



Research

Cite this article: Bear A, Kagan A, Rand DG. 2017 Co-evolution of cooperation and cognition: the impact of imperfect deliberation and context-sensitive intuition. *Proc. R. Soc. B* **284**: 20162326.
<http://dx.doi.org/10.1098/rspsb.2016.2326>

Received: 1 November 2016

Accepted: 14 February 2017

Subject Category:

Evolution

Subject Areas:

evolution, behaviour

Keywords:

cooperation, evolutionary game theory, dual-process

Author for correspondence:

Adam Bear

e-mail: adam.bear@yale.edu

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3703627>.

Co-evolution of cooperation and cognition: the impact of imperfect deliberation and context-sensitive intuition

Adam Bear¹, Ari Kagan¹ and David G. Rand^{1,2,3}

¹Psychology Department, ²Economics Department, and ³School of Management, Yale University, New Haven, CT 06511, USA

AB, 0000-0002-8714-9144

How does cognitive sophistication impact cooperation? We explore this question using a model of the co-evolution of cooperation and cognition. In our model, agents confront social dilemmas and coordination games, and make decisions using intuition or deliberation. Intuition is automatic and effortless, but relatively (although not necessarily completely) insensitive to context. Deliberation, conversely, is costly but relatively (although not necessarily perfectly) sensitive to context. We find that regardless of the sensitivity of intuition and imperfection of deliberation, deliberating undermines cooperation in social dilemmas, whereas deliberating can increase cooperation in coordination games if intuition is sufficiently sensitive. Furthermore, when coordination games are sufficiently likely, selection favours a strategy whose intuitive response *ignores* the contextual cues available and cooperates across contexts. Thus, we see how simple cognition can arise from active selection for simplicity, rather than just be forced to be simple due to cognitive constraints. Finally, we find that when deliberation is imperfect, the favoured strategy increases cooperation in social dilemmas (as a result of reducing deliberation) as the benefit of cooperation to the recipient increases.

1. Introduction

Cooperation is a pervasive feature of human behaviour, yet its adaptive origins pose an evolutionary puzzle. Why would ‘selfish’ evolution lead people to pay personal costs to benefit non-kin? A wide variety of game-theoretic research over the past several decades has provided numerous answers to this question [1–5], many of which involve demonstrating how cooperating can actually be in one’s long-term self-interest because of mechanisms such as direct or indirect reciprocity and signalling (i.e. cooperation can be ‘strategic’) [6–15].

Nevertheless, empirical study of humans’ actual behaviour consistently finds that people cooperate even in contexts when these mechanisms are absent and cooperation is not in one’s material self-interest (e.g. in social dilemmas where the interaction is anonymous and one-shot; for a review, see [16]). To explain ‘pure’ cooperation of this sort, recent research has turned to the proximate psychological mechanisms underlying cooperative behaviour. One prominent strain of this research has adopted a dual-process framework, distinguishing intuitive and deliberative cognitive processes [17,18]. Relatively speaking, intuitive processes are conceptualized as cognitively cheap, fast and automatic, but inflexible (i.e. insensitive to details of one’s particular circumstances); whereas deliberative processes are conceptualized as cognitively expensive, slow and controlled, but flexible (i.e. sensitive to these details) [19–21].

Viewed through this dual-process lens, people’s ‘irrational’ cooperative behaviour in social dilemmas can be explained as resulting from a ‘social heuristic’—an intuitive (and, therefore, inflexible) judgement that typically

does what is adaptive, but occasionally makes errors [17,22–25]. If cooperation is *typically* long-run payoff maximizing, because of mechanisms such as reciprocity and signalling, it can be advantageous to (sometimes) avoid the cost of deliberating about the specific situation one is facing and instead just intuitively cooperate. Importantly, this cost need not be intrinsic to deliberation itself (as stipulated, e.g. by the theory of ego depletion [26]), but can also be operationalized as an opportunity cost that arises from the extra time and attentional demands that come from deliberating, which limit one's ability to engage in other potentially rewarding tasks (see [27]).

This argument has been articulated in a specifically dual-process way by the 'Social Heuristics Hypothesis' (SHH) [22; formalized in [25]). The SHH predicts that people should be more cooperative in social dilemmas when they are induced to rely more on intuitive cognitive processing relative to deliberative cognitive processing—a prediction that has garnered considerable empirical support, as shown by a recent large-scale meta-analysis [28]. Moreover, in further support of the SHH, this meta-analysis found no effect of intuitive versus deliberative cognitive processing on cooperation in coordination contexts in which cooperation is payoff-maximizing (e.g. repeated interactions) and thus it could be in one's self-interest to deliberately cooperate.

Implicit in this theory, however, is an extreme dichotomy between intuitive and deliberative cognition: deliberation is assumed to be perfectly flexible, whereas intuition is assumed to be completely insensitive to the strategic nature of the situation one is facing [22,25]. While this extreme schema may apply to many situations in the real world, there is good reason to believe that it is too extreme in other cases.

With regard to deliberation, there are two primary ways in which this type of cognitive process could sometimes make mistakes and fail to accurately tailor its behaviour to context. First, sometimes the information available to an individual is simply limited. For example, it is sometimes unknown whether you will see a person again and therefore suffer consequences for defecting on them. Thus, even if a person carefully deliberates and analyses all of the information she has available to her in this situation, she may nevertheless come to the wrong conclusion about whether it is worth cooperating. Similarly, a person might get confused while deliberating and reach the wrong conclusion despite having full information available (e.g. experimental participants who mistakenly conclude that mutual cooperation is the highest payoff outcome in a social dilemma [29]). Second, deliberation itself sometimes merely helps people rationalize their intuitive desires, in which case deliberation will merely commit the same mistakes that intuition would [30].

With regard to intuition, an entire research programme separate from the heuristics and biases programme has emphasized how heuristics can be 'fast and frugal,' picking up on salient cues which can inform decision-making [31,32]. In other words, in many settings, intuition may prove quite flexible and accurate, even sometimes surpassing the capacities of deliberation. This may especially be true in domains in which a person has had considerable prior experience or in which there are learnable statistical regularities that distinguish one context from another (e.g. whether or not a person looks like they are from one's in-group).

In this paper, we use a formal evolutionary game-theoretic model to explore the consequences of imperfect

deliberation and flexible intuition on the coevolution of cooperation and cognition. We find that adding imperfect deliberation has little qualitative effect on the evolutionary dynamics. Allowing intuition to be somewhat sensitive, on the other hand, can lead natural selection to favour a qualitatively different type of strategy—one for which deliberation increases cooperation in settings where cooperating can be payoff-maximizing. These findings extend our understanding of the cognitive underpinnings of human decision-making, and generate new, testable empirical predictions.

