Perceived Intentionality Determines when Cooperation is Intuitive*

Juana Castro Santa[1], Filippos Exadaktylos[2,3] & Salvador Soto-Faraco[1,4]

1. Multisensory Research Group, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain
2. Department of Economics, Middlesex University Business School, London, UK
3. School of Agriculture Policy and Development, University of Reading, Reading, UK
4. Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Abstract

Is collaboration the intuitive default for humans? Previous research using Response Times (RTs) as a marker of intuitive and deliberate cognitive processes has been inconclusive. A failure to account for conflict between choices has been suggested to underlie the contradictory results. Here, we introduce the concept of Perceived Intentionality (PI), i.e. beliefs regarding a partner's intentions to cooperate. In a social dilemma game, we show that PI acts as a switch that strongly determines cooperation and defection RTs. Importantly, the effect of choice conflict was drastically reduced once PI and individual differences were considered jointly. We conclude that (a) cooperation is not always intuitive but it depends on PI; (b) the role of choice conflict on RTs is not as important as previously suggested; and c) RTs, though correlational in nature, can still be a useful indicator of cognitive processes, provided that the social context is taken into account.

* We gratefully acknowledge helpful comments from Antonio Espín, David Rand, James Enns and Pablo Brañas-Garza

Human cooperation –sacrificing individual resources to achieve higher collective welfare– poses an evolutionary puzzle in both social and natural sciences: even if selfishness leads to higher evolutionary fitness, cooperation is nevertheless widespread among humans (Axelrod, 1984; Ostrom, 1990). Recently, a case has been made that cooperation emerges as an intuitive impulse and is overridden only when humans employ a deliberate reasoning mode that typically favours selfish outcomes (Gächter, 2012; Rand et al., 2012). This thesis is based on Dual Process Theory, which assumes two separate processes that compete at a cognitive level to guide individual decisions; *intuition* (also referred to as 'System I'), which is a fast and automatic process, and *reflection* ('System II'), which is a slow and controlled process, whereby an individual tries to consider all possible outcomes before reaching a decision (Kahneman, 2011; Stanovich & West, 1998). Applied to the case of cooperation, this paradigm assumes that cooperative behaviour is guided mainly by intuitive, System I, processes whilst reflection shifts individuals' decisions toward selfish choices.

To test this proposition, studies have measured the Response Times (RT henceforth) of choices in social dilemma decisions (i.e. situations where private and social interests are in conflict) to infer relative use of intuition over deliberation. The idea is that intuitive decisions should be faster to make, while deliberate ones demand more time, thus producing longer RTs. However, this approach has produced somehow inconsistent results. While several studies indeed found that individuals make cooperation choices faster than selfish ones (Cappelen et al., 2016; Lotito et al., 2013; Nielsen et al., 2014; Rand et al., 2012; Rubinstein, 2007), others reported the opposite pattern, i.e. that selfish choices are faster on average (Fiedler et al., 2013; Lohse et al., 2017; Piovesan & Wengström, 2009).

Based on these inconsistencies, it has been suggested that RTs might not reflect relative use of intuition and reflection but merely decision conflict, or discriminability (i.e. strength of preference over available options); the more similar the available response options, the longer participants need to make up their mind (Evans et al., 2014; Evans & van de Calseyde, 2017; Krajbich et al., 2015; Nishi et al., 2017). Therefore, what seems to index the use of the slow, deliberate system may instead just reflect complexity of the choice problem, which once controlled for, makes cooperation and selfish decisions indistinguishable in their latencies. Such an argument, along with the mixed results, cast doubt not only on whether RTs are an adequate methodological tool for identifying competing mental functions, but also on whether cooperative acts are intuitive in humans.

