Cognitive-attentional mechanisms of cooperation – with implications for attention-deficit hyperactivity disorder and cognitive neuroscience

Nadja R. Ging-Jehli[a,b], L. Eugene Arnold[c], Trish Van Zandt[a]

Affiliations:

[a] Department of Psychology, The Ohio State University

[b] Carney Institute for Brain Science, Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, Providence, RI

[c] Department of Psychiatry and Behavioral Health, The Ohio State University; Nisonger Center UCEDD

ORCIDs:

Nadja R. Ging-Jehli: 0000-0002-1071-0693; L. Eugene Arnold: 0000-0002-0886-0692; T. Van Zandt: 0000-0003-3576-0803

Correspondence:

Correspondence concerning this article should be addressed to Nadja R. Ging-Jehli, PhD (ORCID: 0000-0002-1071-0693), Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, Rhode Island. nadja@gingjehli.com

Citation

Cognitive-attentional mechanisms of cooperation – with implications for attention-deficit hyperactivity disorder and cognitive neuroscience

Abstract

People's cooperativeness depends on many factors, such as their motives, cognition, experiences, and the situation they are in. To date, it is unclear how these factors interact and shape the decision to cooperate. We present a computational account of cooperation that not only provides insights for the design of effective incentive structures but also redefines neglected social-cognitive characteristics associated with attention-deficit hyperactivity disorder (ADHD). Leveraging game theory, we demonstrate that the source and magnitude of conflict between different motives affected the speed and frequency of cooperation. Integrating eye-tracking to measure motivation-based information processing during decision-making shows that participants' visual fixations on the gains of cooperation rather than its costs and risks predicted their cooperativeness on a trial-by-trial basis. Using Bayesian hierarchical modeling, we find that a situation's prosociality and participants' past experience each bias the decision-making process distinctively. ADHD characteristics explain individual differences in responsiveness across contexts, highlighting the clinical importance of experimentally studying reactivity in social interactions. We demonstrate how the use of eye-tracking and computational modeling can be used to experimentally investigate social-cognitive characteristics in clinical populations. We also discuss possible underlying neural mechanisms to be investigated in future studies.

Keywords: cooperation; computational modeling; social cognition; game theory; experimental economics; Bayesian hierarchical modeling; eye-tracking; computational psychiatry; ADHD

Introduction

Cooperation has never been more relevant in a world shaped by climate change, pandemic outbreaks, and armed conflicts. Moreover, the nature of cooperation has considerably changed due to technological advancements, leading to increased strategic interactions not only between humans but also between humans and artificial agents (Agrawal et al., 2019; Camerer, 2019; March, 2021). Studies (Ostrom, 2000; Rand & Nowak, 2013; Skyrms, 2004; Van Lange et al., 2014) examining the factors that determine cooperation show that people are more cooperative if they believe others are too (Capraro et al., 2020; Capraro & Cococcioni, 2015; Rankin et al., 2000); or if they experienced cooperation before (Bolton et al., 2016; Clark & Sefton, 2001; Duffy & Feltovich, 2006). However, how these factors interact and produce cooperation remains unclear. How do situational determinants, past experiences, and cognitive characteristics shape the decision to cooperate? Answering this question helps to understand why a person cooperates in one situation but not in another. Moreover, different situations confront people with distinct trade-offs between the benefits, risks, and costs of cooperation. How much people attend to the gains of cooperation rather than the costs and risks might predict their cooperativeness. Eventually, examining how people adjust their cooperativeness across contexts might generate insights for interventions that promote (or mitigate) cooperation.

Studying how people adjust their cooperativeness across different contexts might also provide psychiatric insights (Ging-Jehli et al., 2021). Specifically, cooperation requires cognitive control to form beliefs about other’s actions and to balance the gains and risks/costs of cooperation (Frank, 2016; Glimcher & Fehr, 2013; Isler, Gächter, et al., 2021; Isler, Yilmaz, et al., 2021; Kocher et al., 2017; Koenigs & Tranel, 2007; Westbrook et al., 2021). Cognitive control is particularly important when situations repeatedly change as they require more frequent adjustments in beliefs and actions (Cools & D’Esposito, 2011; Frank, 2016; Wiecki & Frank, 2013). The efficiency of these control mechanisms have been shown to depend on the dynamics in neuronal components known as the basal ganglia and the frontostriatal pathway (Cavanagh et al., 2011; Cavanagh & Frank, 2014; Doi et al., 2020; Frank, Samanta, et al., 2007; Isoda & Hikosaka, 2008; van Schouwenburg et al., 2014). These neuronal components are deficient in attention-deficit hyperactivity disorder (ADHD) (Cubillo et al., 2012; Frank, Santamaria, et al., 2007; Ging-Jehli et al., 2021; Krain & Castellanos, 2006; Kumar et al., 2022; McLoughlin et al., 2014; Rubia et al., 2011; Ziegler et al., 2016). Importantly, biological models of dopaminergic influences (Frank, 2016; Moolchand et al., 2022) and neuronal networks of corticobasal ganglia circuits (Durstewitz & Seamans, 2008; Frank, Samanta, et al., 2007; Frank & O'Reilly, 2006) show how gating mechanisms in the basal ganglia facilitate selective information processing and more refined behavioral adaptations across contexts (Jaskir & Frank, 2023; Wiecki & Frank, 2013). Since ADHD has been associated with disruption in these neuronal components, we hypothesized that pronounced behavioral change across contexts (referred to as *over-reactivity*) might represent an important but experimentally understudied social-cognitive characteristic of ADHD as already suggested by other research (Ging-Jehli et al., 2021). This could also explain why ADHD is associated with difficulties in developing and maintaining relationships (American Psychiatric Association, 2000; Ginapp et al., 2023). Specifically, dysfunctions in subcortical gating mechanisms might lead to deficits in reward and emotion processing which can lead to exaggerated over-reactive behavior (Champ et al., 2022; Craig et al., 2017; Ma et al., 2017; Nijmeijer et al., 2008; Uekermann et al., 2010). To date, it is unclear whether over-reactivity indexes a hallmark of ADHD; and whether this hallmark is observable and quantifiable using simulated strategic interactions.

Game theory has been used to develop laboratory paradigms (also known as games) to study human cooperation (Rapoport, 1987; Suleiman et al., 2004; Thielmann et al., 2020; Van Lange et al., 2014; Von Neumann & Morgenstern, 1944). In these games, paired participants simultaneously decide whether to cooperate by choosing between cooperative and non-cooperative options, each option being associated with distinct payoffs. The Prisoner's Dilemma (PD; (Poundstone, 1993; Von Neumann & Morgenstern, 1944) and the Stag-Hunt (SH; (Skyrms, 2004; Snidal, 1985; Weber, 2018) are two game types for studying cooperation (Rapoport & Chammah, 1965; Taylor, 1987; van Baar et al., 2022; Van Lange et al., 2014). The PD exemplifies real-world social dilemmas with incongruent incentive structures (Van Lange et al., 2014). That is, each participant has a monetary incentive to exploit the counterpart's cooperativeness by choosing the non-cooperative option. If both participants choose the non-cooperative option, both receive less than if they had both chosen the cooperative option. The SH exemplifies real-world coordination problems with congruent incentive structures (Skyrms, 2004; Snidal, 1985). That is, each participant has an incentive to match their response to that of their counterpart.

PD and SH games both involve *strategic uncertainty*; i.e., the uncertainty about the counterpart's choice (Weber, 2018). However, they introduce distinct sources of *motivational conflict*; i.e., opposing forces that favor the cooperative versus non-cooperative options (Capraro et al., 2020; Goetze, 1994; Snidal, 1985; Weber, 2018). In SH, the tension between cooperation gain and risk arises from strategic uncertainty. In PD, this tension is augmented by cooperation costs and exploitability due to incongruent incentives. Recent research (Capraro et al., 2020; van Baar et al., 2022) suggests that PD games activate mindsets of distrust because incongruent incentives impose conflicts of interest (Bolton et al., 2016; Schul et al., 2004). If this hypothesis is valid, such mindsets should not be activated in SH games due to the congruency of incentives. Hence, different conflict sources might affect the structure of the decision-making process differently. The activation of these different mindsets might also affect how much (overt) attention participants allocate to gains versus risks and costs of cooperation.

How variation in motivational conflict affects the decision-making process of cooperation remains unstudied. Past research focused on one conflict source (mostly PD games) and examined the effect of stress (Alós-Ferrer & Garagnani, 2020; Belloc et al., 2019; Capraro & Cococcioni, 2015; Rand et al., 2014), and/or used summary statistics such as mean reaction times (RTs) or response frequency rates (Evans et al., 2015; Evans & Rand, 2019; Grossmann et al., 2017; Rubinstein, 2007; Spiliopoulos & Ortmann, 2018). Here, we consider a process-oriented account with the diffusion decision model (DDM; Ratcliff, 1978). The DDM is a sequential sampling model that decomposes the decision-making process into distinct mental components (Capra et al., 2020; Forstmann et al., 2016; Ging-Jehli et al., 2021). In these models, choices (and RTs) are produced by the accumulation of evidence for the available response options. The diffusion process describes evidence accumulation toward the cooperative and non-cooperative options (Figure 1a). Sequential sampling models are valuable for their ability to explain behavior from cognitive tasks (Forstmann et al., 2016; Ging-Jehli & Ratcliff, 2020; Logan et al., 2014; Ratcliff & Smith, 2004) and value-based decision tasks similar to the games described above (Busemeyer & Townsend, 1993; Cavanagh et al., 2011; Krajbich et al., 2010; Krajbich & Rangel, 2011; Pedersen & Frank, 2020). They use more information than summary statistics because parameters are estimated by simultaneously considering response frequency rates and the entire RT distributions (Van Zandt & Ratcliff, 1995). Some studies have leveraged sequential sampling models to study altruistic giving or prosocial preferences within one context (Capra et al., 2020; Chen & Krajbich, 2018). In this study, we used the DDM to decompose performance in strategic social settings used to study cooperation.

The DDM involves four main parameters: drift rate (how quickly and in what direction information accumulates), boundary separation (how much information is required for a decision), starting point (how much information is in the system before the accumulation process starts), and nondecision time. These parameters capture different psychological factors leading to RT distributions and choice frequencies. Larger drift rates indicate less motivational conflict, leading to faster (and less skewed) RTs and more frequent cooperative choices. Larger boundary separation represents a more cautious response strategy, resulting in slower (and more skewed) RTs and more cooperative choices. Starting points closer to the cooperative boundary capture people's prejudice to cooperation, resulting in large changes in the tail and leading edge of the RT distributions. Nondecision time subsumes non-decision processes (e.g., perceptual encoding), resulting in shifts of RT distributions. Hence, each model parameter represents a distinct psychological channel through which situational determinants and clinical characteristics can affect the decision-making process.

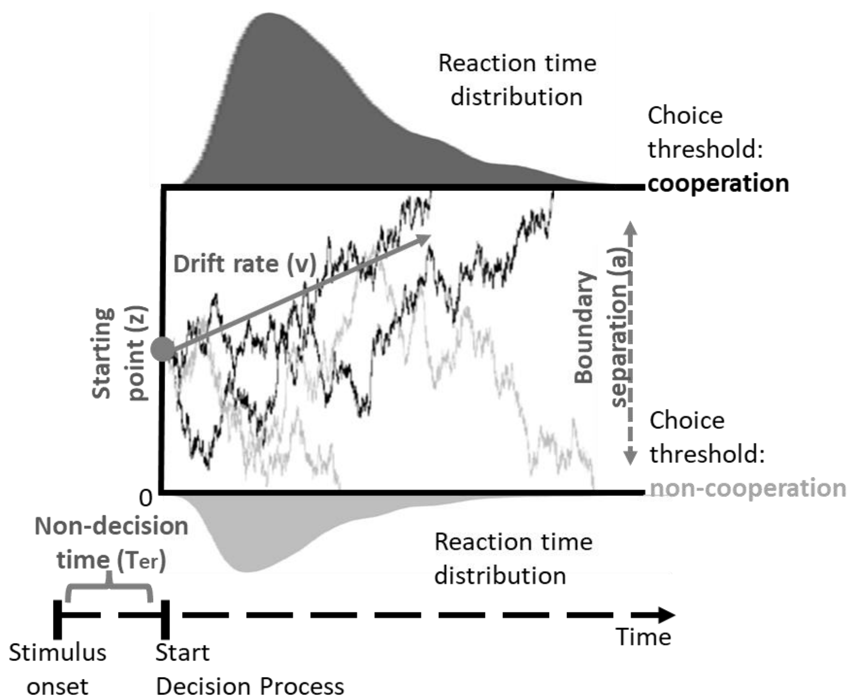

Figure 1. A computational account of cooperation across contexts. In the DDM,(Ratcliff, 1978) choices result from noisy trajectories that have a starting point (z) and that evolve up to a threshold, indicating type and speed of choices. Decisions that lead to cooperative choices are represented in black, ending at the upper boundary (a). Decisions that lead to non-cooperative choices are represented in grey, ending at the lower boundary (0). Drift rate (v) represents the mean rate of evidence accumulation. Shown are four diffusion processes, simulating the time course of evidence accumulation for four rounds of play.

We examined participants' cooperativeness across contexts, accounting for situational determinants (i.e., source and size of motivational conflict; prosociality of the setting; past experiences) and cognitive characteristics (i.e., inattention; hyperactivity-impulsivity). To do so, we developed a cooperation paradigm using different PD and SH games. This created a series of situations with distinct trade-offs between gains and risks (and costs) of cooperation. Participants were rematched with different counterparts at the end of every third round of play. In each round, they decided between cooperative and non-cooperative options.

The application of computational modeling allowed us to examine how different situational determinants affected the decision-making process through distinct cognitive channels; and whether ADHD was characterized by larger parameter changes across situations (a proxy for reactivity across contexts). To do so, we utilized payoff quadrants that have been associated with distinct motives by game theory (Ahn et al., 2001; Axelrod, 1967; Moisan et al., 2018; Rapoport & Chammah, 1965; Van Lange et al., 2014). Specifically, on-diagonal quadrants present cooperation gains, whereas off-diagonal quadrants present cooperation risks and costs (Figure 2a). Using eye-tracking we tested whether participants' cooperativeness was predictable by measures of their visuospatial attention (i.e., fixation duration) directed to cooperation gains relative to risks and costs (Figure 2b).