2. Model

We begin with a basic model framework (adapted from [25]) in which agents are placed in an environment consisting of a mix of social dilemmas (which occur with probability $1 - p$) and cooperative coordination games (which occur with probability p).

In the social dilemmas, agents can cooperate by paying a cost c to give a benefit b to their partner, or they can defect by doing nothing. Thus, in this type of situation, it is always payoff-maximizing to defect. By contrast, it can be payoff-maximizing to cooperate in the coordination games, which represent cooperative interactions involving future consequences (e.g. arising from repeated interactions, reputation effects, signalling or sanctions). When exploitation occurs (i.e. when one player defects while the other cooperates), the possibility of future consequences has two implications. First, it reduces the benefit to the defector (due to, e.g. lost future cooperation, damaged reputation or material punishment). Second, it reduces the cost to the cooperator (owing to, e.g. switching to defection, improved reputation or material rewards). As a result, it becomes payoff-maximizing to cooperate if one's partner also cooperates (thereby creating a coordination structure). For simplicity, we focus on the limiting case where when one player cooperates and the other defects, both receive zero payoffs.

Within this environment, we examine the evolution of agents with dual-process cognition. We do so by allowing agents to decide to cooperate or defect on the basis of one of two types of cognitive process: intuition or deliberation. Deliberation is more sensitive to game type than intuition (as described below), but is also costlier. Specifically, in the real world, deliberation typically takes more time than intuition, requires more cognitive resources (e.g. those involved in flexible domain-general reasoning), and depends on inhibition of more automatic forms of cognition (e.g. those that govern emotions). Each of these factors can impose material costs on the deliberating agent, for example by causing them to miss out on opportunities to act and by preventing them from devoting cognitive resources to other (non-cooperation related) tasks [27].

Importantly, the magnitude of these costs vary based on the specific context of a particular decision (e.g. how much time the actor has to make a decision or what other tasks the actor had to complete at the same time). Thus, in our model the degree to which deliberation is costlier than intuition varies stochastically from interaction to interaction. Specifically, deliberating in a given interaction requires agents to incur a fitness cost d^* (stochastically sampled for each interaction from the uniform distribution $[0, d]$). Each agent then has a metacognitive threshold T that indicates their maximum willingness-to-pay for deliberation, such that they deliberate in interactions where the deliberation

cost is sufficiently small, $d^* \leq T$, but act intuitively when deliberation is too costly, $d^* > T$. (This means that in any given interaction, an agent with threshold T deliberates with probability T/d and uses intuition with probability $1 - T/d$.)

In prior work [25], intuition was assumed to be completely inflexible to game type, so there was a single intuitive response that governed whether an agent would cooperate or defect across all situations. By contrast, when an agent deliberated, she perfectly tailored her strategy to the game type she faced in the current interaction. Here, we instead implement more nuanced characterizations of intuition and deliberation, allowing each to vary in the degree of sensitivity that they have to detect context (social dilemma versus coordination game). With probability y_d , deliberation accurately infers its current interaction type, and with probability $1 - y_d$ it makes the opposite (incorrect) judgement (i.e. confuses a social dilemma for a coordination game or vice versa). Likewise, with probability y_i , intuition accurately infers its current interaction type, and with probability $1 - y_i$ it makes the opposite judgement. Both sensitivity parameters y_d and y_i range from 0.5 to 1. A value of 0.5 corresponds to total insensitivity to context: the judgement of social dilemma versus coordination game is totally random (correct with probability 0.5). A value of 1, conversely, is perfectly sensitive, making no errors. Thus, the extreme dichotomy between intuitive and deliberative cognition used in prior work is a special case of the model we present here, with totally insensitive intuition ($y_i = 0.5$) and perfectly sensitive deliberation ($y_d = 1$).¹

Therefore, in addition to their metacognitive threshold T , agents have four additional strategy parameters which specify their cooperation choices. When responding using intuition, agents cooperate with probability S_{iD} when intuition indicates they are in a social dilemma and with probability S_{iC} when intuition indicates they are in a coordination game. Conversely, when responding using deliberation, agents cooperate with probability S_{dD} when deliberation indicates they are in a social dilemma and with probability S_{dC} when deliberation indicates they are in a coordination game. Figure 1 summarizes the parametrization of our model.

Our goal is to analyse the evolutionary outcomes of these five quantities (T , S_{iD} , S_{iC} , S_{dD} , S_{dC}), which together specify an agent's genotype. To do so, we determine the set of Strict Nash Equilibria for our model. A given strategy is Strict Nash Equilibrium if, conditional on the other player using that strategy, you would earn a lower payoff by switching to any other strategy. The evolutionary logic behind this condition is that in a population where all agents use such a strategy, any mutant would be at a disadvantage relative to the rest of the population, and thus no new strategies could invade the population.

For parameter sets where multiple equilibria exist, we use the 'risk-dominance' condition to determine which equilibrium will be favoured by natural selection. Equilibrium strategy X risk-dominates equilibrium strategy Y if X earns a higher payoff than Y in a population composed of half X and half Y (this condition implies that X has a larger basin of attraction than Y , and has been shown to characterize evolutionary outcomes [33]).

We present only the results of these equilibrium and risk-dominance calculations in the main text, and give the derivations in the electronic supplementary material (ESM). In the ESM, we also present calculations showing that the results of these equilibrium analyses align with the outcomes of stochastic evolutionary dynamics.

3. Results

(a) Baseline case

We begin by examining the baseline case of completely insensitive intuition ($y_i = 0.5$) and perfectly sensitive deliberation ($y_d = 1$) explored in [25]. In this limiting case, there are two possible equilibrium strategies. Throughout, we use the following naming convention for strategies. The first word ('intuitive' versus 'dual-process') indicates whether the strategy is completely intuitive (i.e. only uses intuition, $T = 0$) or dual-process (i.e. sometimes deliberates, $T \gg 0$). The second word ('defector' versus 'cooperator' versus 'attender') indicates what the intuitive response is—i.e. whether intuition always defects, always cooperates or attends to information about game type (cooperating when intuition perceives a coordination game and defecting when intuition perceives a social dilemma).

The first equilibrium strategy is the intuitive defector (ID), which is an equilibrium for all values of p , b and c . ID intuitively defects regardless of perceived game type ($S_{iC} = S_{iD} = 0$) and never deliberates ($T = 0$). (As ID never deliberates, its deliberative strategy choices S_{dC} and S_{dD} are undefined.)

