking a friend behind her back, or falsifying a reimbursement request. The participant had either: 1) private knowledge: only the participant knew the transgression; 2) secondary knowledge: the participant knew that an observer knew; or 3) common knowledge. As predicted, participants reported that they would feel more intense embarrassment, guilt, and shame, and would exhibit more intense physical reactions, such as blushing, hanging their heads, and nervous laughter, when the transgression was common knowledge (Fig. 5).

Attributions of Charitability

Why do people admire gifts of charity more when they are anonymous? This judgment, although intuitively compelling, is morally questionable, because a donation does as much good to the beneficiary whether it is public or anonymous, and publicity should incentivize prospective donors to donate more, increasing aggregate welfare.

One can resolve the paradox by noting that ascriptions of charitability are judgments not only of the utilitarian boon to a beneficiary but also of the underlying character of the donor. The way a donor gives provides clues about his or her disposition for generosity, which in turn indicates how valuable the person would be as a cooperative partner (33, 58–60). Some people are transactional altruists, helping to the minimum extent necessary for both parties to come out ahead; others offer more generous terms.

States of mutual knowledge of the donor and beneficiary are relevant to attributions of charitability because they govern the donor's expectations of payback (in esteem or quid pro quo reciprocation) and hence his disposition for generosity. A donor who gives publicly may appear less charitable, particularly when the gift is common knowledge with the beneficiary (each aware of the other's awareness of the donation), because the common knowledge cements the beneficiary's obligation to repay the favor, similar to the way that legal contracts create common knowledge of the obligations between a creditor and debtor (61). In contrast, an anonymous or double-blind gift creates no obligations, because the beneficiary does not know the donor, and so speaks to the donor's generosity. We posited that intermediate states of knowledge—a donor or beneficiary knowing the other's identity, but not vice versa—would elicit intermediate judgments.

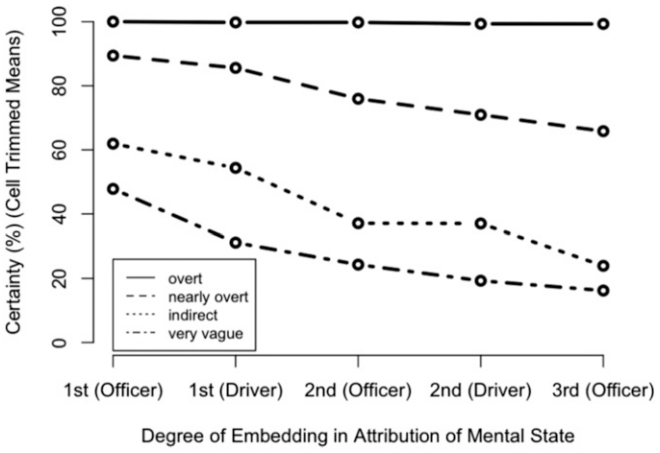

Fig. 4. Participants' ratings of certainty about the intent of a direct proposition (top line) and veiled propositions (bottom 3 lines) on the part of the hearer, the speaker guessing about the hearer's certainty, the hearer guessing about the speaker's guess of their certainty, and so on (*x* axis; labels refer to the scenario in which a driver bribes a police officer). The estimate for each complex proposition is conditioned on acceptance of the propositions embedded within it. Reprinted with permission from ref. 50.