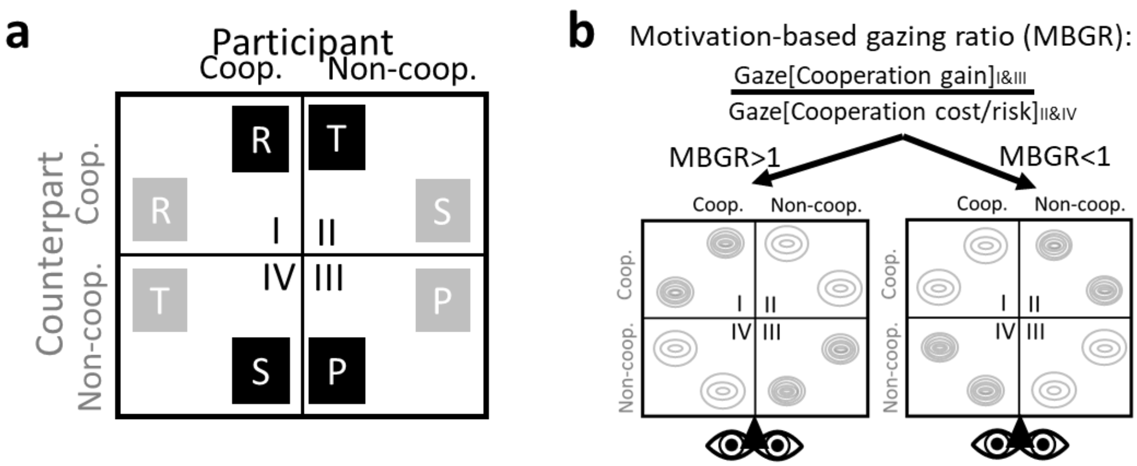

Figure 2. Motivation-based information processing during decisions. a. Payoff matrix (with black boxes corresponding to participants' payoffs and gray boxes corresponding to counterparts' payoffs)

in motivational space: R=reward for cooperation; P=punishment for non-cooperation; T=temptation to exploit others' cooperativeness; S=sucker's payoff when one's own cooperativeness is exploited by others. Roman numerals label the four quadrants. Based on game theory, on-diagonal quadrants (I & III) refer to the gains of cooperation, whereas off-diagonal quadrants (II & IV) refer to the (opportunity) costs and risks of cooperation. **b.** Two distributions of visuo-spatial attention (i.e., fixation duration). The left panel shows a motivation-based gazing ratio (MBGR) larger than one: Cooperation gain attracted more attention than cooperation costs and risks. The right panel shows an MBGR smaller than one: Cooperation gain attracted less attention than cooperation costs and risks.

To disentangle the relative influence of situational determinants and cognitive characteristics on behavior, counterparts were represented by computer players with predefined (but probabilistic) response strategies (see Methods). Computer players have been frequently used to examine cooperation in strategic interactions; that is, interactions in which participants' outcomes not only depend on their own actions but also on counterparts' actions (see for a review: March, 2021; see also: Andreoni & Miller, 1993; Crandall et al., 2018; Haruno & Kawato, 2009; Kangas et al., 2009; Mahon & Canosa, 2012; Shubik et al., 1974). However, one might question whether the fact that participants engaged with computer players would eliminate the social aspect of cooperation. A recent review by March (2021) suggests that participants are less cooperative in strategic interactions with artificial agents but that those situations still involve strategic uncertainty and social-cognitive considerations. The use of computer players was particularly useful for the purpose of our study since we focused on how people's cooperativeness changes across contexts (relative changes) not at the absolute level of cooperativeness per se. Moreover, it allowed us to disentangle the effect of specific situational determinants on specific mental components involved in the decision-making process.

The purpose of this study was to examine how participants' cooperativeness changes across simulated situations that differ in motivational conflict, prosociality, and prior outcomes. By so doing, we expected that individual differences in behavior is predictable by how much (overt) attention participants attribute to the gains versus the risks and costs of cooperation. Ultimately, we hypothesized that our novel paradigm is also useful to study social-cognitive characteristics of ADHD that manifests in over-reactive responses to changes across situations.

Methods

Participant Characteristics

Sixty-eight young adults (35 females), aged 18-35 years, participated. This sample size compares favorably to other neurocognitive studies (Ging-Jehli et al., 2021). Technical problems with data recording led to a data loss of two participants (both from the control group). Therefore, 66 participants (32 controls, 34 ADHD) were included in the analyses.

Inclusion was restricted to those who had: no serious head injury in the previous two years and who were not medicated for seizures. Participants in the ADHD group (n=34) had an ADHD diagnosis according to a semi-structured clinical interview (see Measures: K-SADS) and an ADHD symptom severity T-score of at least 64 (see Measures: CAARS-LV). They were also required to stop stimulants two days prior to the testing. Participants in the control group (n=34) needed an absence of any DSM-5 defined mental health conditions as assessed with the K-SADS; and an ADHD symptom severity T-score of at most 60 on the CAARS-LV. Participants received $20 for the completion of questionnaires and the clinical interview. They then received $20 for participation in the social-cognitive session (performance of the cooperation paradigm) plus any additional money they earned with their counterparts in the cooperation paradigm.

Participants were clinically assessed prior to their visit (detailed below), assigning them to either the ADHD (n=34) or the control (n=32) group. The two groups were similar with respect to age (M_{ADHD}=24 years, SD_{ADHD}=5; $M_{control}$=24 years, $SD_{control}$=4); gender (ADHD: 16 females; control: 17 females); and summed IQ scores (M_{ADHD}=144, SD_{ADHD}=17; $M_{control}$=139, $SD_{control}$=13). The control group had one more year of education than the ADHD group (M_{ADHD}=16, SD_{ADHD}=2; $M_{control}$=17,

$SD_{control}$=3). We provide additional clinical characteristics in Supplemental Table S1 and Supplemental Figure S8.

Study Procedure

The Ohio State University Institutional Review Board (IRB) approved this study. Participants provided written informed consent prior to participation. Participants completed the questionnaires, clinical interview, and the cooperation paradigm in separate study sessions. This study was conducted at a university in the Central Ohio metropolitan area (2 million people). Prospective participants were recruited through online recruiting platforms (e.g., researchmatch.org) and through flyers in local communities.

The Cooperation Paradigm

Our paradigm comprised 300 decisions to which we refer as trials. These 300 trials were subdivided into 100 games that each consisted of three rounds. After three rounds, participants were paired with a different counterpart with whom they had not interacted before. Counterparts were represented by computer players (detailed in subsection: *Counterpart Types*). The pairs' mutual actions determined payments at the end of the session (detailed in subsection: *Payoff Selection*). Specifically, participants and counterparts separately decided between a cooperative and a non-cooperative option in each round. The information above was common knowledge to all participants. We collected choice data and reaction times for each round. We next detail the paradigm, while also referring to the illustrations in Supplemental Figures S1 and S2 that provide an overview of the most important task features.

Procedure of a Game Round

Each round was divided into three stages: 1. Counterparts chose between the two response options; 2. Participants chose between the two options without knowing what their counterpart had chosen; and 3. Participants were informed about the outcome of that round after indicating their choice.

Response Options

In each round, the decision screen displayed a 2-by-2 payoff matrix[1], illustrating the four possible outcomes based on participant's and counterpart's choices. The four quadrants of the matrix associate each option with an on-diagonal quadrant and an off-diagonal quadrant (Figure 2a). The cooperative option was defined as the option with the higher payoffs in the on-diagonal quadrant. The two response options were labeled as options X and Y to avoid framing effects (Camerer, 2003). Moreover, the cooperative option was presented either as option X or Y half of the time. The black boxes contained participants' payoffs, while the grey boxes counterparts' payoffs. The boxes were arranged so that the corners nearest to the center were equally distanced from the screen center. The numbers within the boxes were not perceptible when fixating at the screen center. This ensured that participants had to move their gaze to the location of the boxes to encode the numbers, ensuring the collection of meaningful eye-tracking data.

Conditions

The payoffs associated with each option systematically changed across rounds, varying incentive congruency and strategic uncertainty as follows:

The first two rounds included payoff structures with congruent incentives. These payoff structures are also known as Stag-hunt (SH) games (Skyrms, 2004; Weber, 2018) and an example is illustrated in Figure 3. Matrices with SH structures varied in strategic uncertainty (low, medium, high) based on the relative difference in the minimal payoff of cooperative versus non-cooperative options

[1] We provide a list of all payoff matrices here: https://osf.io/m5xuh/?view_only=9e36d731ebc0404abd86b631261f6a06

(Skyrms, 2004). We therefore created three conditions with congruent incentives by categorizing the SH payoff matrices based on whether the minimal payoff for cooperation was higher, equal, or lower than the minimal payoff for non-cooperation (Figure 3).

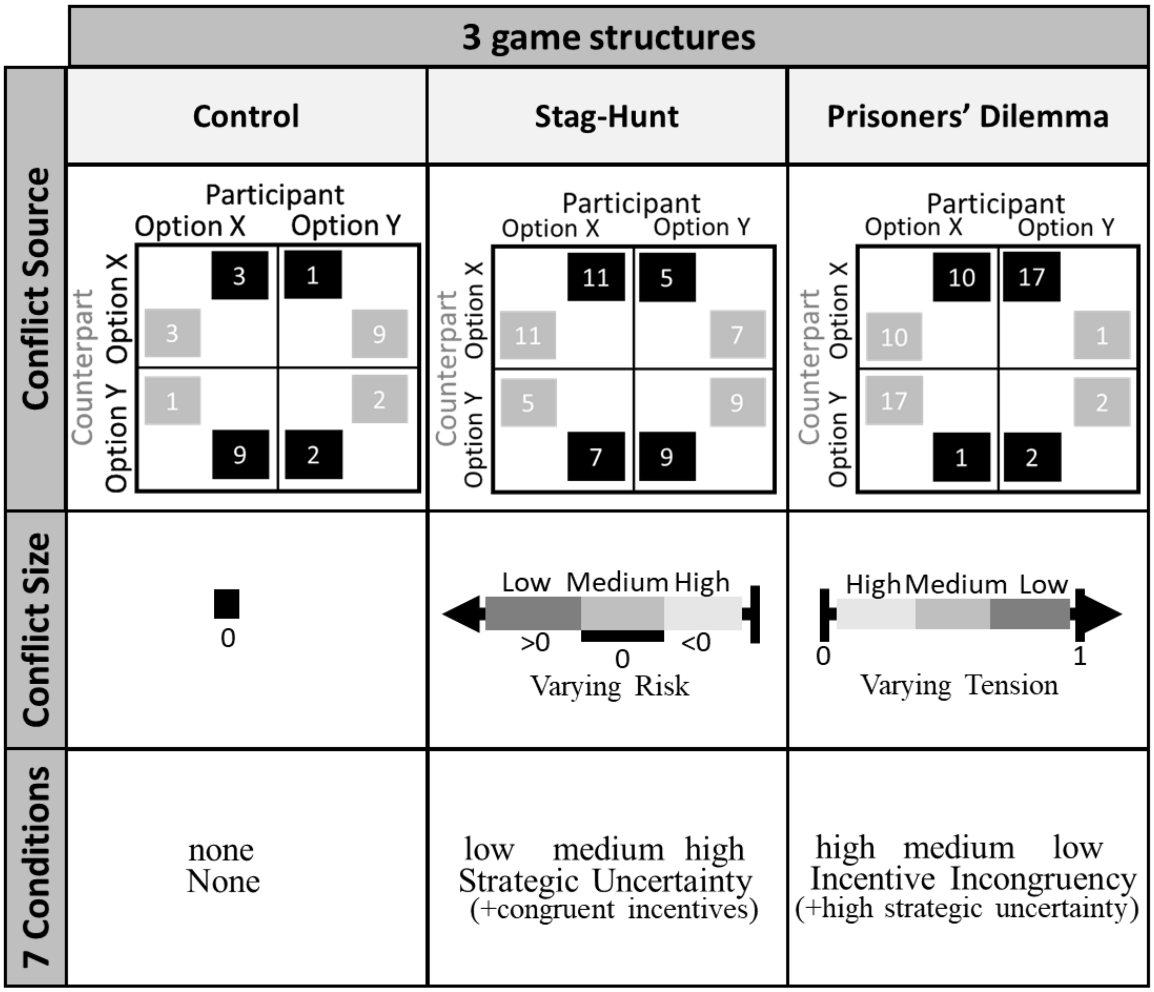

Figure 3. The 3 game structures and the 7 task conditions of our cooperation paradigm. a. The cooperation paradigm with variation in source and size of motivational conflict, leading to seven conditions. The highest on-diagonal payoff pair in the 2x2 payoff matrix defines the cooperative option, either presented as option X or Y. The black boxes in each payoff matrix refer to participants' payoffs, while the gray boxes refer to counterparts' payoffs. Shown are three examples (one for each game structure). The 2x2 payoff matrices varied on a trial-by-trial basis.

The last round included payoff structures with either: 1. High strategic uncertainty and incongruent incentives (either low, medium, or high level), which are also known as Prisoners' dilemma (PD) games (Dawes et al., 1990; Poundstone, 1993; Van Lange et al., 2014); or 2. With no strategic uncertainty and congruent incentives (i.e., providing participants a larger payoff, irrespective of their counterpart's choice). We refer to these payoff structures in 2. as the control condition. Matrices with PD structures varied in how much conflict they induce which can be quantified with the K-index (Moisan et al., 2018; Murphy & Ackermann, 2015). The K-index is based on early theoretical frameworks that provide a simple analytical way to measure the size of conflict in PD games (Axelrod, 1967; Moisan et al., 2018; Rapoport & Chammah, 1965). It is the ratio of the difference between on-diagonal payoffs to the difference between off-diagonal payoffs (Figure 2a). The K-index ranges from zero to one (excluding boundaries) with higher values indicating less tension. We therefore created three conditions with incongruent incentives by categorizing the PD payoff matrices (i.e., payoff matrices of each round with a PD structure) into terciles based on their K-index (Figure 3).

Using SH and PD games with different conflict levels allowed us to examine how cooperation varied between different conflict sources (i.e., tensions between response options caused by the different payoff structures, inducing strategic uncertainty with/without congruent incentives), and conflict sizes (low, medium, high). To summarize, cooperation was relevant in all conditions except those with low or no strategic uncertainty. In those latter two conditions, the cooperative option yielded higher payoffs irrespective of the counterparts' choices.