Findings from various disciplines employing different methodologies such as time pressure (Cappelletti et al., 2011; Rand et al., 2012, 2014; Rand & Kraft-Todd, 2014), ego depletion methods (Cornelissen et al., 2011; Døssing ey al, 2017; Roch et al., 2000; Schulz et al., 2014), as well as structural and functional neural correlates (Fermin et al., 2016; Tabibnia et al., 2008; Yamagishi et al., 2016, 2017) provide the same picture; while most of the studies find that intuition favours cooperation, there seems to be a *consistent inconsistency* since some studies either fail to replicate the results or find results in the opposite direction (for time pressure see: (Bouwmeester et al., 2017; Tinghög et al., 2013; Verkoeijen et al., 2014); for ego depletion: (Hauge et al., 2016); for neuroeconomics: (Knoch et al., 2006; Steinbeis et al., 2012). The literature, therefore, leaves little room for doubt that at least in *some occasions*, cooperation is indeed the result of intuition, while discriminability or the use of RTs cannot fully account for it (for a review see Zaki & Mitchell, 2013, and for a meta-analysis see Rand, 2016). For instance, toddlers, and even infants whose deliberate system has not yet developed fully, spontaneously help others who are visibly struggling with a task (Tomasello, 2009). The key question therefore is not whether the intuitive default is cooperation or selfishness, but rather when and under which circumstances each one is. Here, we contend that above and beyond stimulus-related factors such as payoff, the social context where the choices are made affects the intuitiveness of cooperative vs. selfish actions. Identifying such contextual factors is a fundamental pre-requisite for understanding the human cooperation puzzle. In addition, considering contextual influences might be crucial in order to reconcile the apparently contradictory past results when using RTs as a marker of intuitive versus reflective processes.

The present study focuses on one such factor, namely the *Perceived Intentionality* of others. PI refers to the intentions, beliefs and desires an individual assigns to other actors involved in a given social situation (Dennett, 1987; Tomasello et al., 2005). PI is an integral part of the human decision-making process, intuitive or not, whereby the decision-maker evaluates reality before making a choice. It involves judging mental states and accumulating sensory information on the basis that all other actors involved have their own reference values, intentions and goals and also the ability to change the environment. PI is then used to guide behaviour. For example, people will pass on offering help when seeing someone lets an object fall on purpose, instead of accidently dropping it (Warneken & Tomasello, 2006).

The effects of PI on behaviour are well-known (Barnes et al., 1979; Chuah et al., 2016; Hoffman et al., 2015; Rand et al., 2015) and might well regulate which behaviours are intuitive in a dynamic way. For example, the decision to give way to a customer in the supermarket queue, or give money to a beggar, can be the intuitive choice, or it can be subject to controlled deliberation. The issue will depend on the perceived intentions regarding the behaviour of the other party. Is s/he truly in a hurry? Is s/he really in need of food? In the case of social dilemmas, PI corresponds to the belief a given individual holds about the likelihood that another individual involved in an interaction will be cooperative or selfish. A high PI in a social dilemma, i.e. attributing to the other party a high chance of acting cooperatively, might create a social context where cooperation is the intuitive default. On the contrary, everything else equal, holding a low PI might give rise to a different attribution and the adoption of a different default (i.e., defecting is the intuitive thing to do). Under such a framework, cooperation might be intuitive or not, depending on the PI and, importantly, it can help reconcile the seemingly contradictory results from different studies if, for any reason, they induced different PI levels.

Methods

To address the framework laid out above, we measured the effect of PI on choice RTs in a social dilemma, while controlling for the effect of similarity among choices. We manipulated the PI of cooperation of the game partner with explicit cues in a Prisoners' Dilemma game. Participants made a series of binary choices between a cooperative and a selfish option, with the latter always corresponding to higher expected payoffs. Participants were told their partner will be randomly chosen on each round. Before each choice, participants were presented with the (new) partners' PI (average probability of cooperation over all rounds, 0%, 20%, 50%, 80% and 100%) and the expected returns from cooperation/defection in that trial (Figure 1). We employed a random-effects model to estimate response latency (*log(RT))* as a function of *PI* and discriminability, while controlling for other potential sources of variability. This allowed us to estimate the *net* effects of PI and discriminability on the RTs. Our hypothesis is that PI affects whether cooperation or defection is the more intuitive choice, based on the RTs, for each level of discriminability.