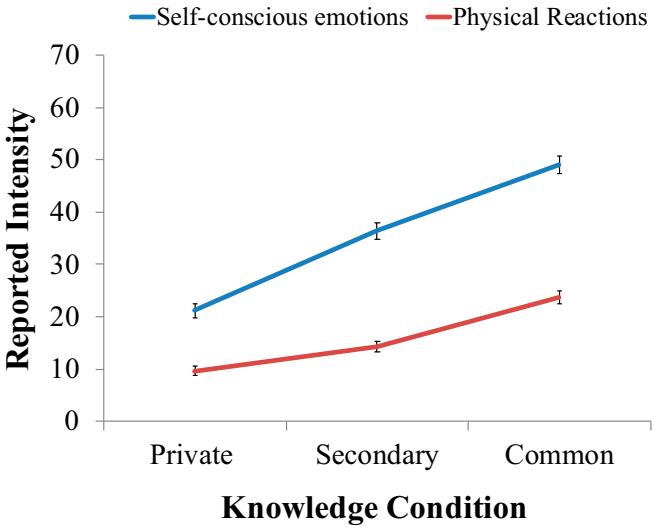

Fig. 5. Average reported self-conscious emotions (embarrassment, guilt, and shame) and physical reactions following an imagined transgression with different degrees of knowledge with observers. Reprinted from ref. 57, with permission from Elsevier.

The predictions were inspired by the medieval scholar Maimonides' "Ladder of Charity," in which forms of charity are ranked in praiseworthiness depending on how the donor generates knowledge of his gift: double-blind gifts are more charitable than single-blind, which are more charitable than public, common knowledge gifts.

Participants read about a donor who gave money to a family in need in one of 3 ways: 1) double-blind, 2) with an exchange of photos (generating common knowledge), or 3) with the donor and beneficiary providing photos to an intermediary, from whom each party could obtain the other's photo confidentially (an option that each in fact exercised), allowing shared but not common knowledge. As predicted, donors who gave double-blind were rated as more charitable than donors who gave with shared knowledge, who were in turn rated as more charitable than donors who gave with common knowledge (Fig. 6). Crucially, in these last 2 cases the donor and beneficiary knew who the other was; what differed was whether each knew that the other knew. This supports the hypothesis that people judge the charitable disposition of donors by tracking the knowledge of the donor, the beneficiary, and third-party observers. In particular, a donor who gives with common knowledge accrues both costs and benefits: on the one hand, he can call in the favor; on the other hand, he forgoes some of the reputational advantages of the donation precisely because he enjoys that perquisite.

Moral Condemnation

Attributions of charitability are just one example of moral judgments that do not track the benefits of an action to other parties. Another example is the way that people value the nature of an action (deontological morality) over its consequences (utilitarian morality) (62). For example, people often judge that stealing, killing, and lying are wrong even when they have net benefits, such as stealing medicine to save someone's life, assisting the suicide of a suffering terminal patient, or comforting a person on her deathbed with an insincere posthumous promise (63). People also enforce morally dubious prohibitions that subtract from human welfare, such as those against contraception, same-sex marriage, and genetically modified foods. Yet another example is condemning a commission more than an omission with the same effect, such as greater censure of starving someone by taking away their food than of refraining from feeding a person who is starving (64–66).

The psychological appeal of deontological morality is a puzzle because it is not an obvious solution to the problem of attaining the benefits of cooperation (67). Why do people gravitate to rules that single out actions, like "Do not lie, kill, or steal" rather than rules that focus on outcomes, such as "Minimize deaths, injuries, and unhappiness"? Deontological rules, although common in moral decisions, rarely appear in matters of prudence, taste, or economics. Heart surgery, for example, is hazardous, but humans do not avoid the danger by committing to a deontological rule like "Never get heart surgery." Here they are consequentialist, using the frequencies of past successes to estimate the costs and benefits.

The logic of coordination applied to the formation of alliances might illuminate the appeal of deontological moralizing. Humans compete over resources, power, and prestige, and form alliances to help one another in such conflicts (62). But while joining a coalition promises advantages, such as spreading the risks and dividing up the spoils, it also brings dangers, such as being exploited by the dominant member, becoming entangled in someone else's fight, finding oneself on the losing side, or stoking a destructive conflict between large, evenly matched coalitions. To avoid these costs, bystanders to a potential conflict can try to join a majority coalition, which requires finding themselves all on the same side, whichever side that is. Moral condemnation offers a solution to this problem. If a community shares a set of acts that they find antisocial, nonconforming, or disgusting, then bystanders can coalesce into a winning coalition via their shared opposition to an offender who commits one.