Sequence of Games

Matrices with PD structures occurred in rounds 3, while those with SH structures occurred in rounds 1 and 2. This allowed us to quantify the effect of participants' prediction errors (i.e., deviations from participants' expected outcomes)[2] on their subsequent cooperativeness. To ensure that participants

[2] Since incentives were aligned in the first two rounds, participants' choices in those rounds revealed their expectations about their counterparts' cooperativeness (see Weber, 2018). Therefore, presenting payoff structures with incongruent incentives at the end of each game allowed us to examine whether positive or negative surprises shifted participants' cooperativeness under incongruent incentives.

did not choose non-cooperation by default in the last game rounds, we sometimes presented a control game with cooperation being the strictly dominant choice (i.e., no motivational conflict).

Composing games with three rounds allowed us to examine how past experiences affected participants' choices in subsequent rounds (and whether participants transferred past experiences to subsequent games). Since incentives were aligned in the first two rounds, participants' choices in those rounds revealed their expectations about their counterparts' cooperativeness. This is because SH structures incentivize participants to choose the option that they believe their counterpart is going to choose (see for a review: Weber, 2018).

Therefore, we defined outcomes in the first two rounds as positive surprises if participants chose non-cooperation, and their counterparts chose cooperation. Similarly, we defined outcomes as negative surprises if participants chose cooperation, and their counterparts chose non-cooperation. Presenting payoff structures with incongruent incentives at the end of each game allowed us to examine whether positive or negative surprises shifted participants' cooperativeness under incongruent incentives. It also ensured that participants had no opportunity to punish/reward their counterparts in subsequent rounds, motives known to affect cooperativeness (Camerer, 2003; Fehr & Gächter, 2000).

Counterpart Types

Participants' counterparts were displayed as computer players that had pre-determined but probabilistic response strategies (being either more or less cooperative). This allowed us to simulate prosocial and antisocial settings. Specifically, more cooperative counterparts chose the cooperative option with probability 1 in rounds 1; and with probability 0.7 in rounds 2 & 3. In contrast, uncooperative counterparts chose the cooperative option with probability 0 in rounds 1; and with probability 0.3 in rounds 2 & 3. At the beginning of the experiment, participants were informed that the computer players had different response strategies. However, they did not know the players' type (i.e., degree of cooperativeness) in a game, and they also did not know how many types there were in the paradigm.

Setting

The 100 tournaments were subdivided into four blocks (each block containing 25 tournaments that each were composed of 3 rounds). In the first and third blocks, counterparts were from the cooperative type, responding predominantly cooperatively. This created two predominantly cooperative settings. In the second and fourth blocks, counterparts were from the uncooperative type, responding predominantly uncooperatively. This created two predominantly uncooperative settings. Importantly, the payoff structures (i.e., 2x2 payoff matrices) in the first and third blocks were identical to those in the second and fourth blocks. Hence, any differences in participants' cooperativeness across blocks could not be explained by any differences in the payoff structures. This setup allowed us to examine the effects of setting on participants' cooperativeness; and to disentangle the effects of setting and participants' past experience (i.e., expected/unexpected positive or negative prior outcomes) on their cooperativeness. Therefore, all participants were presented the same block order.

Payoff Selection

One of the 300 trials was randomly selected, and the outcome of that selected trial was then implemented at the end of the paradigm. Specifically, participants earned the selected outcome in addition to their study payment ($1 to $19).

Eye-tracking

We collected gaze locations using the GP3-HD eye tracker (sampling rate: 150Hz) from Gazepoint (Cuve et al., 2021; Gazepoint, 2021). The eye tracker was attached to the desk in front of the participant's monitor. To position participants in front of the monitor, we used the Gazepoint's software (monitoring window) and physical measurements (i.e., angle: 45°; distance to participants' eyes: 72cm). The standard viewing distance was set to 75cm which corresponds to a visual angle of

0.042 by 0.042 degrees per pixel on the 640 by 480 resolution monitor. A 5-point location scheme calibration was used for the eye tracker, after instructing participants on the cooperation paradigm.

Processing Gaze Data

We used our own Matlab scripts (Version 2021a; (The Math Works, Inc., 2020) to process the gaze data, excluding first any data flagged as invalid (e.g., blinks) by the eye-tracker ((Cuve et al., 2021). We then extracted the gaze data during the presentation of the decision screen (i.e., the time from stimulus onset up to key press). We next applied a trial-by-trial drift correction using the center of the fixation screen as reference as is common practice (Holmqvist et al., 2011). We did not further preprocess gaze data (no smoothing, filtering, interpolation) to prevent data distortions (such as artificial improvement in accuracy and/or precision).

Gaze Measures

The eight boxes on the 2x2 payoff matrix, containing the payoff information, are the eight spatially defined areas of interests (AOIs) that we used to understand how participants acquired information (Capra et al., 2020). It is a common practice (Capra et al., 2020) to distinguish between saccades (i.e., rapid eye movements to shift one's gaze from one location to another) and fixations (i.e., stimulus processing by holding gaze on an area of interest for longer). It has been shown that fixations are typically at least 70ms long and last around 200ms to 300ms at most (Holmqvist et al., 2011; Salvucci & Goldberg, 2000). We therefore extracted the following four gaze measures for each round: 1. The total number of fixations (i.e., uninterrupted fixation on a AOI that lasts at least 70ms); 2. The average fixation duration (i.e., summed duration of fixations divided by the total number of fixations); 3. The number of fixations for each AOI; and 4. The average fixation duration for each AOI.

Motivation-based Gazing Ratio (MBGR)

We were interested in whether the relative magnitude of (overt) attention on cooperation gains versus cooperation costs and risks would predict participants' cooperativeness on a trial-by-trial basis. Specifically, we summed the fixation duration in quadrants I and III and divided that sum by the fixation duration in quadrants II and IV. By so doing, we leverage game theory, associating quadrants I and III with cooperation gains and quadrants II and IV with cooperation costs and risks (Mengel, 2018; Rapoport & Chammah, 1965). We refer to this motivation-based gazing ratio as MBGR. As a sensitivity analysis, we also calculated MBGR based on number of fixations rather than fixation durations which yielded the same results.

Clinical Measures

Schedule for Affective Disorders & Schizophrenia (K-SADS-PL-5)

We administered the K-SADS-PL-5 (Chambers et al., 1985; Kaufman et al., 1997) a conventional semi-structured clinical interview, to formally confirm the absence/presence of ADHD and/or other DSM-5 defined mental health conditions such as mood disorders, anxiety, obsessive compulsive, trauma related disorders, neurodevelopmental, disruptive, and conduct disorders. A trained graduate student, under the guidance of a medical doctor specialized in psychiatry, administered the K-SADS-PL-5 in a separate session prior to the social-cognitive testing.

Conners' Adult ADHD Rating Scale, long version (CAARS-LV)

We administered the established CAARS-LV questionnaire (Conners et al., 1999) to measure symptom severity of total ADHD symptoms (G-section). The CAARS-LV was developed to assess adult ADHD, including severity scores and symptom counts based on the DSM-IV (American Psychiatric Association, 2000). It comprises 66 questions (response options ranging from 0 for *never* to 3 for *very much*). The CAARS-LV is a well-established clinical measure that provides T-scores, ranging from 25 to 90, with a score of 50 representing the average severity level of a symptom in the normative population (scores above 64 represent elevated and possibly clinically significant symptoms).

Non-clinical Measures

Demographic Questionnaire

At the beginning of the study, all participants answered a brief questionnaire indicating age, sex, and years of education.

Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)

We used the WASI-II (Wechsler, 2011) an established IQ measure, to control for any intellectual differences between the control and ADHD group. Specifically, we used the three modules: vocabulary, matrix reasoning, and similarities. Scores for each module range from 40 to 160, and we provide the summed scores across the three modules in the Results section.

Analytical Approach

Before examining performance measures of the cooperation paradigm, we excluded data from the first game (i.e., practice game) for each participant as well as extreme reaction times (RTs) longer than 10 seconds (0.2% of all rounds across all participants. These RT outliers were evenly distributed among participants).

Our analytical approach consisted of three parts: 1. Behavioral analyses, using performance measures and linear and single-trial logistic regressions within a Bayesian hierarchical framework; 2. Process-oriented analyses, using computational modeling within a Bayesian hierarchical framework; and 3. Bayesian' hypothesis testing for examining ADHD-specific social-cognitive characteristics. For all analyses, we used the software R (Version 4.1.0; (R Core Team, 2021) and the brms (Version 2.16.2; (Bürkner, 2017a, 2017b) package as an interface to STAN ((Stan Development Team, 2021) a toolbox for estimating models within a Bayesian hierarchical framework. This approach has multiple advantages discussed elsewhere (Gelman et al., 2012; J. K. Kruschke, 2013). As it is a common practice in Bayesian analyses, we report expected values (point estimates) and 95% credible intervals (CIs) for each model parameter and we refer to the Supplement for a brief description of CIs.

Behavioral Analyses

Analyzing differences in mean RTs across conditions, we used linear mixed regressions with participants' mean reaction times (one for each condition) as outcome. The paradigms' conditions served as a covariate (additionally including the clinical group for the clinical analysis discussed in the last part of the Result section). The general form of the main model can be written as follows:

$$\text{mean RT}_{c,p} \sim \text{Gaussian}(\alpha_{pc}),$$

with the Greek letter indexing the coefficient, *c* referring to conditions and *p* referring to participants. The model was run with 4 chains (each with a total of 8,000 samples which included 2,000 burn-in samples). Model convergence was ensured by examining the trace plots and the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992) which was below 1.1 for all model parameters. We used the default, weakly informative priors set by STAN (i.e., a flat prior for slopes and a half Student's-t with 3 degrees of freedom and a scale parameter depending on the standard deviation of the response for intercepts and variances (Bürkner, 2017a, 2017b). For the assessment of clinical group differences, we included *group* (control versus ADHD) as an additional hierarchical level.

Analyzing participants' cooperativeness, we used single-trial logistic regressions with participants' choices (1=choose cooperative option, 0=choose non-cooperative option) as the binary outcome. The model was specified as follows:

$$\text{choice}_{pcgr} \sim \text{Bernoulli}(\alpha_{pcgr}),$$

whereby the Greek letter indexes the coefficient, *p* indexes participants, *c* indexes conditions, *g* indexes games, *r* indexes round. The model was run with 9 chains (each with a total of 4,000 samples which included 1,000 burn-in samples). Model convergence was ensured by examining the trace plots and the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992) which was below 1.1 for all model parameters. We used weakly informative priors commonly used for logistic regression (Gelman & Hill, 2006; J. Kruschke, 2014; McElreath, 2016). That is, a normal distribution with mean equal to 0 and standard deviation equal to 5 for regression coefficients; an exponential distribution with parameter rate equal to 2 for standard deviations; and the Lewandowski-Kurowicka-Joe distribution (Lewandowski et al., 2009) with scale parameter equal to 2 for correlations between pairs of regression coefficients. For the assessment of clinical group differences, we included *group* (control versus ADHD) as an additional hierarchical level.

Computational Modeling

We leveraged the DDM (Ratcliff, 1978) to decompose performance (i.e., choices and the entire RT distributions of cooperative and non-cooperative choices) into latent mental components.

Modeling value-based decisions. While the DDM has originally been used to only account for decision processes in cognitive tasks (i.e., noisy sensory environments(Forstmann et al., 2016), research has also established its usefulness for understanding decision dynamics in tasks that include value-based decisions like those in our cooperation paradigm (Cavanagh et al., 2011; Krajbich et al., 2010; Krajbich & Rangel, 2011; Pedersen & Frank, 2020). For those value-based decisions, the boundaries represent the two response options (Figure 1a: cooperative vs. non-cooperative choice). Hence, boundary separation (a) captures the trade-off between making speedy versus consistent decisions (i.e., the propensity to choose the same course of action across rounds with similar payoff structures; see also discussion by (Pedersen et al., 2021). Moreover, drift rates capture decision difficulty (i.e., quality of evidence accumulation) induced by task stimuli. In our context, stimulus difficulty arises from the conflict of the payoff structure (i.e., 2x2 payoff matrix as stimulus). Hence, we modeled cooperative decisions by assuming that the systematic variation of the payoff structures across rounds affects the evidence accumulation process (drift rate). Quantifying the starting point relative to the boundary separation measures biases towards cooperative vs non-cooperative choices (i.e., predispositions towards a choice prior to the presentation of a payoff matrix). The DDM assumes that RTs are the sum of decision processes and non-decision processes such as perceptual encoding and motor execution (Ratcliff & Smith, 2004). The non-decision times are denoted as T_{er}, which determines the location of the RT distribution (i.e., speed of the fastest responses). Because T_{er} only leads to shifts in the location of the RT distribution, it is often treated as a constant (Ratcliff & Smith, 2004).

Base model. We started our computational analysis with a model that included the four main DDM parameters (a, z, v, T_{er}). In so doing, we modeled the decision process as a diffusion process (i.e., including within-trial variability in drift rate) but abstained from introducing additional variability parameters (i.e., variability in T_{er}, variability in z, across-trial variability in v). This is because an initial analysis showed that the inclusion of these additional variability parameters did not improve data fits. Instead, it added unnecessary model complexity, comprising risks of overfitting and trade-offs across parameters.

Next, we found that a simpler model, fixing T_{er} to a constant, provided similar estimates to those of the initial model (Supplemental Table S2 & S3). Under the simpler model we treated T_{er} as an additional datapoint. Specifically, for each participant, response type (cooperation, non-cooperation), and condition, we calculated 70% of the fastest RT and subtracted that value from each RT. This shifted each RT distribution to the same location, introducing the assumption that perceptual encoding and motor execution was stimulus-independent (i.e., different values in the 2x2 payoff matrix do not affect perceptual encoding of the decision screen). This approach has already been used in other applications (Ratcliff & Smith, 2004).