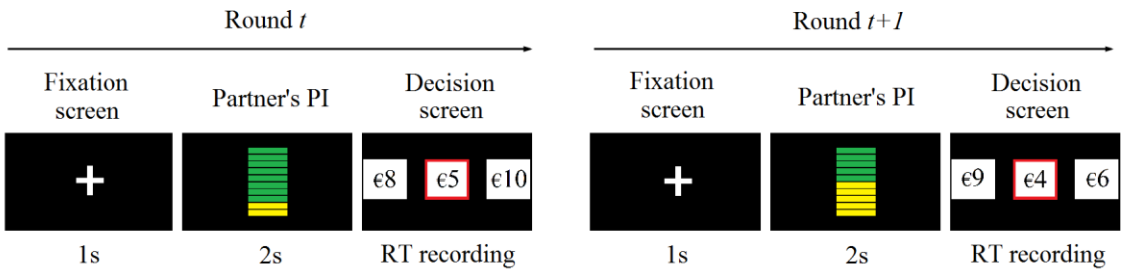

Figure 1. Schematic trial sequence. The figure shows two consecutive trials. Subjects were presented with 100 such trials (in addition to 5 training rounds). After a fixation screen, subjects were presented with a representation of the overall partner's intentions for two seconds. Subsequently, they were presented with the decision screen which comprised the two choices (a low and a high monetary value, presented on the left and right), plus a penalty/reward value (in the centre). Choosing the low value always led to higher expected monetary returns and corresponded to defection strategy. Choosing the high value was the cooperative strategy. RTs were recorded from the presentation of the decision screen until the moment the participants pressed the key denoting their decision.

Participants and procedures: Students from Pompeu Fabra University (UPF) were recruited via e-mail (N=50, average age 22.32; 48% female), from a participants' database. All of them were native Spanish speakers, or Spanish-Catalan bilinguals. The experiment was conducted in the CBC Lab of UPF, lasted approximately 1 hour, and 6 or 8 participants were tested in each session (in individual booths to prevent visual contact or any communication between them). They were paid in cash at the end of the session in private (mean payoff was €10,0 including a show-up fee of €3). The experiment was approved by the UPF Ethics committee and all participants provided written informed consent prior to their participation. The protocol was custom programmed using Psychotoolbox libraries (running in Matlab R2015B).

Prisoner's Dilemma game: Participants played a discrete version of the Traveller's Dilemma Game (Basu, 1994), which is in essence a Prisoner's Dilemma Game. The game was presented as a coordination game, whereby two players are asked to choose between two prices, low and high. If both participants choose the same price, either low or high, then they both end up with a payoff equal to that price. However, if one player chooses low and the other high, then the one who chooses low ends up with a payoff equal to the low plus a "reward", which is represented by a third price (such that the sum of the reward and the low price exceeds the high price). The one who chooses the high price, ends us with a payoff that equals the low price minus a "penalty", which is equal to the "reward", i.e. the third price. Therefore, the high price corresponded to cooperation and the low price corresponded to defection, the latter always exhibited higher expected monetary payoffs, irrespectively of the partner's decision (an English translation of the instructions is available in the Supplementary Information).

PI, decisions and RT recording: Subjects played a total of 100 rounds, in addition to 5 practice rounds. They were led to believe that they were playing against each other. At the beginning of each round participants received a visual cue about the likely strategy adopted by their partner assigned in that round (i.e. their partner's PI), represented with a graphic display of the ratio of trials the partner player had chosen to cooperate overall in the game, similar to a battery charge indicator (see Figure 1). Following the cue, a payment structure was shown for her decision. Both the cue to PI and the payment structure were selected randomly without replacement from the pre-defined payment matrix, on a trial by trial basis. PI could be 0% (partner never cooperates), 20%, 50%, 80% or 100% (partner always cooperates) with equal probability. That is, each participant faced a total of 20 rounds for each PI level. RTs were recorded from the moment the payment structure was displayed on the screen until a response key associated to a decision was pressed. One round was randomly selected at the end of the game to determine payoffs.

Statistical analysis. We conducted a panel data analysis that takes each participant as the unit and each round as time. A Breusch-Pagan test confirmed the superiority of a random effects model over a pooled OLS one (p<0.001), and the Hausman test confirmed the superiority of a random effects model over a fixed effects model (p=0.1262), giving equally consistent but more efficient estimators. Hence, we used a random effects model. After excluding incomplete observations, and one outlier (RT>mean RT + 3sd), we were left with 4600 observations (46 individuals ***x*** 100 decisions). Our model estimates *log(RT).* The natural logarithm is used due to the heavily right-skewed decision time's data (Howell, 2016). The basic predictors are *PI* and discriminability, the latter indexed as the difference in expected value terms (PI ***x*** return) between the chosen and the unchosen option, which we denote as *dEV*. *Coop* is included in the model as a binary variable taking the value of 1 if the choice has been cooperation and 0 if the choice has been defection. We expect by hypothesis that cooperation and defection choices will follow opposite trends with respect to PI, hence we also include the corresponding interaction term (*Coop* ***x*** *PI*). The interaction of *Coop* with *dEV* is also included because *dEV* takes positive values for defection decisions and negative values for cooperation decisions, since by design, defection had always higher expected value than cooperation. Therefore, discriminability for defection choices will translate in a negative coefficient of *dEV* on *log(RT)* and in a positive one for cooperation choices. Lastly, two further terms are used to capture well known sources of variability: sequential variations in RT across the experiment (variable *Round*, indicating the trial position in the experiment sequence, from 1 to 100), and individual differences in response latency (variable $Ind.RT_i$, measured as the mean average RT of each individual across all 100 decisions). In sum, we estimated the following model:

$$log(RT)_{it} = \beta_0 + \beta_1 PI_t + \beta_2 dEV_t + \beta_3 Coop_t + \beta_4 PI_t \, \boldsymbol{x} \, Coop_t + \beta_5 dEV_t \, \boldsymbol{x} \, Coop_t + \beta_5 Round_t + \beta_6 Ind.RT_i + u_i + \varepsilon_{it},$$

where i indexes participants, t indexes decision choices and u_i and ε_{it} are the between- and within-participants error terms respectively. To analyze the potential effect of individual characteristics we added in the model the individual level variable $IndCoop_i$ denoting the individual tendency to response to difference in PI. Therefore, we estimated a second model where we included $IndCoop_i$ and its interaction with $Coop_t$ as predictors of $log(RT)_{it}$. In a third model, we explored the role of $IndCoop_i$ as a potential moderator of the effect of *PI* on $log(RT)_{it}$ and thus we included the interaction terms $IndCoop_i$ **x** PI_t and $IndCoop_i$ **x** PI_t **x** $Coop_t$. Analyses were performed with STATA 14 statistical software.

Results

Overall, participants chose to cooperate 21.2% of the time, and the latency of cooperation choices (mean=2.97 sec, sd=3.32) was longer on average than that of selfish choices (mean=2.65 sec, sd=2.56; z=4.69, p< .001). To study the role of discriminability, we used the difference in expected value (return) between the chosen and the unchosen option, which we denote as *dEV* (for similar methodology, see Krajbich et al., 2015). Our data confirms previous results: discriminability, or similarity of options in monetary terms, drives RTs up. As seen from Figure 2, mean observed response times are on average higher, the closer *dEV* is to zero, a point indicating equal payoffs between cooperation and defection options in expected return. The regression analysis confirms the result (Model 1, Table S1): everything else equal, the effect of *dEV* on *logRT* is positive for cooperation choices (b=.036, p<.001) and negative for defection ones (b=-.018, p<.001). The linear prediction of the model controlling for PI and other confounding factors is also depicted in Figure 2 (dashed lines).

Figure 2. The dots represent the observed individuals' mean *log(RT)* as a function of *dEV* (i.e. expected value of the chosen option minus the expected value of the

unchosen option), separately for cooperation and defection decisions (blue, red respectively). Dotted lines represent the prediction from the regression model (see text). The background bars are the histogram of the expected values. Please note that *dEV* takes negative values for cooperation decision and positive values for defection decisions, since defection was always the payoff maximizing option. Therefore, for both cooperation and defection choices, the more similar the available option, the higher the RTs.

After confirming the anticipated effect of similarity of choices on RTs, we now turn to the role of PI. According to our hypothesis, we expect mean RTs for both cooperation and defection decisions to be strongly dependent on, and follow different trends as a function of, PI. Figure 3 plots the observed mean *log(RT)* as a function of *PI*, together with the linear prediction of Model 1. The RT slopes of cooperation and defection decisions across the levels of PI are clearly in opposite directions. For cooperation decisions, the effect of *PI* on *log(RT)* is negative (b=-.442, p<.001), while for defection decisions it is positive (b=.217, p<.001). Defection decisions are faster under low PI but, as PI increases, defection decisions gradually become slower, reflecting the possible engagement of the deliberative system, as predicted. The opposite holds for cooperation decisions; response latency speeds up as PI increases suggesting an increasing use of the intuitive system. In fact, despite the fact that our experimental context was overall favorable for defection choices, in line with our prediction, when PI is the highest (100%) response latencies for cooperation decisions are significantly faster than those for the defection decisions (even after controlling for similarity of choices).