But to coordinate successfully, bystanders need common knowledge of who is "in the wrong," which cannot be automatic given the countless ways that one person can offend others. Deontological morality offers the advantage of defining public signals for coordinated disapproval—moral focal points—in the form of discrete acts, such as violence, theft, deception, betrayal, and nonconformist sexual or dietary practices (67). Utilitarian welfare affords less possibility for consensus, because bystanders

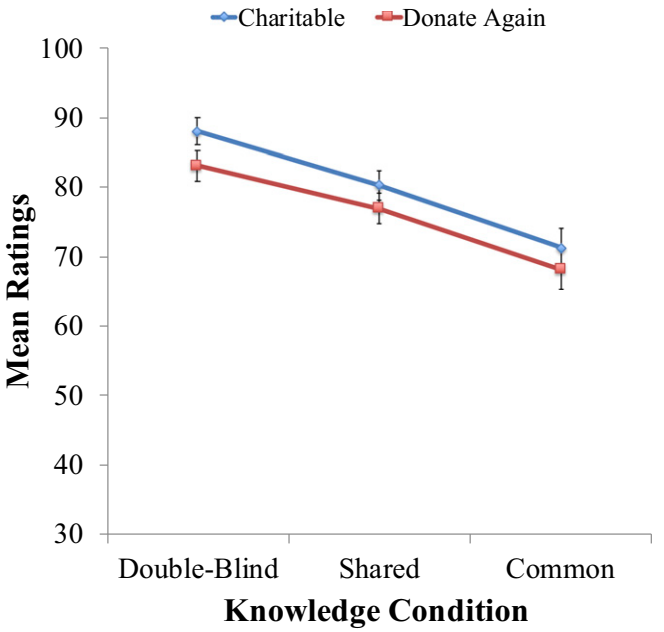

Fig. 6. Average ratings of donor charitability and likelihood of donating again, depending on the state of knowledge between donor and beneficiary. Reprinted with permission from ref. 61.

may differ in whose welfare, which kinds, and which time frames matter most.

We suggest that this may also explain many of the blind spots of human moralizing. People show less outrage at omissions and indirect offenses, even when their foreseeable harms are no less certain, because without a conspicuous action, the violations do not generate the common knowledge necessary to coordinate condemnation (68, 69). Moreover, humans are ingenious in devising victimless taboos and trivial outrages whose main function is to single out miscreants for a snowballing mob to denounce, each denouncer joining to avoid becoming the denounced (20, 70). Today, moral panics are accelerated by social media, which can rapidly generate common knowledge among millions of people. These moral perversities may be explained as a strategy for enjoying the benefits and avoiding the costs of conflict amid dynamically shifting alliances.

Outstanding Questions

The perspective we have presented raises a number of questions for future research:

What are the perceptual cues that give rise to common knowledge?

Can physical displays, like eye contact, blushing, crying, laughing, and facial expressions, which are jointly conspicuous to the expresser and perceiver, be explained as common knowledge generators?

Can laughter in particular be explained as an involuntary public signal of private belief that can deflate an unpopular norm maintained by false common knowledge (also known as pluralistic ignorance, in which a norm persists because everyone mistakenly believes that everyone else values it) (70)?

Which everyday intuitions (e.g., something being "out there") correspond to common knowledge?

What kinds of public displays (posters, billboards, pamphlets, demonstrations, sit-ins, speeches) are perceived as generating common (as opposed to merely shared) knowledge?

Do different forms of electronic communication differ in their ability to generate common knowledge, such as telephones, broadcast and cable television, emails, cc'd emails, bcc'd emails, email discussion lists, tweets, retweets, Facebook postings, blog posts, comments, and viral videos and memes?

Can the phenomenon of electronic mobbing and shaming be explained by the ability of social media to generate the common knowledge that underlies coordinated moral condemnation within a clique or faction?

How intuitive or reflective are people's inferences about common knowledge?

When do young children begin to represent common knowledge and use it for coordination?

Which other animal species distinguish common knowledge from private and shared beliefs? In nonhuman animals, can eye contact and vocal calls in threatening, mating, and predation be explained as common-knowledge generators in coordination games?

Does variation and impairment in the ability to mentalize (as in the autism spectrum) also affect the ability to detect common knowledge, or is common knowledge ascertained by different and perhaps simpler skills?