For this study, we were interested in treating T_{er} as a constant due to the following reasons: first, we did not have theoretical reasons to assume that the perceptual encoding of 2x2 payoff matrices varied as a function of the paradigms' conditions: the decision screen remained the same across all rounds of the paradigms, changing only the values in the matrices across rounds. Second, it was theoretically meaningful to assume that boundary separation (a) could serve as a potential channel for the effects of situational determinants (i.e., setting, prior outcomes) on the underlying decision-making process. However, boundary separation (a) and T_{er} can be negatively correlated (Ratcliff & Smith, 2004). To avoid that boundary separations were simply larger due to smaller T_{er} for some blocks, fixing T_{er} was the more conservative analytical approach.

Our final base model included the three main DDM parameters: a, z, and v. We introduced one drift rate (v) for each of the paradigm's conditions (Figure 1a). Specifically, the base model captured choices and RTs for each condition with the Wiener first passage time likelihood function (Wfpt) which can be specified as follows:

$$(\text{choice}, \text{RT})_{p,c} \sim \text{Wfpt}(a_p, z_p, v_{p,c}, T_{er}|_{constant}),$$

with *c* referring to the paradigm's conditions, and *p* referring to participants. This base model then served as a foundation for the comparison of different model specifications described next. All models included the same weakly informative priors, placing mass on broad range of reasonable values for DDM parameters: v ~ Cauchy(0,5); a ~ Normal(1.5, 1); zb ~ Normal(0.5, 0.2); σ ~ Student's-t(3, 0, 10); r ~ *LKJ*(2), with v referring to drift rate, a referring to boundary separation, zb referring to starting point bias (i.e., starting point relative to the boundary), σ referring to standard deviations, r referring to correlations between pairs of regression coefficients, and *LKJ* referring to the Lewandowski-Kurowicka-Joe distribution (Lewandowski et al., 2009).

Model comparison. We systematically assessed the performance of different model versions. These models included theoretically meaningful but distinct assumptions about the cognitive channels (i.e., DDM parameters) through which the situational determinants (i.e., setting and prior outcome) affected the underlying decision-making process (Table 1). Specifically, we tested whether the setting and/or the prior outcomes modulated boundary separation (a) and/or starting point (z). To identify the model specification that best accounted for the data, we compared model performances using the Watanabe-Akaike Information Criterion (WAIC; (Watanabe, 2010) as well as out-of-sample predictions estimated by Bayesian leave-one-out (LOO) cross-validation (McElreath, 2016; Vehtari et al., 2017) implemented through the loo R package (Vehtari et al., 2017). Hence, the best-fitting model was the one that converged ($\hat{R}$ statistic was below 1.1 for all parameters), and that best accounted for the observed behavioral patterns with respect to both the leave-one-out cross-validation, and posterior predictive checks (provided in the Supplement).

After establishing the best-fitting model, we extended that model with our trial-based gaze measure of interest, namely the motivation-based gazing ratio (MBGR; see section: Eye-tracking) which is a continuous variable. This allowed us to assess whether trial-based drift rates (v) changed not only as a function of the paradigm's conditions but also as a function of the attentional magnitude on cooperation gains versus costs and risks. Indeed, including the gaze measures (MBGR) improved the model as indicated by a reduction in the WAIC from 63,480 (i.e., Table 1: best-fitting model) to 53,799. Therefore, the final best-fitting model discussed in this manuscript included drift rates specified as:

$$v_{ct} = \alpha_c + \beta_c * MBGR_{ct},$$

where *t* refers to trials, *c* refers to condition, and *MBGR* (continuous variable) refers to the mean-centered motivation-based gazing ratio. Note that the attentional measure was hierarchically mean-centered before entering it as a covariate (Jackman, 2009). Boundary separation (a_s) and starting points (z_{st}) were specified as:

$$a_s = \alpha_{[s]},$$

$$z_{st} = \alpha_{[s]} + \beta_{[s]t} * prior\ outcome_{[s]t(where\ r>1)} + \delta_t * game\ start_{t(where\ r=1)}$$

where *t* refers to trials, *s* (index variable) refers to setting (i.e., the paradigm's four blocks), and *r* refers to the three rounds of each game. We also show good parameter recovery based on additional sensitivity analyses in Supplemental Figure S4 and S5.

Fitting procedure. As for the behavioral analyses, we fitted the DDM models to data within a Bayesian hierarchical framework implemented through brms and STAN (Bürkner, 2017b; Stan Development Team, 2021). All models were run with 4 chains (each with a total of 6,000 samples which included 2,000 burn-in samples). Model convergence was ensured by examining the trace plots and the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992) which was below 1.1 for all model parameters. All models included the same weakly informative priors specified in the subsection *Base model*. We also exemplify our modeling script for this study (including sample data)[3].

Model validation. We used posterior predictive checks reported in the Supplement, assessing that the models captured the observed data (i.e., choice and response time patterns) well.

Bayesian Hypothesis Testing

We used Bayesian hypotheses testing to examine group-specific clinical differences. We report Bayes' Factors (*BF*) and posterior probabilities (*PP*) as it is common (van Doorn et al., 2021);

[3] https://osf.io/m5xuh/?view_only=9e36d731ebc0404abd86b631261f6a06

(Andraszewicz et al., 2015). Moreover, we report expected values (point estimates) and 95% credible intervals (CIs) for each model parameter and we refer to the Supplement for a brief description of CIs. We also provide posterior means and credible intervals (CIs) of the best-fitting DDM model in the Supplementary Tables S4 & S5.

Results

We first provide a summary of reaction times and cooperation rates across conditions before we introduce a computational account of the underlying decision-making process of cooperation.

Descriptive Analyses

Cooperative choices are associated with distinct gaze patterns. Participants' cooperativeness decreased from 95% under no conflict to 40% under high strategic uncertainty (Figure 4a). Introducing incongruent incentives (keeping strategic uncertainty high), decreased cooperation to 10% and 15% for high and low incongruency, respectively. Bayesian hypothesis tests suggested that this 5%-difference in cooperation rate (CR) was systematic; posterior probability (*PP*) that ($CR_{\text{high incongruency}}$<$CR_{\text{low incongruency}}$)=1.00.[4]

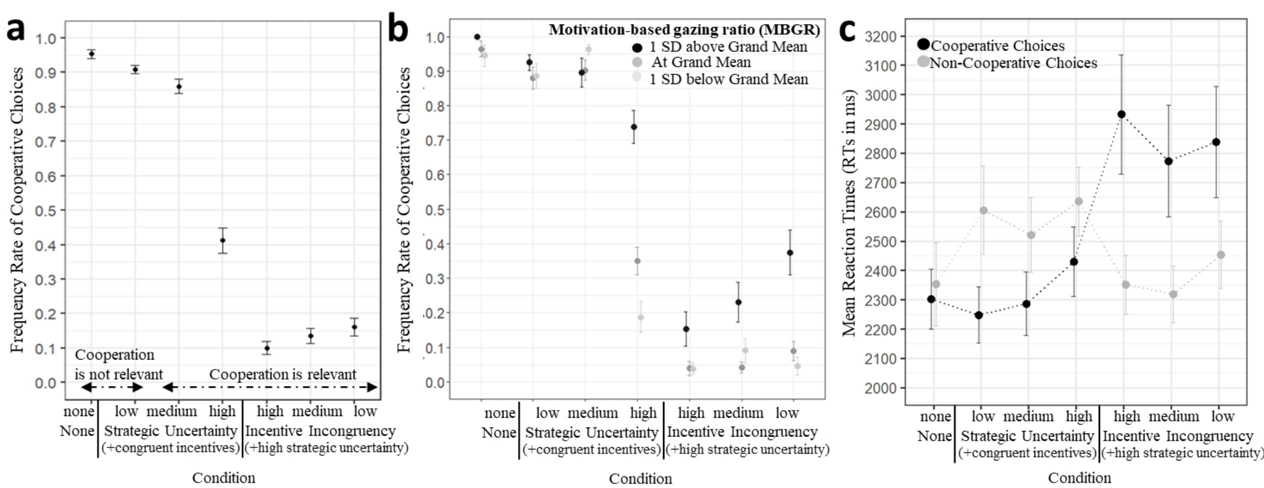

Figure 4. Aggregated performance by choice and condition. a. Mean frequency rates of cooperative choices by condition. **b.** Mean frequency rates of cooperative choices by condition for three MBGR levels (averaged over participants) with vertical bars representing standard errors. MBGR was a continuous variable but trichotomized for illustration purposes. **c.** Mean reaction times by condition (averaged over participants) with vertical bars representing standard errors.

[4] Bayes' Factor (*BF*) = 200.34; mean difference (log odds): -0.52, 95%-CI = [-0.85,-0.19].

Figure 4b illustrates how MBGR affected cooperation frequencies for each condition. For illustration purposes, we show the effects for three MBGR levels (even though MBGR was treated as a continuous variable in any statistical analyses; see Methods). Participants were more cooperative when they focused on the gains rather than the risks and costs of cooperation as shown in Figure 4b by the black dots that are above the grey dots for the conditions for which cooperation is more relevant.[5]

The speed of cooperative choices depends on source and size of motivational conflict. Whether cooperative choices were faster than non-cooperative choices depended on the source of conflict (Figure 4b). Under congruent incentives (SH games), cooperative choices were on average 280ms (*SD*=78ms) faster than non-cooperative choices and the estimated posterior probability (*PP*) that mean $RT_{cooperation}$<mean $RT_{non\text{-}cooperation}$=1.00.[6] Under incongruent incentives (PD games), cooperative choices were on average 469ms (*SD*=33ms) slower than non-cooperative choices and the estimated *PP* that mean $RT_{cooperation}$>mean $RT_{non\text{-}cooperation}$=1.00.[7] We provide additional descriptive results in Supplementary Figure S9.

Diffusion Decision Model Analyses

Using the DDM, we examined how situational determinants (i.e., condition, setting, prior outcome, and MBGR) influenced the decision-making process. This allowed us to simultaneously consider the frequency and speed of cooperative and non-cooperative choices. We use the term

[5] Averaging estimates across all conditions, the estimated posterior probability (PP) that CR for higher (1SD above the mean) MBGR is larger than CR for MBGR at the mean = 1.00; corresponding *BF* = 507.47; estimated mean difference (log odds): 5.13, 95%-CI = [2.30, 7.88]. We provide detailed statistics in the Supplementary Table S4 and Figure S10 also shows the effects of MBGR on accuracy for quartiles of MBGR values.
[6] *BF* = 140.18; log mean difference: 0.23, 95%-CI = [0.02, 0.46].
[7] *BF* = 856.14; log mean difference: -0.41, 95%-CI = [-0.75, -0.10].

setting to refer to blocks of games in which counterparts' response strategy was either cooperative or non-cooperative (see Methods).

We compared a sequence of nested models (Table 1), fitting each to the data using hierarchical Bayesian modeling. Each model is based on different, theoretically meaningful assumptions about the psychological channel through which the determinants affected the decision-making process. Specifically, stimulus-driven attributes like motivational conflict (payoffs) and gaze patterns might influence drift rates which reflect the speed of stimulus processing. Prior outcomes might influence starting points which reflects biases of subsequent decisions towards cooperation or non-cooperation. The setting might contribute to biases in starting points and/or affect participants' boundary separation which reflects participants' overall response cautiousness. Leave-one-out cross-validation was used to compare the fits of the different models. We then chose the best-fitting model in Table 1 and expanded it to include our gaze measure of interest (MBGR). Details are provided in the Methods.

The model around which our results are organized is specified in Table 1 and in the Methods (subsection: Computational Modeling). Posterior predictive checks, parameter recovery analyses, and the results from other models (leading to the same findings reported here) are presented in the Supplement. Posterior means and 95% credible intervals (CIs) are illustrated in Figure 5 and reported in Supplementary Table S5. The parameter, nondecision time (Ter), was computed from the fastest RTs (detailed in the Methods).

Table 1. Diffusion decision model (DDM) selection for behavioral data.

Model	Parameters			WAIC		LOO cross-validation		
Label	Drift rate (v) by	Starting point (z) by	Boundary separation (a) by	Estimate	SE	Δelpd	Δse	Rank
1	condition	setting & prior outcome & game start	setting	63480	351	0	0	1
2	condition	prior outcome & game start	setting	63867	356	-187	28	2
3	condition	round & prior outcome	setting type	65695	351	-1101	51	3
4	condition	prior outcome & game start	setting type	65733	353	-1119	52	4
5	condition	prior outcome	setting type	65765	356	-1135	54	5
6	condition	setting type	prior outcome	65816	356	-1160	57	6
7	condition	fixed	setting type	66054	358	-1277	59	7
8	condition	fixed	setting type & prior outcome	66069	359	-1288	60	8
9	condition	round & prior outcome	constant	66106	356	-1305	58	9
10	condition	setting type	constant	66151	359	-1327	62	10
11	condition	prior outcome & setting type	constant	66163	359	-1334	62	11
12	condition	fixed	prior outcome	66201	362	-1351	65	12
13	condition	prior outcome	constant	66234	361	-1369	64	13
Base	condition	fixed	constant	66415	359	-1457	65	14
14	condition	constant	constant	66628	365	-1563	70	15

Different DDMs (one model per row, detailed in the Methods). Models are presented in decreasing order of fitness, with the best-fitting model first (indicated by the last column). *Condition* refers to the paradigm's seven conditions (Figure 3). *Setting* refers to the paradigm's four blocks. *Prior outcome* refers to the four possible outcomes in the previous round ({cooperation, cooperation}, {non-cooperation, non-cooperation}, {non-cooperation, cooperation}, {cooperation, non-cooperation} with the first term in the curly brackets referring to the participant's choice and the second term referring to the counterpart's choice). *Setting type* refers to the two block types based on counterparts' response strategy (prosocial vs. antisocial). *Game start* equals one for rounds 1, and equals two for rounds 2 & 3. This allowed us to introduce separate starting point biases for the beginning of new games (see Methods). *Fixed* means that the starting point was set to be equidistant from each decision boundary (excluding starting point biases). *Constant* means that we kept the parameter constant throughout the paradigm, introducing one parameter per participant. *WAIC* = Watanabe-Akaike information criterion for which lower values indicate better fits. *LOO* = leave-one-out cross-validation. Δelpd = difference in expected log pointwise predictive density as compared to the best-fitting model. Δse = difference in expected standard errors as compared to the best-fitting model.