The second possible equilibrium strategy is the dual-process cooperator (DC), which is an equilibrium when $p > c/b$ (i.e. when coordination games are sufficiently likely). DC intuitively cooperates regardless of how intuition perceives the game type ($S_{iC} = S_{iD} = 1$). When deliberating, DC overrides the intuitively cooperative response in favour of defection when deliberation perceives the game to be a social dilemma ($S_{dD} = 0$), but maintains the intuitively cooperative response when deliberation perceives the game to be a coordination game ($S_{dC} = 1$). Finally, the maximum cost DC is willing to pay is $T = c(1 - p)$. This is the optimal deliberation threshold for DC agents because the benefit for deliberating for DC is the ability to avoid cooperating (and, thus, avoid incurring a cost c) in the fraction $1 - p$ of interactions that are social dilemmas. In other words, $c(1 - p)$ is DC's expected payoff gain from deliberating, and so deliberation is disadvantageous when it is costlier than this value, leading to $T = c(1 - p)$.

A risk-dominance calculation provides the condition for DC to be favoured by selection (see ESM for exact condition). The favoured strategy for the baseline case, as well as the corresponding equilibrium value of T , is shown as a function of p in figure 2*a*. We see that once coordination games become sufficiently likely ($p = 0.3$ for the parameters used in figure 2*a*), DC overtakes ID and T jumps from zero to a high level, $c(1 - p)$. Then as the likelihood of coordination games p increases further, T decreases (as social dilemmas become more rare and therefore deliberation becomes less useful).

(b) Imperfect deliberation

We now examine the effect of weakening deliberation's accuracy at distinguishing social dilemmas from coordination games, such that $0.5 < y_d < 1$. (For tractability, we focus on the case where intuition remains completely insensitive, $y_i = 0.5$; in the ESM, we show that allowing intuition to be sensitive does not qualitatively alter the consequences of imperfect deliberation.) How do errors in deliberation

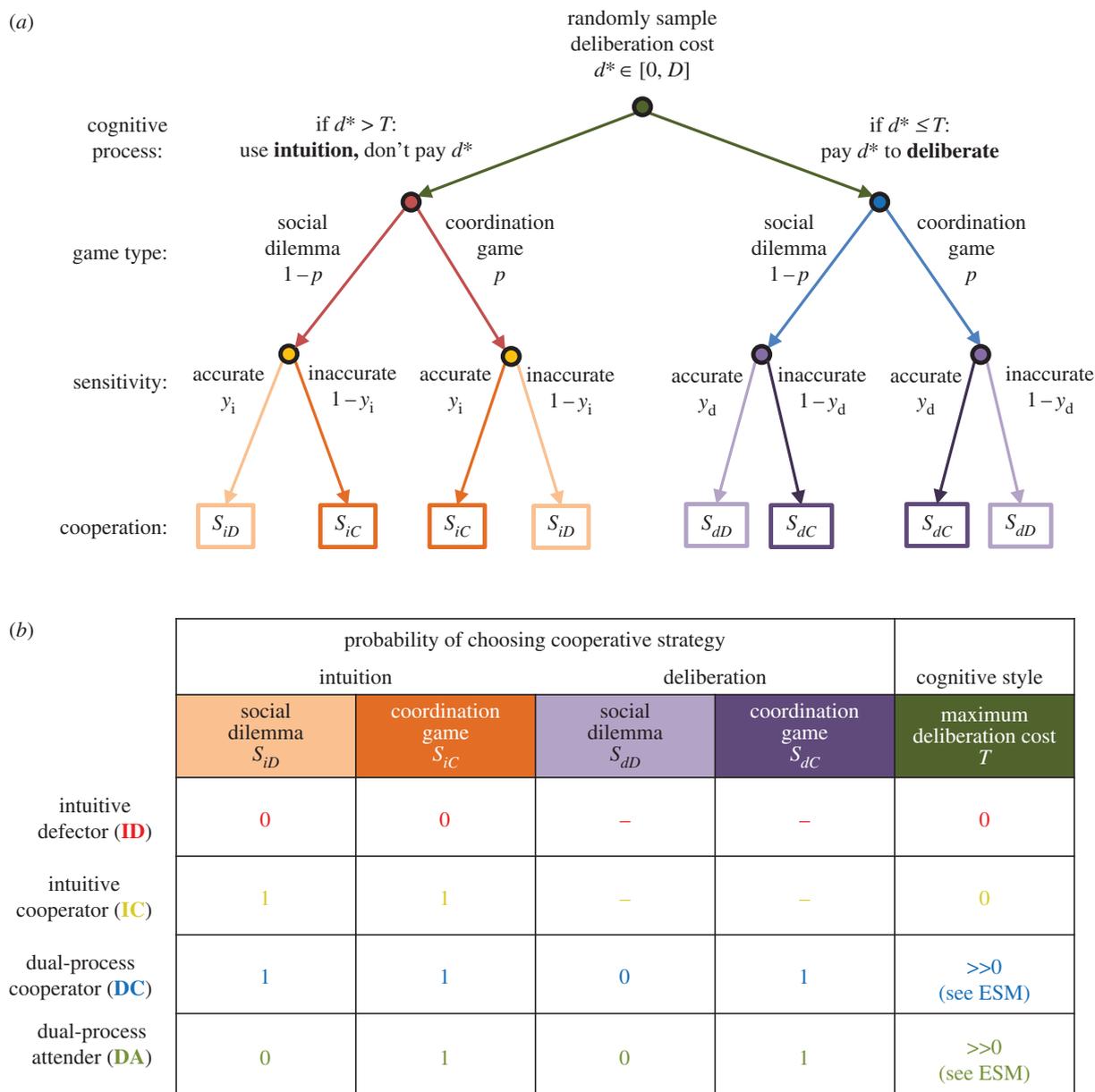


Figure 1. (a) Summary of an agent's decision process in the model. First, a deliberation cost is sampled, and on the basis of this and the agent's maximum willingness to pay to deliberate, the agent either uses intuition or deliberation. The agent then tries to implement her desired cooperation strategy for the given interaction, but sometimes makes errors and uses the opposite intuitive or deliberative strategy, depending on the sensitivity of the cognitive process being used. (b) Summary of the four possible equilibrium strategies discussed in the main text. (Online version in colour.)

influence the strategies that are selected by evolution and the amount of deliberation that agents are willing to engage in?

As in the baseline case, ID and DC are the only two possible equilibria across the entire $0.5 < y_d < 1$ range. The value of y_d does, however, have a substantial quantitative impact on the maximum cost T that DC is willing to pay to deliberate. This optimal T is influenced by y_d in two ways.