Conclusion: The Strategic Logic of Social Interactions Shapes Human Social Psychology

We have presented 6 puzzles of social life and identified a phenomenon underlying them all: a sensitivity to common knowledge. Why should this logical distinction matter in so many social domains? We have suggested that insight may be found in the game theory of coordination dilemmas.

As a highly social animal, humans have many opportunities to coordinate for mutual gain. Natural selection may have favored cognitive abilities that improved the ability to coordinate, just as it favors the ability to execute other fitness-critical social tasks, such as nurturing kin and trading favors. Game theory identifies the payoff structure that makes a social interaction a coordination dilemma, and implies that humans can most effectively coordinate if they are able to distinguish common knowledge from private and shared knowledge. It thereby ties together diverse social situations that conform to the logic of coordination, and predicts that in all of them we will find that people are attuned to the cues of common knowledge. The overall result is the distinctively human obsession with publicity, privacy, salience, secrecy, outrage, shame, hypocrisy, discretion, and taboo.

1. A. M. Colman, B. D. Pulford, C. L. Lawrence, Explaining strategic coordination: Cognitive hierarchy theory, strong Stackelberg reasoning, and team reasoning. *Decision (Wash. D.C.)* **1**, 35–58 (2014).
2. O. S. Curry, D. A. Mullins, H. Whitehouse, Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Curr. Anthropol.* **60**, 47–69 (2019).
3. Y. N. Harari, *Sapiens: A Brief History of Humankind* (Penguin Random House, New York, 2014).
4. M. A. Nowak, Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
5. T. C. Schelling, Bargaining, communication, and limited war. *J. Conflict Resolut.* **1**, 19–36 (1957).
6. D. Lewis, *Convention: A Philosophical Study* (Harvard University Press, Cambridge, MA, 2008).
7. M. Gilbert, Rationality and salience. *Philos. Stud.* **57**, 61–77 (1989).
8. R. Sugden, Thinking as a team: Towards an explanation of nonselfish behavior. *Soc. Philos. Policy* **10**, 69–89 (1993).
9. N. A. Dalkiran, M. Hoffman, R. Paturi, D. Ricketts, A. Vattani, "Common knowledge and state-dependent equilibria" in *Proceedings of the Symposium on Algorithm Game Theory*, M. Serna, Ed. (Springer, New York, 2012), pp. 84–95.
10. D. Monderer, D. Samet, Approximating common knowledge with common beliefs. *Games Econ. Behav.* **1**, 170–190 (1989).
11. K. A. Thomas, P. DeScioli, O. S. Haque, S. Pinker, The psychology of coordination and common knowledge. *J. Pers. Soc. Psychol.* **107**, 657–676 (2014).
12. R. J. Aumann, Agreeing to disagree. *Ann. Stat.* **4**, 1236–1239 (1976).
13. N. Eilan, C. Hoerl, T. McCormack, J. Roessler, *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology* (Oxford University Press, New York, 2005).
14. J. Geanakoplos, Common knowledge. *J. Econ. Perspect.* **6**, 53–82 (1992).
15. A. Rubinstein, The electronic mail game: Strategic behavior under "almost common knowledge". *Am. Econ. Rev.* **79**, 385–391 (1989).
16. H. H. Clark, *Arenas of Language Use* (University of Chicago Press, Chicago, 1992).
17. H. H. Clark, *Using Language* (Cambridge University Press, New York, 1996).
18. N. V. Smith, *Mutual Knowledge* (Academic Press, Cambridge, MA, 1982).
19. S. Pinker, *The Stuff of Thought: Language as a Window into Human Nature* (Viking, New York, 2007).
20. R. Willer, K. Kuwabara, M. W. Macy, The false enforcement of unpopular norms. *Am. J. Sociol.* **115**, 451–490 (2009).
21. M. S.-Y. Chwe, *Rational Ritual: Culture, Coordination, and Common Knowledge* (Princeton University Press, Princeton, NJ, 2013).
22. A. Baltag, L. S. Moss, S. Solecki, "The logic of public announcements, common knowledge, and private suspicions" in *Readings in Formal Epistemology*, H. Arló-Costa, V. F. Hendricks, J. van Benthem, Eds. (Springer, Cham, Switzerland, 2016), pp. 773–812.
23. J. Y. Halpern, Y. Moses, Knowledge and common knowledge in a distributed environment. *J. Assoc. Comput. Mach.* **37**, 549–587 (1990).
24. C. Frith, U. Frith, Theory of mind. *Curr. Biol.* **15**, R644–R646 (2005).
25. S. Nichols, S. P. Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Oxford University Press, New York, 2003).
26. R. Saxe, L. Young, "Theory of Mind: How brains think about thoughts" in *The Handbook of Cognitive Neuroscience*, K. Ochsner, S. Kosslyn, Eds. (Oxford University Press, New York, 2013), pp. 204–213.
27. S. Pinker, *The Language Instinct: How the Mind Creates Language* (William Morrow and Company, New York, 1994).
28. O. Curry, M. J. Chesters, 'Putting ourselves in the other fellow's shoes': The role of 'theory of mind'in solving coordination problems. *J. Cogn. Cult.* **12**, 147–159 (2012).
29. P. Kinderman, R. Dunbar, R. P. Bentall, Theory-of-mind deficits and causal attributions. *Br. J. Psychol.* **89**, 191–204 (1998).
30. J. Perner, H. Wimmer, "John thinks that Mary thinks that…" attribution of second-order beliefs by 5-to 10-year-old children. *J. Exp. Child Psychol.* **39**, 437–471 (1985).