Situational determinants affect the decision-making process. Parameters extracted from the best-fitting model demonstrated that variation in conflict and MBGR affected drift rates (Figure 5a). As strategic uncertainty increased, those with lowest MBGRs showed the largest change in drift rates towards non-cooperative choices. This indicates that more attention on costs/risks of cooperation (rather than gains) was associated with more difficult cooperative decisions. Under incongruent incentives, non-cooperative decisions were easier (larger negative drift rates) with little variation across levels of conflict for MBGR below the grand mean. For MBGR above or at the grand mean, drift rates increased as incongruency decreased. For MBGR above the grand mean, the estimated posterior mean $v_{\text{high incongruency}}$=-0.50 (SD=0.13), estimated posterior mean $v_{\text{low incongruency}}$=-0.25 (SD=0.13), and the *PP* that the estimated posterior mean $v_{\text{high incongruency}}$<mean $v_{\text{low incongruency}}$=1.00; *BF* > 500. For MBGR at the grand average, the estimated posterior mean $v_{\text{high incongruency}}$=-0.88 (SD=0.03), estimated posterior mean $v_{\text{low incongruency}}$=-0.75 (SD=0.03), and the *PP* that the estimated posterior mean mean $v_{\text{high incongruency}}$< mean $v_{\text{low incongruency}}$=1.00; *BF* > 500.

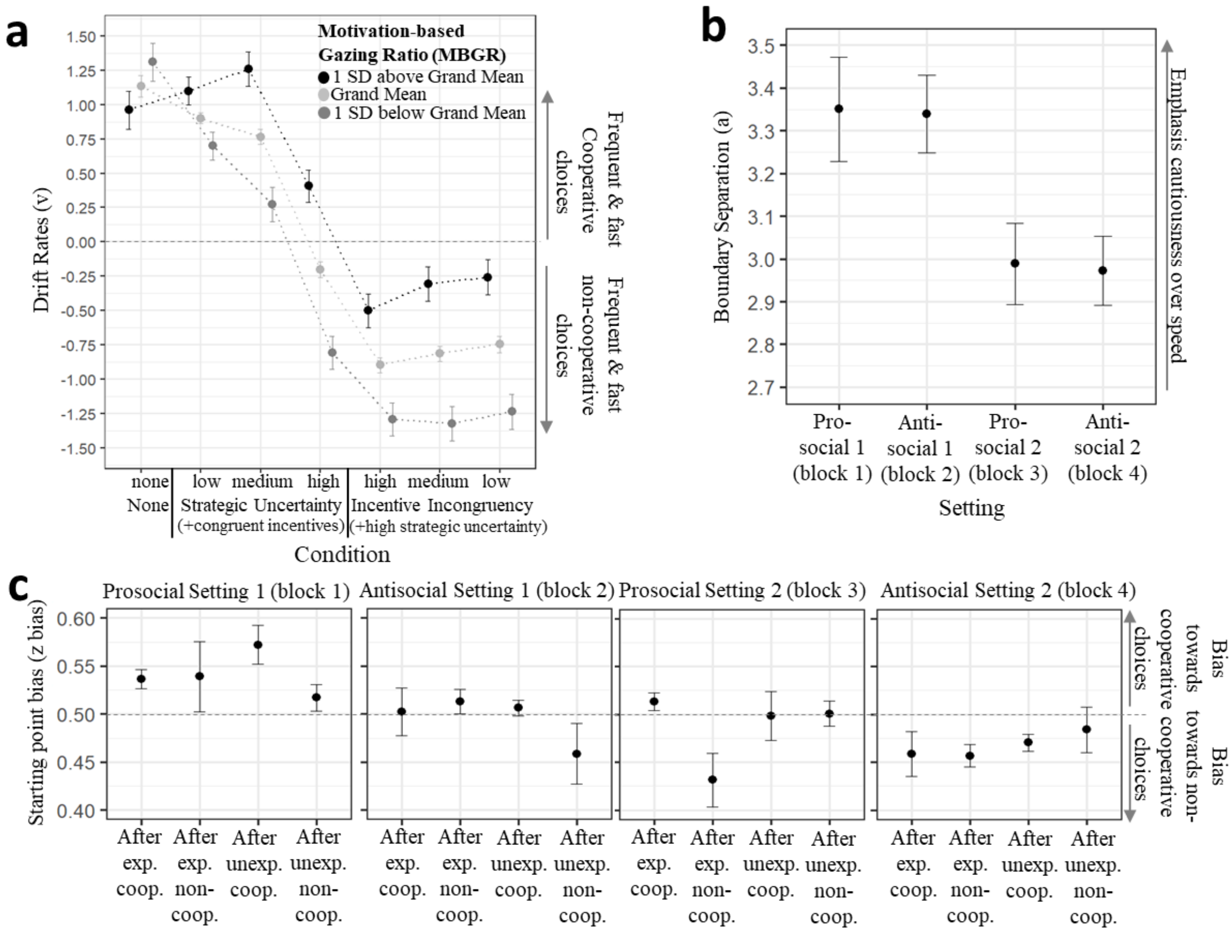

Figure 5. Estimated DDM parameters across all participants. a. Drift rate by condition for three MBGR levels. MBGR was a continuous variable but trichotomized for illustration purposes. Shown are the estimated posterior distributions (with posterior means as points and corresponding 95% credible intervals as vertical bars) of the drift rates. **b.** Boundary separation by setting (i.e., blocks of games; see Methods). Shown are the estimated posterior distributions (with posterior means as points and corresponding 95% credible intervals as vertical bars) of the boundary separations. **c.** Starting point biases (for rounds 2 and 3) towards cooperation (values>0.5) or non-cooperation (values<0.5) by setting and prior outcome. Shown are the estimated posterior distributions (with posterior means as points and corresponding 95% credible intervals as vertical bars) of the starting point biases.

Responses were equally cautious (and consistent) in prosocial and antisocial settings (i.e., blocks of games; see Methods) as indicated by the lack of change in posterior mean boundary separations across settings (Figure 5b). Participants had larger boundary separations for the first half of the task (blocks 1 & 2) than for the second half of the task (blocks 3 & 4) which could reflect training effects.

Setting and prior outcome both influenced the starting point, indicating bias toward cooperation or non-cooperation even before stimulus presentation (Figure 5c). In the first prosocial setting, participants were biased towards cooperation if their counterparts unexpectedly cooperated in the previous round. In the first antisocial setting, participants were biased towards non-cooperation if their counterparts unexpectedly non-cooperated in the previous round. To start, in the first setting, all responses (irrespective of the previous outcome) were biased towards cooperation. To the last setting, however, all responses were biased towards non-cooperation.

Clinical groups differ in performance. In what follows, asterisks refer to group differences in model parameters (MP) based on Bayesian hypothesis tests and displayed in the respective figures. *BF* is the Bayes' Factor with hypothesis H_1:(MP_{ADHD}>$MP_{control}$) in the numerator and hypothesis H_0:(MP_{ADHD}≤$MP_{control}$) in the denominator. *PP* is the posterior probability that MP_{ADHD}>$MP_{control}$.

The ADHD group was more cooperative than the control group under no conflict (*1: *BF*=5.58, *PP*=0.85), medium strategic uncertainty (*2: *BF*=17.35, *PP*=0.95), and low incongruent incentives (*3: *BF*=5.24, *PP*=0.84; Figure 6a). They also had a 321ms (*SD*=139ms) slower mean RT than the control group when averaging over all responses and conditions [MRT_{ADHD}=2558ms; $MRT_{control}$=2237ms; *PP* that MRT_{ADHD}>$MRT_{control}$=1.00]. Under incongruent incentives, mean RTs of the control group became faster as conflict decreased (Figure 6b). The opposite relationship was found for the ADHD group. Group-specific differences in mean RTs were due to differences in the number of gaze fixations (Figure 6c). The ADHD group fixated more on cooperation gains (rather than risks and costs) under incongruent incentives than did the control group (Figure 6d).

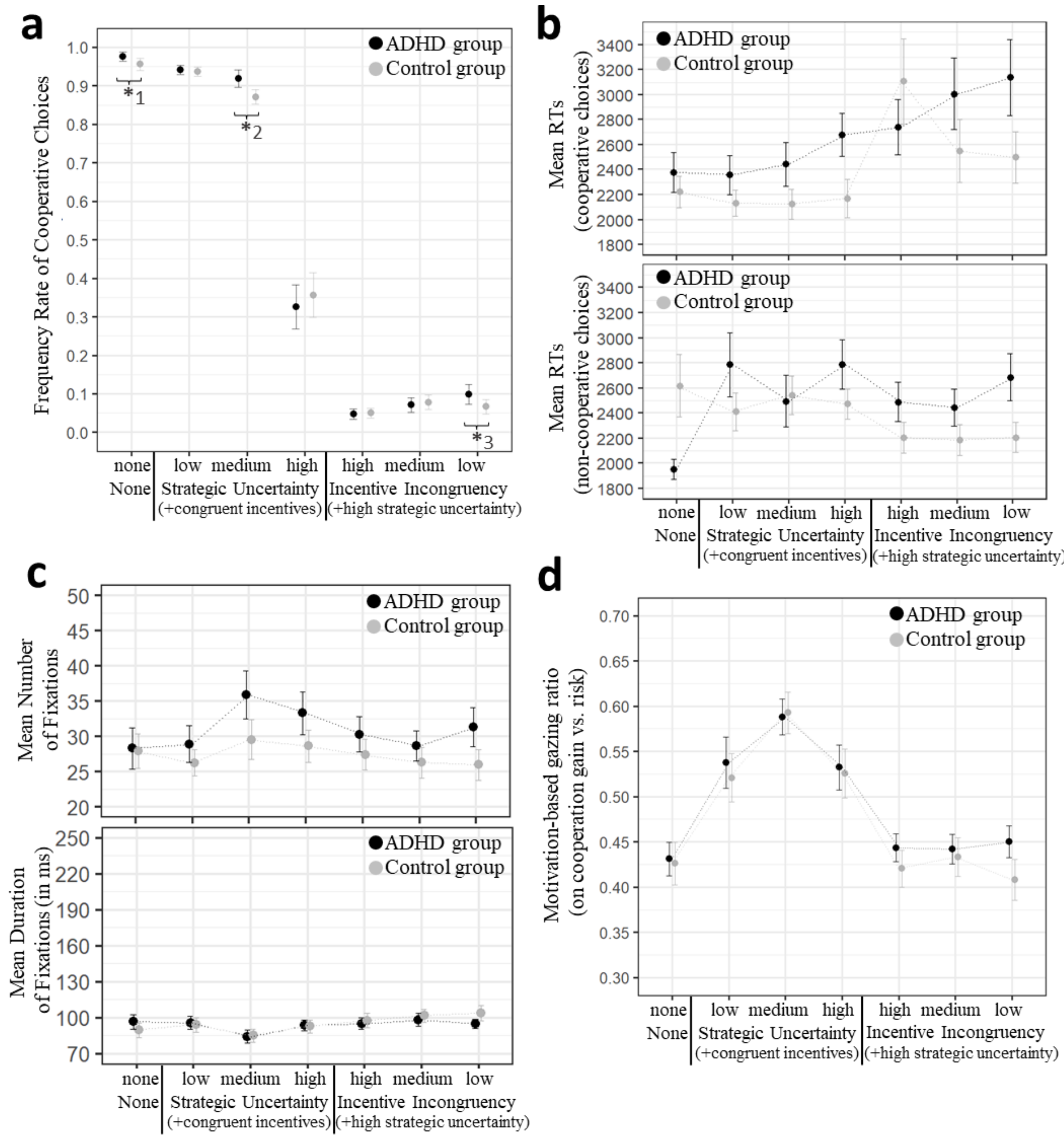

Figure 6. ADHD-specific group differences in performance. ***a.*** Frequency rates of cooperative choices by condition. Shown are the estimated marginal posterior distributions (with posterior means as points and corresponding 95% credible intervals as vertical bars) of the frequency rates from single-trial logistic regression (see Methods). Asterisks refer to group-specific differences detailed in the text. **b.** Mean reaction times (in milliseconds; averaged by subjects) by group for cooperative choices (top panel), and non-cooperative choices (bottom panel) with vertical bars representing standard errors. **c.** Top panel: Average number of fixations per round (averaged by subject) by condition with vertical bars representing standard errors. Bottom panel: Average fixation duration per round (averaged by subject) by condition with vertical bars representing standard errors. **d.** Motivation-based gazing ratio by group and condition with vertical bars representing standard errors.

ADHD-specific differences in latent decision components. The ADHD group had larger drift rates under low incongruent incentives (*1: *BF*=10.45, *PP*=0.91; Figure 7a). This indicates more frequent and faster cooperative choices. Moreover, the ADHD group had larger boundary separations than the control group across all blocks (*1: *BF*=29.30, *PP*=0.97; *2: *BF*=6.26, *PP*=0.86; *3: *BF*=3.66, *PP*=0.79; *4: *BF*=6.70, *PP*=0.87; Figure 7b). This indicates more cautious response strategies, which led to longer (more skewed) RTs and more frequent cooperative choices.

ADHD symptom severity explains individual differences. Higher ADHD symptom severity was associated with larger contextual changes in both boundary separation (Figure 7c) and starting points (Figure 7d). We also refer to Supplementary Figure S8 for a distribution of CAARS-LV scores across all participants.