First, because deliberation is sometimes inaccurate, it becomes less beneficial to deliberate in order to avoid wasting the cost of cooperation when in social dilemmas. Specifically, in the $1-y_d$ fraction of cases that deliberation erroneously perceives that a social dilemma is a coordination game (and prescribes continued cooperation), nothing is gained by deliberating. As such, the benefit of deliberation from the baseline model $c(1-p)$ is transformed into $c(1-p)y_d$ here—agents only receive the benefits of defection in social dilemmas in the y_d fraction of cases in which deliberation is accurate.

Second, deliberation errors can lead agents to miss out on the benefits of cooperation in coordination games.

Specifically, if a deliberating agent incorrectly perceives a coordination game to be a social dilemma (which happens with probability $1-y_d$), she will defect. Assuming that the other agent avoids making the same error and cooperates (which happens with probability $1-(T/d)(1-y_d)$), the first agent will miss out on the $b-c$ benefit of coordination in the p per cent of interactions that are coordination games.

Putting this all together allows us to calculate the overall expected value of (imperfect) deliberation, and therefore the corresponding optimal deliberation threshold T for the DC strategy. Unlike in the baseline case, however, this expected value can be negative (for example, when y_d is sufficiently small or p is sufficiently large). As an agent's deliberation threshold T is constrained to be non-negative, this means that the equilibrium value of T for the DC strategy is equal to the expected value of deliberation when that value is non-negative, and 0 when it is negative.

In this latter case where $T=0$, the DC agent simply becomes an intuitive cooperator (IC) who is no longer

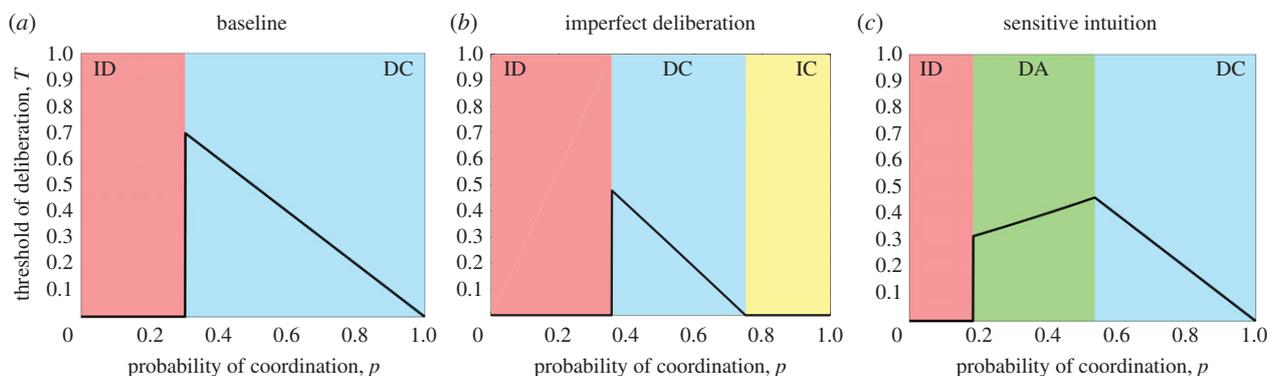


Figure 2. Summary of key results from the model for $b = 4$, $c = 1$ and $d = 1$. Each panel shows the equilibrium strategy and associated threshold of deliberation T , as a function of the probability p that any given interaction is a cooperative coordination game (as opposed to a social dilemma). (a) Results from the baseline model in which deliberation is perfectly sensitive to context ($y_d = 1$) and intuition is completely insensitive to context ($y_i = 0.5$). (b) Results when deliberation is imperfectly sensitive to context ($y_d = 0.9$) and intuition completely insensitive to context ($y_i = 0.5$). (c) Results when intuition is somewhat sensitive to context ($y_i = 0.75$) and deliberation perfectly sensitive to context ($y_d = 1$). (Online version in colour.)

‘dual-process’. For clarity, we refer to the DC equilibrium with $T = 0$ as IC, although this is not technically a different equilibrium from DC (just a special case of the DC equilibrium). The non-deliberative IC strategy always plays the cooperative strategy because the risk of accidentally defecting in a coordination game is too great—which is the same ‘error management’ logic [34,35] used by Delton *et al.* [24] to explain cooperation in one-shot games (an example of a social dilemma). This is illustrated in figure 2b, which shows the favoured strategy and corresponding equilibrium value of T for $y_d = 0.9$, as a function of p . A more general characterization of the evolutionary outcomes over the full $[p \times y_d]$ space is shown in figure 3.

Another interesting difference in DC’s deliberation threshold T between the baseline case and the case with imperfect deliberation is its dependence on b . When deliberation is perfect ($y_d = 1$), the equilibrium value of $T = c(1 - p)$ does not depend on b : because the only consequence of deliberating is being able to avoid the cost of cooperation in social dilemmas, the extent to which cooperation is beneficial is irrelevant. This formulation makes the prediction that, within the region where DC is the risk-dominant equilibrium, making cooperation more beneficial in a social dilemma will have no effect on cooperation.

When $y_d < 1$, however, the equilibrium deliberation threshold T does contain b . Specifically, T decreases as b increases. This is because when deliberation is imperfect, deliberating can be costly by leading to defection in coordination games—and the larger b is, the worse it is to miss out on the benefit of cooperating in coordination games. The consequence of this fact is that as b increases, DC agents become less inclined to deliberate, and therefore become more likely to cooperate in social dilemmas.

In sum, the extent to which deliberation is imperfect does not affect the set of possible equilibrium strategies. Making deliberation inaccurate does, however, (i) reduce the extent to which DC agents deliberate (as one would expect, given that the more imperfect deliberation is, the less useful it is), in the extreme leading to a version of DC which never deliberates and just intuitively cooperates all the time; and (ii) makes it so that DC agents are sensitive to the benefit of cooperation b in social dilemmas, becoming less likely to deliberate (and therefore more likely to cooperate) as b increases.

(c) Sensitive intuition

We next consider how improving intuition’s sensitivity, such that $0.5 < y_i < 1$, affects the evolutionary outcomes. (For tractability, we focus on the case where deliberation is perfect, $y_d = 1$; in the ESM, we show that allowing deliberation to be imperfect does not qualitatively alter the consequences of sensitive intuition.) Recall that, as for deliberation, in our model agents have two intuitive responses S_{iD} and S_{iC} corresponding to separate behaviours when intuition perceives the situation it faces as a social dilemma and coordination game, respectively. So, for example, an agent might defect when intuition judges that the interaction is a social dilemma ($S_{iD} = 0$), but intuitively cooperate when it believes it is a coordination game ($S_{iC} = 1$). Alternatively, agents need not attend to the difference between these two situations: they may choose to always intuitively defect ($S_{iD} = 0$ and $S_{iC} = 0$) or always intuitively cooperate ($S_{iD} = 1$ and $S_{iC} = 1$).