31. Academian, Unrolling social metacognition: Three levels of meta are not enough. Lesswrong, August 25. (2018). https://www.lesswrong.com/posts/K4eDzqS2rbcBDsCLZ/unrolling-social-metacognition-three-levels-of-meta-are-not. Accessed 30 March, 2019.
32. D. A. Sabien, Common knowledge and miasma. Medium, August 25. (2018). https://medium.com/@ThingMaker/common-knowledge-and-miasma-20d0076f9c8e. Accessed 30 March 2019.
33. N. Baumard, J.-B. André, D. Sperber, A mutualistic approach to morality: The evolution of fairness by partner choice. *Behav. Brain Sci.* **36**, 59–78 (2013).
34. J. Mehta, C. Starmer, R. Sugden, Focal points in pure coordination games: An experimental investigation. *Theory Decis.* **36**, 163–185 (1994).
35. J. Mehta, C. Starmer, R. Sugden, The nature of salience: An experimental investigation of pure coordination games. *Am. Econ. Rev.* **84**, 658–673 (1994).
36. N. Bardsley, J. Mehta, C. Starmer, R. Sugden, Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Econ. J. (Lond.)* **120**, 40–79 (2009).
37. N. Bardsley, A. Ule, Focal points revisited: Team reasoning, the principle of insufficient reason and cognitive hierarchy theory. *J. Econ. Behav. Organ.* **133**, 74–86 (2017).
38. D. J. Butler, A choice for 'me' or for 'us'? Using we-reasoning to predict cooperation and coordination in games. *Theory Decis.* **73**, 53–76 (2012).
39. A. M. Colman, N. Gold, Team reasoning: Solving the puzzle of coordination. *Psychon. Bull. Rev.* **25**, 1770–1783 (2018).
40. G. A. Miller, P. E. Nicely, An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* **27**, 338–352 (1955).
41. J. M. Darley,B. Latané, Bystander intervention in emergencies: Diffusion of responsibility. *J. Pers. Soc. Psychol.* **8**, 377–383 (1968).
42. B. Latané, J. M. Darley, Group inhibition of bystander intervention in emergencies. *J. Pers. Soc. Psychol.* **10**, 215–221 (1968).
43. B. Latané, J. M. Darley, *The Unresponsive Bystander: Why Doesn't He Help?* (Appleton-Century-Croft, New York, 1970).
44. B. Latané, S. Nida, Ten years of research on group size and helping. *Psychol. Bull.* **89**, 308–324 (1981).
45. P. Fischer *et al*., The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychol. Bull.* **137**, 517–537 (2011).
46. A. Diekmann, Volunteer's dilemma. *J. Conflict Resolut.* **29**, 605–610 (1985).
47. A. Diekmann, "Volunteer's dilemma: A social trap without a dominant strategy and some empirical results" in *Paradoxical Effects of Social Behavior: Essays in Honor of Anatol Rapoport*, A. Diekmann, P. Mitter, Eds. (Physica-Verlag, Heidelberg, 1986), pp. 187–197.
48. K. A. Thomas, J. De Freitas, P. DeScioli, S. Pinker, Recursive mentalizing and common knowledge in the bystander effect. *J. Exp. Psychol. Gen.* **145**, 621–629 (2016).
49. S. Pinker, M. A. Nowak, J. J. Lee, The logic of indirect speech. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 833–838 (2008).