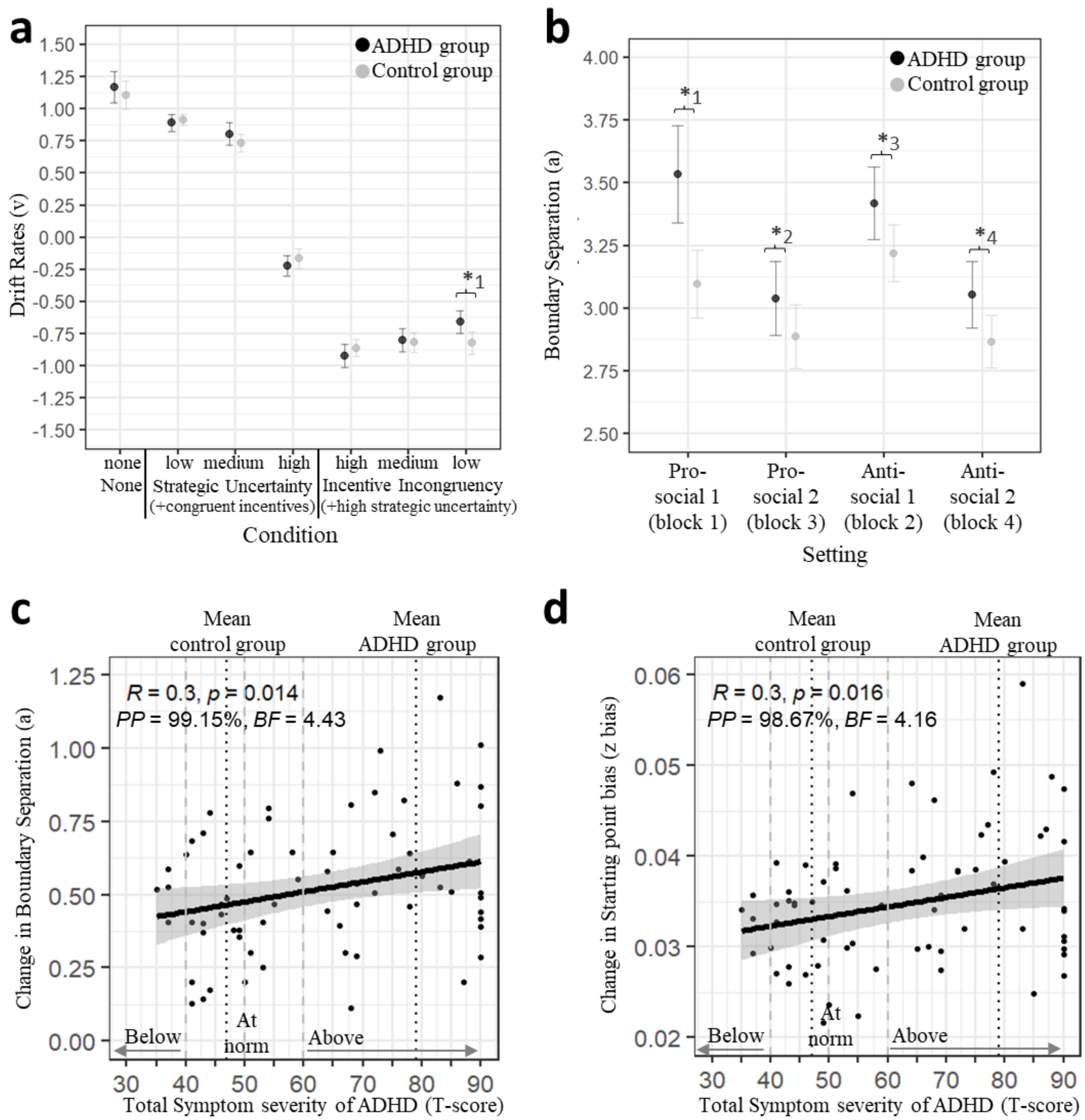

Figure 7. Clinical differences in DDM parameters. a. Drift rates by condition and group (for motivation-based gazing ratio at grand average). Shown are the estimated posterior distributions (with posterior means as points and corresponding 95% credible intervals as vertical bars) of the drift rates. Asterisks refer to group-specific differences detailed in the text. **b.** Boundary separation by setting and group. The numbers in brackets of the x-axis labels refer to the order of presented blocks. Shown are the estimated posterior distributions (with posterior means as points and corresponding 95% credible intervals as vertical bars) of the boundary separations. Asterisks refer to group-specific differences detailed in the text. **c.** Relationship between mean change (absolute values) in boundary separation across settings and ADHD severity scores (T-scores from CAARS-LV). Mean change was calculated by averaging the change in boundary separation between two consecutive blocks. Shown are estimated posterior means (solid lines) and corresponding 95% credible intervals (shaded intervals). **d.**

Relationship between mean change in starting point bias (absolute difference between starting point and 0.5 which represents an unbiased starting point) to unbiased starting point of 0.5 in absolute values across settings and prior outcomes) and ADHD severity scores. Mean change was calculated by averaging the absolute differences across settings and prior outcomes. Shown are estimated posterior means (solid lines) and corresponding 95% credible intervals (shaded intervals).

Discussion

We used an integrated Bayesian hierarchical approach to examine how situational determinants (source and size of motivational conflict, prosociality of the setting), and participants' past experience and cognitive characteristics shaped their cooperativeness. Developing a paradigm with games derived from game theory, we found that the source and magnitude of motivational conflict affected the speed and frequency of cooperative choices. The setting's prosociality and participants' past experience led to systematic biases in the decision-making process that were distinguishable with a process-oriented, computational model analysis. Integrating eye-tracking to measure motivation-based information processing *during* decisions, we found that participants' cooperativeness was higher if they gazed more on the gains of cooperation rather than costs and risks. ADHD characteristics explained some individual differences in responsiveness across contexts. This highlights the clinical importance of studying social-cognitive reactivity with experimental paradigms.

Cooperation across contexts

Cooperative choices were faster than non-cooperative choices under congruent incentives, but slower under incongruent incentives. Focusing on other social behavior (e.g., altruistic giving, group-dynamics in collective decisions), past research suggested that RTs are faster for choices that are more frequent (Capra et al., 2020; Krajbich et al., 2015). Our results do not support this hypothesis, suggesting that this view is too simplistic for understanding cooperation in these games. Under high strategic uncertainty, cooperative choices were less frequent but faster than non-cooperative choices (Figure 4). In the absence of motivational conflict, cooperative choices were more frequent and as fast as non-cooperative choices. Hence, the source of motivational conflict (PD vs. SH games) determined the relative speed of cooperative choices.

Our findings contribute to a century-long question about whether cooperation is instinctively ingrained in humans. In search for answers, prior studies either externally induced stress to evoke instinctive responses and/or they relied on RTs, linking decision speed to distinct mindsets (deliberation versus intuition) based on Kahneman's dual-process theory (De Neys, 2021; Kahneman, 2011). Some studies (Evans & Rand, 2019; Kieslich & Hilbig, 2014; Rand, 2016; Rand et al., 2012; Rubinstein, 2007) found that cooperative choices were faster than non-cooperative choices, while other studies (Capraro & Cococcioni, 2016; Grossmann et al., 2017; Lohse, 2016) found the opposite. Our results suggest that considering the source of motivational conflict might help to reconcile this mixed evidence.

Attentional and cognitive aspects of cooperation

Combining eye-tracking with cooperation games, we found that participants' cooperativeness could be predicted by how much they looked at the gains rather than risks and costs of cooperation. Moreover, the extent to which people looked at cooperation gains varied according to the source and size of conflict (Figure 6d). This interaction supports the idea that distinct patterns of information processing guide subsequent choices (Capra et al., 2020; Chen & Fischbacher, 2016; Chen & Krajbich, 2018; Jiang et al., 2016; Stewart et al., 2016; van Baar et al., 2022). Compared to other studies though, we focused on how participants' cooperativeness changed across contexts using both PD and SH games.

The DDM (Ratcliff, 1978) analysis decomposed the latent decision-making process into quantifiable mental components with established psychological interpretations. Each parameter represents a channel through which situational determinants could influence the decision-making process. This analysis showed that different conflict sources biased the acquisition of information (how much information people gather about cooperation gain versus risks and costs), determining decision difficulty as captured by drift rate (Figure 5a).

Past studies have demonstrated that participants' beliefs about others' intentions and actions influence their prosocial choices (Castro Santa et al., 2018; Ging-Jehli et al., 2020; van Baar et al., 2022), and that people's cooperativeness is a dynamic process that depends on recent experiences (Capraro & Cococcioni, 2015; Nishi et al., 2016). However, the relative importance of the setting and past experiences remained unclear. Our results showed that the prosociality of the setting and previous outcomes affected people's decision-making process by influencing the starting point of decision processes (Figure 5c). In the first prosocial setting, participants were biased towards cooperation regardless of the prior outcome, indicating an optimistic belief about others (Ging-Jehli et al., 2020; Lerner, 1980). These positive biases then turned into negative biases when participants were exposed to antisocial settings. When they subsequently experienced unexpectedly cooperative counterparts, their cooperativeness increased. However, counterparts' unexpected non-cooperation influenced participants' subsequent cooperativeness more strongly than counterparts' unexpected cooperation. Our analytical approach provides avenues for future research to disentangle different situational and individual determinants of cooperation.

Clinical implications

Our work promotes the use of game-theoretical paradigms as a tool for studying clinical social-cognitive characteristics. Compared to the control group, the ADHD group increased cooperation more when the benefits of cooperation increased (Figure 6a). This effect was mediated by more visual attention to cooperation gains rather than risks/costs (Figure 6d). Moreover, the ADHD group accommodated a more cautious response strategy (larger boundary separation across all blocks) than the control group. This might seem counter-intuitive at first since ADHD has been commonly associated with less cautious responses (see for a review: Ging-Jehli et al., 2021). However, most studies have used (neuro)cognitive tests where participants' outcomes did not critically depend on their interactions with others. Considering that individuals with ADHD have more difficulties in developing and maintaining social relationships (Hoza et al., 2005; Landau & Moore, 1991) and a higher propensity for being peer rejected (Craig et al., 2017; Nijmeijer et al., 2008; Uekermann et al., 2010), it seems plausible that they apply more caution when deciding whether to cooperate with counterparts.

The process-oriented analysis with the DDM showed that higher ADHD severity was associated with larger parameter changes across contexts. Hence, reactivity to changes, an important ADHD characteristic noted in a recent review on neurocognitive testing (Ging-Jehli et al., 2021) is quantifiable with our integrative approach. This reactivity to changes became increasingly important at higher severity of ADHD symptoms. Previous studies have examined social-cognitive characteristics of ADHD using primarily emotion recognition tasks or tasks with hypothetically described social problems (Ma et al., 2017; Nijmeijer et al., 2008; Uekermann et al., 2010). None of these studies involved social interactions (participants' outcomes were not affected by others' choices). Mentalizing about others' hypothetical actions might not be the same as making consequential choices based on others' actual choices. Hence, game theoretical derived tasks provide avenues to experimentally study systems for social processes such as cooperation and over-reactivity; characteristics that have not been experimentally studied in clinical populations.

Neuroscientific implications

Using sequential sampling models, together with game theoretical paradigms, provide great potential to dissociate the role of distinct striatal mechanisms involved in value-based decision-making. For instance, it is known that phasic activity in midbrain dopamine neurons encode prediction errors and information about subjective values of presented options (Frank, 2016; Glimcher & Fehr, 2013).

Moreover, cortico-striatal interactions mediate belief updating and adaptive behavior (Westbrook et al., 2021). However, how these neural subprocesses interact to produce behavior has yet to be examined. Process-oriented sequential sampling models can help to formalize this examination because behavior can be decomposed into distinct psychological components (Forstmann et al., 2016; Ging-Jehli et al., 2021; Pedersen et al., 2021). These computational components can then be associated with neural components that express distinct spatiotemporal dynamics.

In our study, prior experience affected behavior through the model parameter (i.e., psychological channel) *starting point bias*. Past studies (Rilling et al., 2002, 2008) found greater activity in caudate and nucleus accumbens after mutual cooperation than unreciprocated cooperation in a PD game. We would therefore expect that greater activity in caudate and nucleus accumbens would be associated with larger starting point biases. Moreover, by recording striatal activity during task performance, future studies could separate the neural subcomponent associated with social prediction errors (i.e., integrating information from prior experience hypothesized to affect starting point bias) from that associated with model parameter *drift rate* (i.e., integrating information about reward and opportunity costs presented in the payoff matrix).

We found that the source of motivational conflict affected behavior through the model parameter *drift rates.* It is known that phasic activity in midbrain dopamine neurons encode information about subjective values of presented options, among others (Glimcher & Fehr, 2013). Moreover, different dynamics of tonic dopaminergic activity has been proposed to signal opportunity costs (Frank, 2016). The use of different game structures, all designed for the study of cooperation, allows one to examine how dynamical neural responses change as motivational trade-offs and opportunity costs systematically change across different game structures (while keeping constant the perceptual object presented to participants). Lastly, since SH games incentivize participants to form an accurate belief about their counterpart to select the optimal response, other considerations are present in the PD games that involve fear of exploitation and greed. Neural differences between SH and PD games could be used to examine how different affects (greed, fear) bias the belief formation process during decision-making.

Summary and future directions

This study showed how an integrative approach composed of eye-tracking, computational modeling, and a new experimental paradigm provided insights for the underlying psychological processes that lead to cooperation in social strategic interactions. We further illustrated the importance of studying clinical characteristics associated with ADHD in simulated strategic interactions (characterized by the fact that participants' outcomes not only depended on their own actions but also on their counterparts' actions). Our study also contributed to a century-long question about whether cooperation is instinctively ingrained in humans. Specifically, we showed that the source and magnitude of motivational conflict affected the relative speed and frequency of cooperative versus non-cooperative choices. The prosociality of a setting and participants' past experience led to distinct biases in their decision-making process. Moreover, cooperativeness was higher if participants gazed more on the gains of cooperation rather than costs and risks.

Our findings have at least four implications for the design of incentive structures for promoting cooperative interactions (at least with artificial agents). First, reducing strategic uncertainty in first encounters is important as it determines whether people incorporate unrelated past outcomes when deciding whether to cooperate. Second, the setting has lasting effects so that participants are less cooperative in antisocial settings even if their counterparts previously demonstrated their cooperativeness. Third, the setting's prosociality and prior outcomes interactively affect decisions. Specifically, unexpected cooperation by others in prosocial settings (or unexpected non-cooperation by others in antisocial settings) induces initial biases (indexed by the DDM parameter starting point) to reciprocate that behavior. Increasing the latency of decisions (e.g., introducing waiting time or decreasing urgency signals for fast responding) mitigate the effects of starting point biases which can therefore lead to less or more cooperation (depending on the setting). Fourth, how past experiences shape future actions depends on characteristics associated with ADHD; as well as on the size and source of motivational conflict.