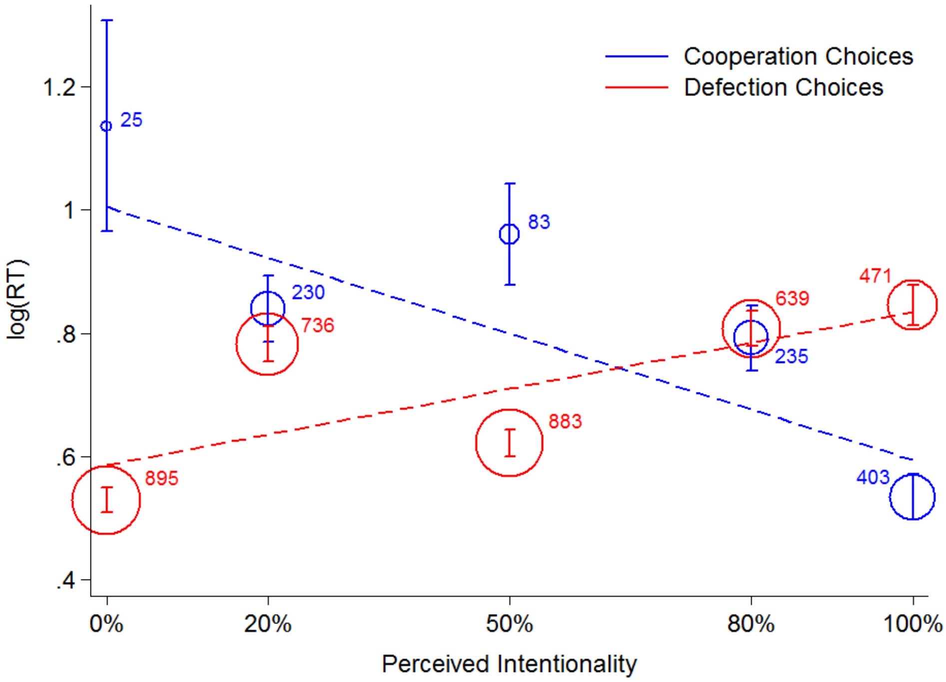

Figure 3. The position of circles represents the observed mean log(RT) for cooperation (blue) and defection (red) decisions as a function of PI, and their size represents number of observations. The predicted values from the Regression Model 1 are represented by the dotted lines. As can be seen, RTs for cooperation and defection decisions follow opposite trajectories with respect to PI, so that the higher the PI, the faster the RTs for cooperation decisions and the slower the RTs for defection decisions.

It is worth noting that despite the general pattern described above, the relationship between PI and RTs (for both cooperation and defection decisions) displays a non-linearity at PI=50%, which represents the highest uncertainty regarding co-player's intentions. Under this uncertain condition defection decisions speed up, even compared with conditions of less likely cooperation (PI=20%). This non-linearity was not anticipated in our initial hypothesis, but might reflect aversion to uncertainty, making defection the default (more intuitive) choice in this particular case. Note however that, as discussed, the effect of PI on RTs does not simply reflect uncertainty, given the clear opposing tendencies in the zero uncertainty cases (PI=100% and PI=0%) as a function of choice made (for PI=0%, defection decisions are the fastest to take, whilst cooperation decisions need the most time to take).

Once the role played by PI as a switch for intuition of collaboration/defection has been settled, we turn the focus on how individual differences might be affecting this switch. It is well-established that individuals differ in their propensity toward cooperation: while some are ready to cooperate in a given situation as long as they assume reciprocity from their partners (i.e. they are conditional cooperators), others tend to defect, or "free-ride" on others' cooperative behaviour (see Gächter, 2007 for a review of the literature). PI, or in this context beliefs about others' cooperative behaviour, might affect the behaviour of these two groups differently. Whereas positive expectations that others will cooperate enhance cooperation propensity in conditional cooperators, they have either null effects or even increase free riders' propensity to defect, as they consider this as an opportunity for maximizing personal gains (Fischbacher & Gächter, 2010). In much the same way that such individual preferences moderate the effect that beliefs have on *behaviour*, in the current dataset individuals' preferences might moderate the *intuitiveness* of each kind of behaviour under varying PI. For conditional cooperators, high PI may make cooperation more intuitive, while for free riders it may induce costs for the otherwise intuitive defection.