As in the baseline case, we again find that ID and DC are possible equilibria, and DC’s threshold for deliberation is unaffected by having $y_i > 0.5$: it remains $T = c(1 - p)$. Recall that these strategies do not attend to whether their intuition perceives the game they are facing to be a social dilemma or a coordination game—ID always intuitively defects ($S_{iD} = 0$ and $S_{iC} = 0$) and DC always cooperates when using intuition ($S_{iD} = 1$ and $S_{iC} = 1$).

When y_i is sufficiently large, however, we find the emergence of a qualitatively new equilibrium that we will call the ‘dual-process attender’ (DA). DA’s intuitive response *does* discriminate on the basis of interaction type—intuitively defecting in interactions that intuition perceives as social dilemmas ($S_{iD} = 0$) and intuitively cooperating in interactions that intuition perceives as coordination games ($S_{iC} = 1$). Moreover, DA also engages in deliberation ($T \gg 0$; see ESM for full condition). As with DC, a deliberating DA agent cooperates when deliberation indicates it is playing a coordination game and defects when deliberation indicates it is playing a social dilemma, $S_{dD} = 0$ and $S_{dC} = 1$.

The DA strategy thus exhibits a qualitatively new function for deliberation. Although the deliberative responses for DC and DA are the same, the intuitions that deliberation sometimes overrides are different. Unlike DC, which only uses deliberation to override its intuitively cooperative response in social dilemmas (saving c in these instances by defecting), DA uses deliberation to sharpen the accuracy of an intuition

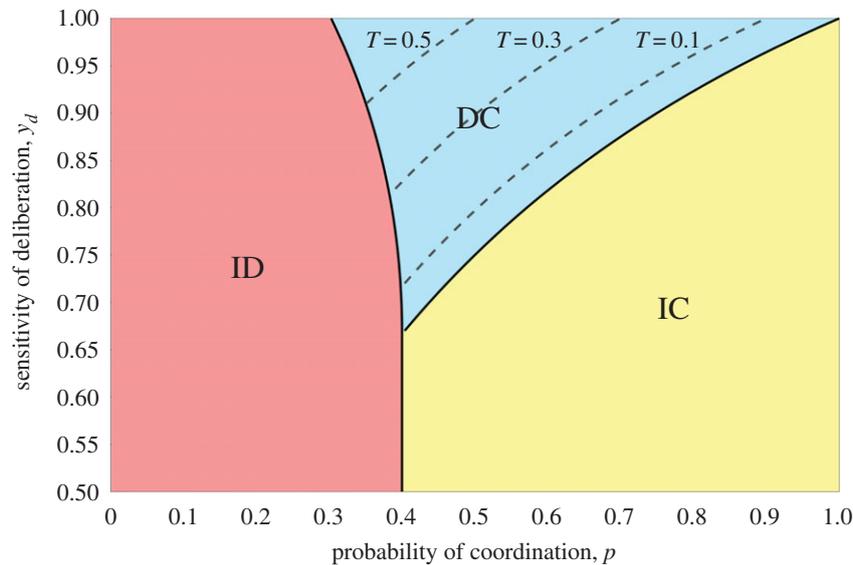


Figure 3. Risk-dominant strategies for $b = 4$, $c = 1$ and $d = 1$ as the probability of cooperative coordination games p and sensitivity of deliberation y_d are varied (with intuition fixed to be fully insensitive to context, $y_i = 0.5$). Dotted lines indicate isodines for values of the deliberation threshold T within the DC region ($T = 0$ everywhere in the ID and IC regions). Note that IC and DC are not technically different equilibria: IC is merely the special case of DC where $T = 0$. (Online version in colour.)

that is already attending to game type. Consequently, the value of deliberation for DA is inversely related to the accuracy of intuition: as the sensitivity of intuition y_i improves, deliberation becomes less and less useful. By contrast, the value of deliberation for DC does not depend on y_i .

The influence of p , the fraction of interactions that are coordination games, on each of these strategies also importantly differs. For DC, because the value of deliberation lessens as there are fewer and fewer social dilemmas, T is strictly decreasing in p . By contrast, for DA, the value of deliberation results from correcting errors in intuition, which can either occur when intuition misperceives a social dilemma to be a coordination game (thereby leading the agent to waste c by accidentally cooperating) or when intuition misperceives a coordination game to be a social dilemma (thereby leading the agent to throw away $b - c$ by accidentally defecting, in cases where the other agent cooperates). As there is an asymmetry in the costliness of these intuitive errors, such that errors in coordination games are costlier than in social dilemmas ($b - c > c$; as per error management theory [24,34,35]), the overall benefit of deliberating increases as coordination games become more common. Thus, the equilibrium value of T is actually increasing in p for DA agents.

An interesting corollary of this result is that deliberation can have positive consequences for cooperation when agents play DA. For both DC agents and DA agents, deliberation undermines cooperation in social dilemmas by overriding intuitive cooperation. For DC agents, this was deliberation's only function, and so deliberation only worked to reduce cooperativeness. For DA agents, however, deliberation can also *promote* cooperation by overriding (mistaken) intuitive defection when in coordination games.

The favoured strategy and corresponding equilibrium value of T is shown as a function of p for $y_i = 0.75$ in figure 2c, and a more general characterization of the evolutionary outcomes over the full $[p \times y_d]$ space is shown in figure 4. Notably, DC (which ignores intuition's sensitivity and always intuitively cooperates) is the favoured equilibrium for high values of p , even when intuition is quite sensitive (y_i is quite large). This indicates that the conclusions of the baseline model are quantitatively unaffected by

sensitive intuition, so long as p is sufficiently large. Furthermore, this illustrates an important point about the adaptive significance of cognitive ability: even if an agent has the cognitive capacity to attend to meaningful differences, it is not necessarily adaptive to do so. This follows from the asymmetry in costs between the errors of defecting in a coordination game and cooperating in a social dilemma. As the former is costlier, it can be adaptive for agents to err on the side of cooperation even if they are reasonably sure they could get away with defecting (as per [24]).

Also of note is the transition, for low values of p , from ID to DA as intuition becomes more sensitive (y_i increases). This transition may seem counterintuitive: improving intuition (increasing y_i) actually *reduces* the extent to which agents rely on intuition, leading a purely intuitive agent (ID) to be outcompeted by an only partially intuitive agent (DA). How could making intuition better lead to a decrease in its use?