Future research is needed to address current limitations of this study. For instance, the DDM provided a good account for the data, but other models could do so as well. We found that over-reactivity, as quantified by the larger change in model parameters across contexts, was more prevalent for participants with higher ADHD severity scores. While we would expect rational agents to incorporate past outcomes into their decision-making process, ADHD was associated with extensive generalization based on past outcomes. Future studies are needed to replicate these results, investigating also whether these pronounced parameter changes are linked to stronger emotional responses. Moreover, while we operationalized over-reactivity as a task-evoked response, future research could also consider integrating mood into the working definition of over-reactivity. Our approach could be used as a training tool to teach individuals how to downregulate reactivity to others' responses, or how to incorporate other-regarding concerns into one's own decisions. Moreover, we found that distinct gaze patterns during information processing (i.e., before the decisions), predicted participants' cooperative choices. It remains to be answered whether these distinct gaze patterns are a consequence or a cause of cooperative choices. Future studies could examine this relationship closer by systematically manipulating information processing dynamics during decision-making. Lastly, this study focused on younger adults (aged 18 to 35) and future research is needed to generalize our results for a broader population of individuals with ADHD. For instance, evidence suggests that individuals with ADHD are more likely to drop out of school (Barbaresi et al., 2007; Fried et al., 2016; Mirza et al., 2018) or experience negative academic outcomes (Arnold et al., 2020). Some studies also suggest that ADHD is associated with lower intelligence scores (Kuntsi et al., 2004; Voigt et al., 2006) albeit evidence remains mixed (Jepsen et al., 2009; Wood et al., 2011). Although participants in the present study involved individuals from a broad range of ADHD symptom severities, they were also characterized by intelligence scores above the average and a higher number of years of education. These demographic characteristics make it challenging to generalize our findings to the broader ADHD population. Especially, our sample may not be representative of the cognitive characteristics found in those with lower intelligence scores or who have had fewer educational opportunities. It could be that individuals with ADHD and co-morbid intellectual disabilities and/or fewer educational opportunities show stronger effects in the discussed findings. Alternatively, it could also be that they demonstrate additional qualitative differences (e.g., task strategies) compared to this participant pool. Future studies might examine a more heterogenous sample to assess the extent and nature of potential specific effects to this study sample.

Declaration

Funding

This research was supported by the Swiss National Science Foundation (P1SKP1_184033) awarded to N.R. Ging-Jehli, and Clinical and Translational Science award 8UL18TR000090-05 from the National Center for Translational Sciences to the Ohio State University.

Conflicts of interest/Competing interests

N. R. Ging-Jehli has received research funding from the Swiss National Science Foundation. L. E. Arnold has received research funding from Curemark, Forest, Lilly, Neuropharm, Novartis, Noven, Shire, Supernus, Roche, and YoungLiving (as well as NIH and Autism Speaks), has consulted with CHADD, Gowlings, Neuropharm, Organon, Pfizer, Sigma Tau, Shire, Tris Pharma, and Waypoint, and been on advisory boards for Arbor, Ironshore, Novartis, Noven, Otsuka, Pfizer, Roche, Seaside Therapeutics, Sigma Tau, Shire. This material is based on work performed by T. Van Zandt while serving at the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Ethics approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Insitutional Review Board of The Ohio State University (Study Number: 2020B0085).

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable

Availability of data and materials

Aggregated data and materials are available at OSF (https://osf.io/m5xuh/?view_only=9e36d731ebc0404abd86b631261f6a06) and none of the experiments was preregistered.

Code availability

Modeling Code is uploaded to OSF (https://osf.io/m5xuh/?view_only=9e36d731ebc0404abd86b631261f6a06).

Authors' contributions

N. Ging-Jehli designed and programmed the study design (including experimental paradigms), conducted the analysis and wrote the paper. L.E. Arnold contributed to the study protocol and the write up of the paper. T. Van Zandt contributed to the study setup, the analysis, and the write up of the paper.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2019). *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Ahn, T. K., Ostrom, E., Schmidt, D., Shupp, R., & Walker, J. (2001). Cooperation in PD Games: Fear, Greed, and History of Play. *Public Choice*, *106*(1), 137–155. https://doi.org/10.1023/A:1005219123532

Alós-Ferrer, C., & Garagnani, M. (2020). The cognitive foundations of cooperation. *Journal of Economic Behavior & Organization*, *175*, 71–85. https://doi.org/10.1016/j.jebo.2020.04.019

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders.* (4th ed.). Author.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An Introduction to Bayesian Hypothesis Testing for Management Research. *Journal of Management*, *41*(2), 521–543. https://doi.org/10.1177/0149206314560412

Andreoni, J., & Miller, J. H. (1993). Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. *The Economic Journal*, *103*(418), 570–585. https://doi.org/10.2307/2234532

Arnold, L. E., Hodgkins, P., Kahle, J., Madhoo, M., & Kewley, G. (2020). Long-Term Outcomes of ADHD: Academic Achievement and Performance. *Journal of Attention Disorders*, *24*(1), 73–85. https://doi.org/10.1177/1087054714566076

Axelrod, R. (1967). Conflict of interest: An axiomatic approach. *Journal of Conflict Resolution*, *11*(1), 87–99. https://doi.org/10.1177/002200276701100107

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2007). Long-Term School Outcomes for Children with Attention-Deficit/Hyperactivity Disorder: A Population-

Based Perspective. *Journal of Developmental & Behavioral Pediatrics*, *28*(4), 265. https://doi.org/10.1097/DBP.0b013e31811ff87d

Belloc, M., Bilancini, E., Boncinelli, L., & D'Alessandro, S. (2019). Intuition and Deliberation in the Stag Hunt Game. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-50556-8

Bolton, G. E., Feldhaus, C., & Ockenfels, A. (2016). Social Interaction Promotes Risk Taking in a Stag Hunt Game. *German Economic Review*, *17*(3), 409–423. https://doi.org/10.1111/geer.12095

Bürkner, P.-C. (2017a). Advanced Bayesian Multilevel Modeling with the R Package brms. *arXiv:1705.11123 [Stat]*. http://arxiv.org/abs/1705.11123

Bürkner, P.-C. (2017b). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. (2019). The economics of artificial intelligence: An agenda. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *Artificial intelligence and behavioral economics* (pp. 587–608). The University of Chicago Press.

Capra, C. M., Croson, R. T., Rigdon, M. L., & Rosenblat, T. S. (2020). *Handbook of Experimental Game Theory*. Edward Elgar Publishing.

Capraro, V., & Cococcioni, G. (2015). Social setting, intuition and experience in laboratory experiments interact to shape cooperative decision-making. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1811), 20150237. https://doi.org/10.1098/rspb.2015.0237

Capraro, V., & Cococcioni, G. (2016). Rethinking spontaneous giving: Extreme time pressure and ego-depletion favor self-regarding reactions. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep27219

Capraro, V., Rodriguez-Lara, I., & Ruiz-Martos, M. J. (2020). Preferences for efficiency, rather than preferences for morality, drive cooperation in the one-shot Stag-Hunt game. *Journal of Behavioral and Experimental Economics*, *86*, 101535. https://doi.org/10.1016/j.socec.2020.101535

Castro Santa, J., Exadaktylos, F., & Soto-Faraco, S. (2018). Beliefs about others' intentions determine whether cooperation is the faster choice. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-25926-3

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), Article 11. https://doi.org/10.1038/nn.2925

Chambers, W. J., Puig-Antich, J., Hirsch, M., Paez, P., Ambrosini, P. J., Tabrizi, M. A., & Davies, M. (1985). The Assessment of Affective Disorders in Children and Adolescents by Semistructured Interview: Test-Retest Reliability of the Schedule for Affective Disorders and Schizophrenia for School-Age Children, Present Episode Version. *Archives of General Psychiatry*, *42*(7), 696–702. https://doi.org/10.1001/archpsyc.1985.01790300064008

Champ, R. E., Adamou, M., & Tolchard, B. (2022). Seeking connection, autonomy, and emotional feedback: A self-determination theory of self-regulation in attention-deficit hyperactivity disorder. *Psychological Review*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/rev0000398

Chen, F., & Fischbacher, U. (2016). Response time and click position: Cheap indicators of preferences. *Journal of the Economic Science Association*, *2*(2), 109–126. https://doi.org/10.1007/s40881-016-0026-6

Chen, F., & Krajbich, I. (2018). Biased sequential sampling underlies the effects of time pressure and delay in social decision making. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-05994-9

Clark, K., & Sefton, M. (2001). Repetition and signalling: Experimental evidence from games with efficient equilibria. *Economics Letters*, *70*(3), 357–362. https://doi.org/10.1016/S0165-1765(00)00381-5

Conners, C., Erhardt, D., & Sparrow, E. (1999). *Conners' Adult ADHD Rating Scales (CAARS) technical manual*. Multi-Health Systems.

Cools, R., & D'Esposito, M. (2011). Inverted-U–Shaped Dopamine Actions on Human Working Memory and Cognitive Control. *Biological Psychiatry*, *69*(12), e113–e125. https://doi.org/10.1016/j.biopsych.2011.03.028

Craig, A. B., Grossman, E., & Krichmar, J. L. (2017). Investigation of autistic traits through strategic decision-making in games with adaptive agents. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-05933-6

Crandall, J. W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J.-F., Cebrian, M., Shariff, A., Goodrich, M. A., & Rahwan, I. (2018). Cooperating with machines. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-017-02597-8

Cubillo, A., Halari, R., Smith, A., Taylor, E., & Rubia, K. (2012). A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with Attention Deficit Hyperactivity Disorder (ADHD) and new evidence for dysfunction in adults with ADHD during motivation and attention. *Cortex*, *48*(2), 194–215. https://doi.org/10.1016/j.cortex.2011.04.007

Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2021). Validation of Gazepoint low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01654-x

Dawes, R. M., van de Kragt, A. J. C., & Orbell, J. M. (1990). *Cooperation for the benefit of us—Not me, or my conscience* (p. 110). University of Chicago Press.

De Neys, W. (2021). On Dual- and Single-Process Models of Thinking. *Perspectives on Psychological Science*, *16*(6), 1412–1427. https://doi.org/10.1177/1745691620964172

Doi, T., Fan, Y., Gold, J. I., & Ding, L. (2020). The caudate nucleus contributes causally to decisions that balance reward and uncertain visual information. *eLife*, *9*, e56694. https://doi.org/10.7554/eLife.56694

Duffy, J., & Feltovich, N. (2006). Words, Deeds, and Lies: Strategic Behaviour in Games with Multiple Signals. *The Review of Economic Studies*, *73*(3), 669–688. https://doi.org/10.1111/j.1467-937X.2006.00391.x

Durstewitz, D., & Seamans, J. K. (2008). The Dual-State Theory of Prefrontal Cortex Dopamine Function with Relevance to Catechol-O-Methyltransferase Genotypes and Schizophrenia. *Biological Psychiatry*, *64*(9), 739–749. https://doi.org/10.1016/j.biopsych.2008.05.015

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General*, *144*(5), 951–966. https://doi.org/10.1037/xge0000107

Evans, A. M., & Rand, D. G. (2019). Cooperation and decision time. *Current Opinion in Psychology*, *26*, 67–71. https://doi.org/10.1016/j.copsyc.2018.05.007

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Frank, M. J. (2016). Computational Cognitive Neuroscience Approaches to Deconstructing Mental Function and Dysfunction. In A. D. Redish & J. D. Gordon, *Computational Psychiatry: New Perspectives on Mental Illness* (Vol. 20, pp. 101–120). MIT Press.

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, *120*(3), 497–517. https://doi.org/10.1037/0735-7044.120.3.497

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold Your Horses: Impulsivity, Deep Brain Stimulation, and Medication in Parkinsonism. *Science*, *318*(5854), 1309–1312. https://doi.org/10.1126/science.1146157

Frank, M. J., Santamaria, A., O'Reilly, R. C., & Willcutt, E. (2007). Testing Computational Models of Dopamine and Noradrenaline Dysfunction in Attention Deficit/Hyperactivity Disorder. *Neuropsychopharmacology*, *32*(7), Article 7. https://doi.org/10.1038/sj.npp.1301278

Fried, R., Petty, C., Faraone, S. V., Hyder, L. L., Day, H., & Biederman, J. (2016). Is ADHD a Risk Factor for High School Dropout? A Controlled Study. *Journal of Attention Disorders*, *20*(5), 383–389. https://doi.org/10.1177/1087054712473180

Gazepoint. (2021). *GP3 Eye-Tracker*. GP3 Eye-Tracker. Retrieved from https://www.gazept.com

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211. https://doi.org/10.1080/19345747.2011.618213

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472.

Ginapp, C. M., Greenberg, N. R., Macdonald-Gagnon, G., Angarita, G. A., Bold, K. W., & Potenza, M. N. (2023). The experiences of adults with ADHD in interpersonal relationships and online communities: A qualitative study. *SSM - Qualitative Research in Health*, *3*, 100223. https://doi.org/10.1016/j.ssmqr.2023.100223

Ging-Jehli, N. R., & Ratcliff, R. (2020). Effects of aging in a task-switch paradigm with the diffusion decision model. *Psychology and Aging*, *35*(6), 850–865. https://doi.org/10.1037/pag0000562

Ging-Jehli, N. R., Ratcliff, R., & Arnold, L. E. (2021). Improving neurocognitive testing using computational psychiatry—A systematic review for ADHD. *Psychological Bulletin*, *147*(2), 169–231. https://doi.org/10.1037/bul0000319

Ging-Jehli, N. R., Schneider, F. H., & Weber, R. A. (2020). On self-serving strategic beliefs. *Games and Economic Behavior*, *122*, 341–353. https://doi.org/10.1016/j.geb.2020.04.016

Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision Making and the Brain*. Academic Press.

Goetze, D. (1994). Comparing Prisoner's Dilemma, Commons Dilemma, and Public Goods Provision Designs in Laboratory Experiments. *Journal of Conflict Resolution*, *38*(1), 56–86. https://doi.org/10.1177/0022002794038001004

Grossmann, I., Brienza, J. P., & Bobocel, D. R. (2017). Wise deliberation sustains cooperation. *Nature Human Behaviour*, *1*(3), Article 3. https://doi.org/10.1038/s41562-017-0061

Haruno, M., & Kawato, M. (2009). Activity in the Superior Temporal Sulcus Highlights Learning Competence in an Interaction Game. *Journal of Neuroscience*, *29*(14), 4542–4547. https://doi.org/10.1523/JNEUROSCI.2707-08.2009

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hoza, B., Mrug, S., Gerdes, A. C., Hinshaw, S. P., Bukowski, W. M., Gold, J. A., Kraemer, H. C., Pelham Jr., W. E., Wigal, T., & Arnold, L. E. (2005). What Aspects of Peer Relationships Are Impaired in Children With Attention-Deficit/Hyperactivity Disorder? *Journal of Consulting and Clinical Psychology*, *73*(3), 411–423. https://doi.org/10.1037/0022-006X.73.3.411

Isler, O., Gächter, S., Maule, A. J., & Starmer, C. (2021). Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-93412-4

Isler, O., Yilmaz, O., & John Maule, A. (2021). Religion, parochialism and intuitive cooperation. *Nature Human Behaviour*, *5*(4), Article 4. https://doi.org/10.1038/s41562-020-01014-3

Isoda, M., & Hikosaka, O. (2008). Role for Subthalamic Nucleus Neurons in Switching from Automatic to Controlled Eye Movement. *Journal of Neuroscience*, *28*(28), 7209–7218. https://doi.org/10.1523/JNEUROSCI.0487-08.2008

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. John Wiley & Sons.