50. J. J. Lee, S. Pinker, Rationales for indirect speech: The theory of the strategic speaker. *Psychol. Rev.* **117**, 785–807 (2010).
51. A. P. Fiske, The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychol. Rev.* **99**, 689–723 (1992).
52. N. Haslam, "Research on the relational models: An overview"in *Relational Models Theory: A Contemporary Overview*, N. Haslam, Ed. (Erlbaum Associates, Mahwah, NJ, 2004), pp. 27–57.
53. J. S. Beer, E. A. Heerey, D. Keltner, D. Scabini, R. T. Knight, The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *J. Pers. Soc. Psychol.* **85**, 594–604 (2003).
54. D. Keltner, B. N. Buswell, Embarrassment: Its distinct form and appeasement functions. *Psychol. Bull.* **122**, 250–270 (1997).
55. D. Sznycer *et al*., Shame closely tracks the threat of devaluation by others, even across cultures. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2625–2630 (2016).
56. J. P. E. Tangney, J. L. Tracy, "Self-conscious emotions" in *Handbook of Self and Identity*, M. Leary, J. Tangney, Eds. (Guilford Press, New York, ed. 2, 2012), pp. 446–478.
57. K. A. Thomas, P. DeScioli, S. Pinker, Common knowledge, coordination, and the logic of self-conscious emotions. *Evol. Hum. Behav.* **39**, 179–190 (2018).
58. P. Barclay, Biological markets and the effects of partner choice on cooperation and friendship. *Curr. Opin. Psychol.* **7**, 33–38 (2016).
59. R. Noë, P. Hammerstein, Biological markets. *Trends Ecol. Evol. (Amst.)* **10**, 336–339 (1995).
60. R. L. Trivers The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
61. J. De Freitas, P. DeScioli, K. A. Thomas, S. Pinker, Maimonides' ladder: States of mutual knowledge and the perception of charitability. *J. Exp. Psychol. Gen.* **148**, 158–173 (2019).
62. P. DeScioli, R. Kurzban, A solution to the mysteries of morality. *Psychol. Bull.* **139**, 477–496 (2013).
63. J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (Vintage, New York, 2012).
64. J. Baron, Nonconsequentialist decisions. *Behav. Brain Sci.* **17**, 1–10 (1994).
65. P. DeScioli, J. Christner, R. Kurzban, The omission strategy. *Psychol. Sci.* **22**, 442–446 (2011).
66. P. Singer, *The Life You Can Save: How to Do Your Part to End World Poverty* (Penguin Random House, New York, 2010).
67. P. DeScioli, R. Kurzban, Mysteries of morality. *Cognition* **112**, 281–299 (2009).
68. F. Cushman, L. Young, M. Hauser, The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychol. Sci.* **17**, 1082–1089 (2006).
69. P. DeScioli, R. Bruening, R. Kurzban, The omission effect in moral cognition: Toward a functional explanation. *Evol. Hum. Behav.* **32**, 204–215 (2011).
70. D. Centola, R. Willer, M. Macy, The emperor's dilemma: A computational model of self-enforcing norms. *Am. J. Sociol.* **110**, 1009–1040 (2005).