The answer lies in considering why ID agents do not deliberate. It is only valuable for intuitively defecting agents to deliberate and switch to cooperation when (i) they are in coordination games and (ii) the other agent *also* deliberates and overrides their selfish response. This dynamics sets up a second-order coordination problem, where each agent's willingness to deliberate is linearly dependent on the other agent's willingness to deliberate. In most parameter regions, agents maximize their payoffs by deliberating slightly less than their partners, leading selection to an equilibrium in which nobody deliberates at all (ID). (See [25] for a detailed discussion of this race to the bottom in deliberativeness.)

The situation changes, however, when intuition becomes sufficiently sensitive. Here, DA becomes the favoured equilibrium, and DA agents cooperate when intuition perceives that they are in a coordination game ($S_{iC} = 1$). Because of this, one's partner may cooperate even if they do not deliberate; and therefore, there is value in deliberating even if one's partner does not. This (at least partially) resolves the second-order coordination problem regarding deliberation, and leads to selection favouring non-zero deliberation. Thus, we see that making intuition more sensitive can, paradoxically, *increase* the value of deliberation for intuitively discriminating

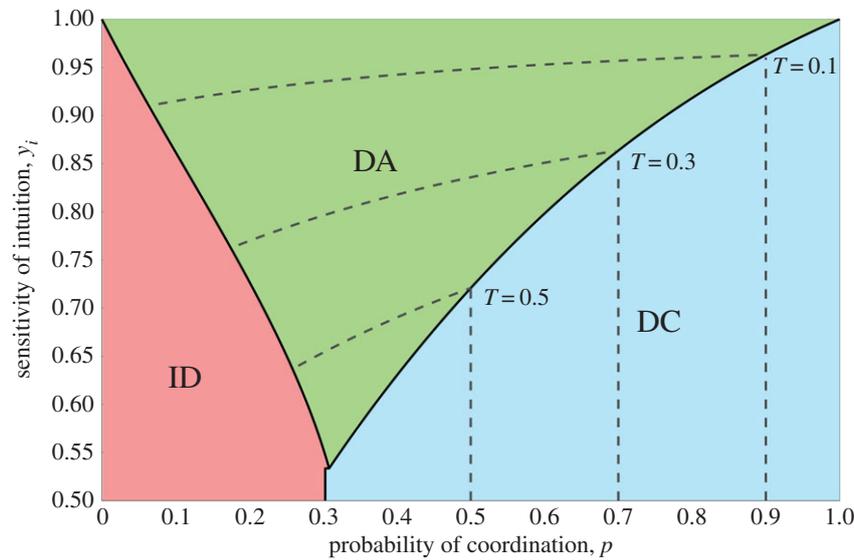


Figure 4. Risk-dominant strategies for $b = 4$, $c = 1$ and $d = 1$ as the probability of cooperative coordination games p and sensitivity of intuition y_i are varied (with deliberation fixed to be perfectly sensitive to context, $y_d = 1$). Dotted lines indicate isoclines for values of the deliberation threshold T ($T = 0$ everywhere in the ID region). (Online version in colour.)

Table 1. Deliberation's effects on cooperation for the four equilibrium strategies.

strategy	social dilemmas	coordination games
intuitive defector (ID)	no effect	no effect
intuitive cooperator (IC)	no effect	no effect
dual-process cooperator (DC)	reduces cooperation	no effect
dual-process attender (DA)	reduces cooperation	increases cooperation

agents—who want to ensure that they do not miss out on the $b - c$ payoff from mutual cooperation in coordination games.

4. Discussion

Here, we have explored the consequences of imperfect deliberation and sensitive intuition for the co-evolution of cognition and cooperation. We find that the following four observations hold regardless of how imperfect deliberation is, and how sensitive intuition is. First, when the probability of coordination games p is sufficiently small, selection will favour a strategy that has a totally insensitive intuition to defect in all situations, and that never deliberates (ID). Second, when p is sufficiently large, selection will favour a strategy whose intuition is totally insensitive and chooses cooperation regardless of the type of interaction it perceives (either DC that does deliberate or IC that does not, depending on how imperfect deliberation is). Third, deliberation only ever reduces cooperation in social dilemmas (for DC and DA) or has no effect (for IC and ID), but never increases social-dilemma cooperation. Fourth, deliberation only ever increases cooperation in coordination games (for DA) or has no effect (for IC, ID and DC), but never decreases coordination game cooperation.

From the first and second observations, we see that selection can favour intuitions that indiscriminately cooperate or defect across interactions, even when intuition is sensitive and therefore discrimination is possible. Although greater intelligence and cognitive complexity is often favoured by selection [36], simplicity can also be adaptive. Even when

intuition can perceive context reasonably well without exerting any cognitive cost, evolution may still select for a simpler strategy that does not condition its behaviour on this information. This result illustrates the importance of studying cognitive processes not only at a proximate level, but also at an ultimate level: sometimes, simple cognition is actively built to be simple by design, and not because of cognitive constraints. In other words, inflexible intuitions that are observed in the real world may have gotten this way through adaptive selection, rather than cognitive compromise.

The third and fourth observations generate clear theoretical predictions about the consequences of experimentally manipulating the use of intuition versus deliberation (see table 1 for a summary). These predictions are borne out by empirical evidence from economic cooperation games. This can be seen in the results of a meta-analysis of 67 such experiments in which cognitive processing was manipulated [28]: in social-dilemma settings, deliberation was found to reduce cooperation, whereas in coordination settings, deliberation had no effect on cooperation.

While these four observations from our model do not depend on the extent of imperfect deliberation and sensitive intuition, other features do, lead to important consequences for the evolutionary outcomes. To start, although weakening deliberation's accuracy does not change the two possible equilibria that can be favoured by selection, it does affect some important features of the extent to which the DC strategy chooses to deliberate.

First, it makes DC's deliberation sensitive to the benefit of cooperation: the more beneficial it is to receive cooperation

from one's partner, the less inclined non-perfect deliberators are to deliberate (out of fear of misperceiving a coordination game as a social dilemma and accidentally defecting; error management, cf. [24,34,35]). As a result, the DC strategy with imperfect deliberation is more likely to cooperate in social dilemmas as the benefit of cooperation increases, a pattern that fits empirical data (e.g. [37]).

Second, imperfect deliberation undermines deliberation's value, reducing the amount of deliberation that DC engages in. Even a fairly modest reduction in deliberation's accuracy can, under the right set of values for the other parameters, lead the DC equilibrium to collapse into a purely IC strategy, which places no value on the discrimination capabilities of deliberation (and thus never deliberates). This result can also be explained by error management: because missing out on the benefits of cooperation in coordination games is costlier than cooperating in social dilemmas, it can be payoff-maximizing to cooperate even when there is good reason to believe that one could get away with defecting. Given this, we might expect deliberation in the real world to be highly tuned to contextual factors since, otherwise, it would be of little use.