Jaskir, A., & Frank, M. J. (2023). On the normative advantages of dopamine and striatal opponency for learning and choice. *eLife*, *12*, e85107. https://doi.org/10.7554/eLife.85107

Jepsen, J. R. M., Fagerlund, B., & Mortensen, E. L. (2009). Do Attention Deficits Influence IQ Assessment in Children and Adolescents With ADHD? *Journal of Attention Disorders*, *12*(6), 551–562. https://doi.org/10.1177/1087054708322996

Jiang, T., Potters, J., & Funaki, Y. (2016). Eye-tracking Social Preferences. *Journal of Behavioral Decision Making*, *29*(2–3), 157–168. https://doi.org/10.1002/bdm.1899

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

Kangas, B. D., Berry, M. S., Cassidy, R. N., Dallery, J., Vaidya, M., & Hackenberg, T. D. (2009). Concurrent performance in a three-alternative choice situation: Response allocation in a Rock/Paper/Scissors game. *Behavioural Processes*, *82*(2), 164–172. https://doi.org/10.1016/j.beproc.2009.06.004

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making*, *9*(6), 510–522.

Kocher, M. G., Martinsson, P., Myrseth, K. O. R., & Wollbrant, C. E. (2017). Strong, bold, and kind: Self-control and cooperation in social dilemmas. *Experimental Economics*, *20*(1), 44–69. https://doi.org/10.1007/s10683-015-9475-7

Koenigs, M., & Tranel, D. (2007). Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game. *Journal of Neuroscience*, *27*(4), 951–956. https://doi.org/10.1523/JNEUROSCI.4606-06.2007

Krain, A. L., & Castellanos, F. X. (2006). Brain development and ADHD. *Clinical Psychology Review*, *26*(4), 433–444. https://doi.org/10.1016/j.cpr.2006.01.005

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), Article 10. https://doi.org/10.1038/nn.2635

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*(1), Article 1. https://doi.org/10.1038/ncomms8455

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Kumar, U., Arya, A., & Agarwal, V. (2022). Altered functional connectivity in children with ADHD while performing cognitive control task. *Psychiatry Research: Neuroimaging*, *326*, 111531. https://doi.org/10.1016/j.pscychresns.2022.111531

Kuntsi, J., Eley, T. c., Taylor, A., Hughes, C., Asherson, P., Caspi, A., & Moffitt, T. e. (2004). Co-occurrence of ADHD and low IQ has genetic origins. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *124B*(1), 41–47. https://doi.org/10.1002/ajmg.b.20076

Landau, S., & Moore, L. A. (1991). Social Skill Deficits in Children with Attention-Deficit Hyperactivity Disorder. *School Psychology Review*, *20*(2), 235–251. https://doi.org/10.1080/02796015.1991.12085548

Lerner, M. J. (1980). *The belief in a just world. In The Belief in a just World*. Springer.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, *121*(1), 66–95. https://doi.org/10.1037/a0035230

Lohse, J. (2016). Smart or selfish – When smart guys finish nice. *Journal of Behavioral and Experimental Economics*, *64*, 28–40. https://doi.org/10.1016/j.socec.2016.04.002

Ma, I., Lambregts-Rommelse, N. N. J., Buitelaar, J. K., Cillessen, A. H. N., & Scheres, A. P. J. (2017). Decision-making in social contexts in youth with ADHD. *European Child & Adolescent Psychiatry*, *26*(3), 335–344. https://doi.org/10.1007/s00787-016-0895-5

Mahon, P. G., & Canosa, R. L. (2012). Prisoners and chickens: Gaze locations indicate bounded rationality. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 401–404. https://doi.org/10.1145/2168556.2168647

March, C. (2021). Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *Journal of Economic Psychology*, *87*, 102426. https://doi.org/10.1016/j.joep.2021.102426

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315372495

McLoughlin, G., Palmer, J. A., Rijsdijk, F., & Makeig, S. (2014). Genetic Overlap between Evoked Frontocentral Theta-Band Phase Variability, Reaction Time Variability, and Attention-Deficit/Hyperactivity Disorder Symptoms in a Twin Study. *Biological Psychiatry*, *75*(3), 238–247. https://doi.org/10.1016/j.biopsych.2013.07.020

Mengel, F. (2018). Risk and Temptation: A Meta-study on Prisoner's Dilemma Games. *The Economic Journal*, *128*(616), 3182–3209. https://doi.org/10.1111/ecoj.12548

Mirza, H., Roberts, E., AL-Belushi, M., AL-Salti, H., AL-Hosni, A., Jeyaseelan, L., & AL-Adawi, S. (2018). School Dropout and Associated Factors Among Omani Children with Attention-Deficit Hyperactivity Disorder: A Cross-Sectional Study. *Journal of Developmental & Behavioral Pediatrics*, *39*(2), 109. https://doi.org/10.1097/DBP.0000000000000522

Moisan, F., ten Brincke, R., Murphy, R. O., & Gonzalez, C. (2018). Not all Prisoner's Dilemma games are equal: Incentives, social preferences, and cooperation. *Decision*, *5*(4), 306–322. https://doi.org/10.1037/dec0000079

Moolchand, P., Jones, S. R., & Frank, M. J. (2022). Biophysical and Architectural Mechanisms of Subthalamic Theta under Response Conflict. *Journal of Neuroscience*, *42*(22), 4470–4487. https://doi.org/10.1523/JNEUROSCI.2433-19.2022

Murphy, R. O., & Ackermann, K. A. (2015). Social preferences, positive expectations, and trust based cooperation. *Journal of Mathematical Psychology*, *67*, 45–50. https://doi.org/10.1016/j.jmp.2015.06.001

Nijmeijer, J. S., Minderaa, R. B., Buitelaar, J. K., Mulligan, A., Hartman, C. A., & Hoekstra, P. J. (2008). Attention-deficit/hyperactivity disorder and social dysfunctioning. *Clinical Psychology Review*, *28*(4), 692–708. https://doi.org/10.1016/j.cpr.2007.10.003

Nishi, A., Christakis, N. A., Evans, A. M., O'Malley, A. J., & Rand, D. G. (2016). Social Environment Shapes the Speed of Cooperation. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep29622

Ostrom, E. (2000). Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives*, *14*(3), 137–158. https://doi.org/10.1257/jep.14.3.137

Pedersen, M. L., & Frank, M. J. (2020). Simultaneous Hierarchical Bayesian Parameter Estimation for Reinforcement Learning and Drift Diffusion Models: A Tutorial and Links to Neural Data. *Computational Brain & Behavior*, *3*(4), 458–471. https://doi.org/10.1007/s42113-020-00084-w

Pedersen, M. L., Ironside, M., Amemori, K., McGrath, C. L., Kang, M. S., Graybiel, A. M., Pizzagalli, D. A., & Frank, M. J. (2021). Computational phenotyping of brain-behavior dynamics underlying approach-avoidance conflict in major depressive disorder. *PLOS Computational Biology*, *17*(5), e1008955. https://doi.org/10.1371/journal.pcbi.1008955

Poundstone, W. (1993). *Prisoner's Dilemma/John Von Neumann, Game Theory and the Puzzle of the Bomb*. Anchor Books.

R Core Team. (2021). *R: A language and environment for statistical computing* (4.1.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rand, D. G. (2016). Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation. *Psychological Science*, *27*(9), 1192–1206. https://doi.org/10.1177/0956797616654455

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), Article 7416. https://doi.org/10.1038/nature11467

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*(1), Article 1. https://doi.org/10.1038/ncomms4677

Rankin, F. W., Van Huyck, J. B., & Battalio, R. C. (2000). Strategic Similarity and Emergent Conventions: Evidence from Similar Stag Hunt Games. *Games and Economic Behavior*, *32*(2), 315–337. https://doi.org/10.1006/game.1999.0711

Rapoport, A. (1987). Research paradigms and expected utility models for the provision of step-level public goods. *Psychological Review*, *94*(1), 74–83. https://doi.org/10.1037/0033-295X.94.1.74

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., Murdock, C. D., & Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, *46*(5), 1256–1266. https://doi.org/10.1016/j.neuropsychologia.2007.11.033

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A Neural Basis for Social Cooperation. *Neuron*, *35*(2), 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Rubia, K., Cubillo, A., Woolley, J., Brammer, M. J., & Smith, A. (2011). Disorder-specific dysfunctions in patients with attention-deficit/hyperactivity disorder compared to patients with obsessive-compulsive disorder during interference inhibition and attention allocation. *Human Brain Mapping*, *32*(4), 601–611. https://doi.org/10.1002/hbm.21048

Rubinstein, A. (2007). Instinctive and Cognitive Reasoning: A Study of Response Times. *The Economic Journal*, *117*(523), 1243–1259. https://doi.org/10.1111/j.1468-0297.2007.02081.x

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71–78. https://doi.org/10.1145/355017.355028

Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, *86*(5), 668–679. https://doi.org/10.1037/0022-3514.86.5.668

Shubik, M., Wolf, G., & Poon, B. (1974). Perception of Payoff Structure and Opponent's Behavior in Related Matrix Games. *Journal of Conflict Resolution*, *18*(4), 646–655. https://doi.org/10.1177/002200277401800406

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Snidal, D. (1985). Coordination versus Prisoners' Dilemma: Implications for International Cooperation and Regimes. *American Political Science Review*, *79*(4), 923–942. https://doi.org/10.2307/1956241

Spiliopoulos, L., & Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics*, *21*(2), 383–433. https://doi.org/10.1007/s10683-017-9528-1

Stan Development Team. (2021). *Stan Modeling Language Users Guide and Reference Manual* (2.28) [Computer software]. https://mc-stan.org

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye Movements in Strategic Choice. *Journal of Behavioral Decision Making*, *29*(2–3), 137–156. https://doi.org/10.1002/bdm.1901

Suleiman, R., Budescu, D., Messick, D. M., & Fischer, I. (2004). *Contemporary approaches to social dilemma research*. Cambridge University Press.

Taylor, M. (1987). *The possibility of cooperation*. Cambridge University Press.

The Math Works, Inc. (2020). *MATLAB* (Version 2021a) [Computer software]. The Math Works, Inc.

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30. https://doi.org/10.1037/bul0000217

Uekermann, J., Kraemer, M., Abdel-Hamid, M., Schimmelmann, B. G., Hebebrand, J., Daum, I., Wiltfang, J., & Kis, B. (2010). Social cognition in attention-deficit hyperactivity disorder

(ADHD). *Neuroscience & Biobehavioral Reviews*, *34*(5), 734–743. https://doi.org/10.1016/j.neubiorev.2009.10.009

van Baar, J. M., Nassar, M. R., Deng, W., & FeldmanHall, O. (2022). Latent motives guide structure learning during adaptive social choice. *Nature Human Behaviour*, *6*(3), Article 3. https://doi.org/10.1038/s41562-021-01207-4

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Van Lange, P. A. M., Balliet, D., Parks, C. D., & Van Vugt, M. (2014). *Social Dilemmas The Psychology of Human Cooperation*. Oxford University Press.

van Schouwenburg, M. R., Onnink, A. M. H., ter Huurne, N., Kan, C. C., Zwiers, M. P., Hoogman, M., Franke, B., Buitelaar, J. K., & Cools, R. (2014). Cognitive flexibility depends on white matter microstructure of the basal ganglia. *Neuropsychologia*, *53*, 171–177. https://doi.org/10.1016/j.neuropsychologia.2013.11.015

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, *2*(1), 20–54. https://doi.org/10.3758/BF03214411

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Voigt, R. G., Barbaresi, W. J., Colligan, R. C., Weaver, A. L., & Katusic, S. K. (2006). Developmental dissociation, deviance, and delay: Occurrence of attention-deficit–hyperactivity disorder in individuals with and without borderline-to-mild intellectual disability. *Developmental Medicine and Child Neurology*, *48*(10), 831–835. https://doi.org/10.1017/S0012162206001782

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

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Weber, R. A. (2018). *Organizational Coordination: A Game-Theoretic View*. https://doi.org/10.1184/R1/6571286.v1

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence–second edition*. NCS Pearson.

Westbrook, A., Frank, M. J., & Cools, R. (2021). A mosaic of cost–benefit control over cortico-striatal circuitry. *Trends in Cognitive Sciences*, *25*(8), 710–721. https://doi.org/10.1016/j.tics.2021.04.007

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*, *120*(2), 329–355. https://doi.org/10.1037/a0031542

Wood, A. C., Rijsdijk, F., Johnson, K. A., Andreou, P., Albrecht, B., Arias-Vasquez, A., Buitelaar, J. K., McLoughlin, G., Rommelse, N. N. J., Sergeant, J. A., Sonuga-Barke, E. J. S., Uebel, H., Meere, J. J. van der, Banaschewski, T., Gill, M., Manor, I., Miranda, A., Mulas, F., Oades, R. D., … Kuntsi, J. (2011). The relationship between ADHD and key cognitive phenotypes is not mediated by shared familial effects with IQ. *Psychological Medicine*, *41*(4), 861–871. https://doi.org/10.1017/S003329171000108X

Ziegler, S., Pedersen, M. L., Mowinckel, A. M., & Biele, G. (2016). Modelling ADHD: A review of ADHD theories through their predictions for computational models of decision-making and reinforcement learning. *Neuroscience & Biobehavioral Reviews*, *71*, 633–656. https://doi.org/10.1016/j.neubiorev.2016.09.002