We now turn from the imperfection of deliberation to the sensitivity of intuition. As discussed above, selection can favour strategies that ignore sensitive intuition's ability to discriminate. However, once intuition becomes *sufficiently* sensitive, it starts being used in the same discriminating way that deliberation is used. (Recall that DA's intuitive response discriminates between social dilemmas and coordination games, and DA sometimes pays a cost to deliberate in order to improve the accuracy of this discrimination.) The emergence of this qualitatively new dual-process attender equilibrium, which differs from all of the other equilibria in that it makes use of intuition's sensitivity, has important implications for how people's reasoning affects cooperative behaviour in coordination games.

In social dilemmas, as discussed above, empirical work suggests that people become more selfish when they deliberate (for a meta-analysis, see [28]). Both the DC and DA equilibria predict this pattern of results. For DC, deliberation's only function is to override a cooperative response in social dilemmas. For DA, deliberation allows the agent to better tailor her strategy to the situation at hand, which leads her to defect more than intuition would in social dilemmas.

By contrast, DC and DA produce conflicting predictions regarding the role of deliberation in coordination games. DC agents, who always intuitively cooperate, never change their behaviour when learning that an interaction is a coordination game, whereas DA agents sometimes intuitively defect in these situations and thereby become more cooperative if they stop and deliberate.

This is an important result because in many real-world environments, intuition is likely to be reasonably good at discriminating (and thus, our model predicts that DA will be favoured). For example, simple forms of automatic and intuitive cognition are often quite adept at distinguishing ingroup members (likely potential beneficiaries of reciprocity) from outgroup members (unlikely potential beneficiaries of reciprocity) on the basis of cues like skin colour, accent and other salient perceptual features [38]. But stereotyping on the basis of these cues can lead to systematic inefficiencies, where people of different groups or ethnicities become outcasts in a community even when they could be contributing members of society [39,40]. In this class of situations, in which intuition attends to

superficial features, deliberation can actually promote mutually beneficial cooperation by leading people to think more carefully about the situation they are in and whether these cues are valid for everyone's goals at the moment. Though some evidence already speaks to deliberation's role in reducing prejudice (e.g. studies on implicit bias among people who explicitly reject such bias [41]), more research is needed to directly assess this prediction and its dependence on the sensitivity of intuition.

Other domains, however, are not so rich with cues for intuition to pick up on. For example, in most economic game experiments, participants interact anonymously via the computer and thus have no visual or auditory cues regarding their partner's identity. In these situations, intuition should be relatively insensitive, and therefore our model predicts that DC should be favoured—such that the intuitive response favours cooperation regardless of game type. As a result, deliberation should not increase cooperation, as the intuitive and deliberative cooperative responses are perfectly aligned in coordination games (e.g. the repeated Prisoner's Dilemma). This prediction is supported by the experimentally observed lack of effect of deliberation on cooperation in 'strategic' coordination situations described above [28].

Finally, we note some limitations of this work. Only two types of games (social dilemma and coordination game) are considered, rather than a continuum of different possible games; the deliberation cost is sampled from a simple uniform distribution, rather than a more realistic cost distribution; and cognitive processing is unobservable to other agents, such that agents cannot condition on each other's use of intuition versus deliberation (as per [15,42]). Moreover, the flexibility of deliberation and intuition are assumed to be fixed parameters of the environment rather than dynamic features of individuals or populations, which could evolve or interact with other variables. We also do not consider the possibility that deliberation could alter beliefs, which could in turn lead to increased cooperation. Future work could augment the current model to test the robustness of these simplifications and explore how many of the key results about ease–flexibility trade-offs translate into other domains, involving, e.g. social norm compliance, intertemporal choice (as per [43,44]) or punishment.

In sum, by exploring the dynamics of imperfect deliberation and context-sensitive intuition, the model presented here brings us one step closer to linking proximate dual-process cognition with evolutionary game-theory. In doing so, we generate new predictions and help to clarify the ways in which deliberation can, and cannot, work to improve the greater good.

Ethics. This research was conducted without the participation of any human or animal subjects.

Data accessibility. No data were collected in order to conduct the research presented in this manuscript.

Authors' contributions. A.B., A.K. and D.G.R. designed research, performed research, analysed data and wrote the paper.

Competing interests. The authors declare that they have no competing interests.

Funding. The authors gratefully acknowledge funding from the Templeton World Charity Foundation (grant no. TWCF0209) and the Defense Advanced Research Projects Agency NGS2 programme (grant no. D17AC00005).

Acknowledgements. We thank Joshua Greene, Adam Morris, and Jonathan Phillips for helpful feedback on previous drafts.

Endnotes

¹As we are interested in exploring the trade-off between ease and flexibility, we focus on cases where $y_d > y_r$ —i.e. we assume that deliberation is more sensitive than intuition. In the real world, this

condition may not always hold (e.g. for certain fast and frugal heuristics), but we do not consider these cases, as their evolutionary dynamics would be trivial, with selection clearly favouring intuition because there would be no ease–flexibility trade-off between intuition and deliberation.

References

- Nowak MA. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–1563. (doi:10.1126/science.1133755)
- Lehmann L, Bargum K, Reuter M. 2006 An evolutionary analysis of the relationship between spite and altruism. *J. Evol. Biol.* **19**, 1507–1516. (doi:10.1111/j.1420-9101.2006.01128.x)
- West SA, Griffin AS, Gardner A. 2007 Evolutionary explanations for cooperation. *Curr. Biol.* **17**, R661–R672. (doi:10.1016/j.cub.2007.06.004)
- Sigmund K. 2010 *The calculus of selfishness*. Princeton, NJ: Princeton University Press.
- Bowles S, Gintis H. 2003 Origins of human cooperation (ed. P Hammerstein). *Genetic Cult. Evol. Cooperation*, 429–443. Cambridge, MA: MIT Press.
- Axelrod R, Hamilton WD. 1981 The evolution of cooperation. *Science* **211**, 1390–1396. (doi:10.1126/science.7466396)
- Fudenberg D, Maskin ES. 1990 Evolution and cooperation in noisy repeated games. *Am. Econ. Rev.* **80**, 274–279.
- Fudenberg D, Maskin ES. 1986 The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* **54**, 533–554. (doi:10.2307/1911307)
- Nowak MA, Sigmund K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131)
- Gintis H, Smith EA, Bowles S. 2001 Costly signaling and cooperation. *J. Theor. Biol.* **213**, 103–119. (doi:10.1006/jtbi.2001.2406)
- Jordan JJ, Hoffman M, Bloom P, Rand DG. 2016 Third-party punishment as a costly signal of trustworthiness. *Nature* **530**, 473–476. (doi:10.1038/nature16981)
- Roberts G. 2008 Evolution of direct and indirect reciprocity. *Proc. R. Soc. B* **275**, 173–179. (doi:10.1098/rspb.2007.1134)
- Noë R, Hammerstein P. 1994 Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behav. Ecol. Sociobiol.* **35**, 1–11. (doi:10.1007/BF00167053)
- van Veelen M, García J, Rand DG, Nowak MA. 2012 Direct reciprocity in structured populations. *Proc. Natl Acad. Sci. USA* **109**, 9929–9934. (doi:10.1073/pnas.1206694109)
- Hoffman M, Yoeli E, Nowak MA. 2015 Cooperate without looking: why we care what people think and not just what they do. *Proc. Natl Acad. Sci. USA* **112**, 1727–1732. (doi:10.1073/pnas.1417904112)
- Camerer CF. 2003 *Behavioral game theory: experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Rand DG, Greene JD, Nowak MA. 2012 Spontaneous giving and calculated greed. *Nature* **489**, 427–430. (doi:10.1038/nature11467)
- Zaki J, Mitchell JP. 2013 Intuitive prosociality. *Curr. Directions Psychol. Sci.* **22**, 466–470. (doi:10.1177/0963721413492764)
- Sloman SA. 1996 The empirical case for two systems of reasoning. *Psychol. Bull.* **119**, 3. (doi:10.1037/0033-2909.119.1.3)
- Kahneman D. 2003 A perspective on judgment and choice: mapping bounded rationality. *Am. Psychol.* **58**, 697–720. (doi:10.1037/0003-066X.58.9.697)
- Stanovich KE, West RF. 1998 Individual differences in rational thought. *J. Exp. Psychol. Gen.* **127**, 161–188. (doi:10.1037/0096-3445.127.2.161)
- Rand DG, Peysakhovich A, Kraft-Todd GT, Newman GE, Wurzbacher O, Nowak MA, Green JD. 2014 Social heuristics shape intuitive cooperation. *Nat. Commun.* **5**, 3677. (doi:10.1038/ncomms4677)
- Kiyonari T, Tanida S, Yamagishi T. 2000 Social exchange and reciprocity: confusion or a heuristic? *Evol. Hum. Behav.* **21**, 411–427. (doi:10.1016/S1090-5138(00)00055-6)
- Delton AW, Krasnow MM, Cosmides L, Tooby J. 2011 Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc. Natl Acad. Sci. USA* **108**, 13 335–13 340. (doi:10.1073/pnas.1102131108)
- Bear A, Rand DG. 2016 Intuition, deliberation, and the evolution of cooperation. *Proc. Natl Acad. Sci. USA* **113**, 936–941. (doi:10.1073/pnas.1517780113)
- Baumeister RF, Bratslavsky E, Muraven M, Tice DM. 1998 Ego depletion: is the active self a limited resource? *J. Pers. Soc. Psychol.* **74**, 1252. (doi:10.1037/0022-3514.74.5.1252)
- Kurzban R, Duckworth A, Kable JW, Myers J. 2013 An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* **36**, 661–679. (doi:10.1017/S0140525X12003196)
- Rand DG. 2016 Cooperation, fast and slow: meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychol. Sci.* **27**, 1192–1206. (doi:10.1177/0956797616654455)
- Burton-Chellew MN, El Mouden C, West SA. 2016 Conditional cooperation and confusion in public-goods experiments. *Proc. Natl Acad. Sci. USA* **113**, 1291–1296. (doi:10.1073/pnas.1509740113)
- Haidt J. 2001 The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814. (doi:10.1037/0033-295X.108.4.814)
- Gigerenzer G, Goldstein DG. 1996 Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* **103**, 650. (doi:10.1037/0033-295X.103.4.650)
- Gigerenzer G, Todd PM, Group AR. 1999 *Simple heuristics that make us smart*. Oxford, UK: Oxford University Press.
- Kandori M, Mailath GJ, Rob R. 1993 Learning, mutation, and long run equilibria in games. *Econometrica: J. Econ. Soc.* **61**, 29–56. (doi:10.2307/2951777)
- Haselton MG, Buss DM. 2000 Error management theory: a new perspective on biases in cross-sex mind reading. *J. Pers. Soc. Psychol.* **78**, 81. (doi:10.1037/0022-3514.78.1.81)
- Haselton MG, Nettle D. 2006 The paranoid optimist: an integrative evolutionary model of cognitive biases. *Pers. Soc. Psychol. Rev.* **10**, 47–66. (doi:10.1207/s15327957pspr1001_3)
- McNally L, Brown SP, Jackson AL. 2012 Cooperation and the evolution of intelligence. *Proc. R. Soc. B* **279**, 3027–3034. (doi:10.1098/rspb.2012.0206)
- Capraro V, Jordan JJ, Rand DG. 2014 Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Sci. Rep.* **4**, 6790. (doi:10.1038/srep06790)
- Fiske ST. 2002 What we know now about bias and intergroup conflict, the problem of the century. *Curr. Direc. Psychol. Sci.* **11**, 123–128. (doi:10.1111/1467-8721.00183)
- Greene J. 2014 *Moral tribes: emotion, reason and the gap between us and them*. London, UK: Atlantic Books Ltd.
- Stagnaro MN, Dunham Y, Rand DG. 2017 Profit versus prejudice: harnessing self-interest to reduce in-group bias. *Soc. Psychol. Person. Sci.* (doi:10.1177/1948550617699254)
- Greenwald AG, Banaji MR. 1995 Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* **102**, 4. (doi:10.1037/0033-295X.102.1.4)
- Jordan JJ, Hoffman M, Nowak MA, Rand DG. 2016 Uncalculating cooperation is used to signal trustworthiness. *Proc. Natl Acad. Sci. USA* **113**, 8658–8663. (doi:10.1073/pnas.1601280113)
- Tomlin DA, Rand DG, Ludvig E, Cohen JD. 2015 The evolution and devolution of cognitive control: The costs of deliberation in a competitive world. *Sci. Rep.* **5**, 11002.
- Toupo DFP, Strogatz SH, Cohen JD, Rand DG. 2015 Evolutionary game dynamics of controlled and automatic decision-making. *Chaos* **25**, 073120. (doi:10.1063/1.4927488)