We estimated individual cooperation profiles from participants' behavioural patterns in the game. In particular, we measured individuals' choice tendencies to cooperate as a function of increases in the partner's PI, i.e. their degree of conditional cooperativeness. We quantify this tendency by running a separate logistic regression on the proportion of cooperation choices for each participant across all rounds, using PI as the sole independent variable (100 observations per participant). The resulting coefficient, our measure for individual-level responsiveness to PI (*IndCoop*), was subsequently entered as an individual-level variable into our basic model (Model 1) estimating *log(RT)*. As a first step, we introduced *IndCoop* and its interaction with *Coop* in the model. This allows us to test whether individuals with different cooperation profiles differ in their speed of cooperation and defection choices (Model 2). Then, in order to test whether the *effect of PI* on RTs is different among individuals with different cooperation profiles we introduced the interactions *PI*IndCoop* and *PI*Coop*IndCoop* in a separate model. Lastly, to test whether a moderation effect also exists with respect to the effect discriminability has on RTs, we added in the same model the corresponding interaction, *dEV*Coop*IndCoop* (Model 3).

From Model 2 we see that the more conditionally cooperative participants were, the faster they took their cooperation decisions (*b*=-.664, *p*<.001) and slightly (and marginally significant) slower their defection decisions (*b*=.105, *p*=.062) (Figure S1). Importantly, from Model 3, we see that *PI*IndCoop* (*b*=.370, *p*=.001) and the triple interaction *PI*Coop*IndCoop* (*b*=-1.521, *p*<.001) are both highly significant, meaning that the effect of PI on response times is moderated by the individual differences of cooperation attitudes. On the other hand, we observe only a minor interaction effect between *dEV* and *IndCoop* (*dEV*IndCoop*: *b*=.029, *p*=.065; *dEV*Coop*IndCoop*: *b*=-.060, *p*=.205). In order to visualize this pattern, we replotted the interaction shown in Figures 2 and 3 above, separately for conditional cooperators and free riders (see Figure 4). Conditional cooperators (24 out of 46 individuals, or 52% of our sample) were participants whose coefficient from the individual-level regressions of *PI* over *Coop* was positive and highly significant (*p*<.002 for all but three individuals, for whom *p* was .018, .018 and .021), whereas free riders (the remaining 48% of the sample) were participants with either negative or positive but not statistically significant coefficient at conventional levels (*p*>.1 for all but three individuals, for whom *p* was .065, .071 and .073) see Figure S3a,b.

In Figure 4, we plot the mean *log(RT)* separately for conditional cooperators and free riders against PI (Figure 4a) and *dEV* (Figure 4b), as well as the linear prediction of Model 3. Figure 4a gives a clear picture regarding the differential effect that PI has on cooperation decisions for the two typologies of individuals. The effect of PI for conditional cooperators is strong and negative (*b*= -.452, *p*<.001), while for free riders it has a positive sign, but the effect is less strong and does not reach statistical significance, due probably to the small number of observations (*b*=.166, *p*=.305). The strong effect for conditional cooperators and the minor effect for free riders also explains the overall negative relationship found between PI and RTs of cooperation choices, when we did not account for the individual cooperation profiles (*b*=-.442, *p*<.001). In other words, for conditional cooperators, who to begin with take their cooperation decisions on average 2.1 seconds faster (*p*=.04) compared to free riders, a high PI induces cooperation decisions to be taken considerably faster, which indicates use of the intuitive system. For free riders, if anything, it leads to increased response times. For defection decisions on the other hand the effect is positive and significant for both cooperators (*b*=.375, *p*<.001) and free riders (*b*=.177, *p*<.001), although it is again more pronounced among conditional cooperators: a higher PI makes defection decisions less intuitive across all individuals.

Focusing on the effect of discriminability on mean *log(RT)* (Figure 4b) for each group, the picture is clearly different. Both types of individual display the same RT pattern for cooperation and defection decisions, but the effects are now considerably smaller in magnitude. Additionally, regarding cooperation decisions, the effect of *dEV* is significant only for conditional cooperators (*b*=.023, *p*=.012) and it is stronger but only close to significance for free riders (*b*=.040, *p*=.093), again maybe due to small number of observations. Similarly, for defection decisions, the effect is small among free riders (*b*=-.023, *p*<.001) and virtually no effect is observed among cooperators (*b*=-.008, *p*=.269).

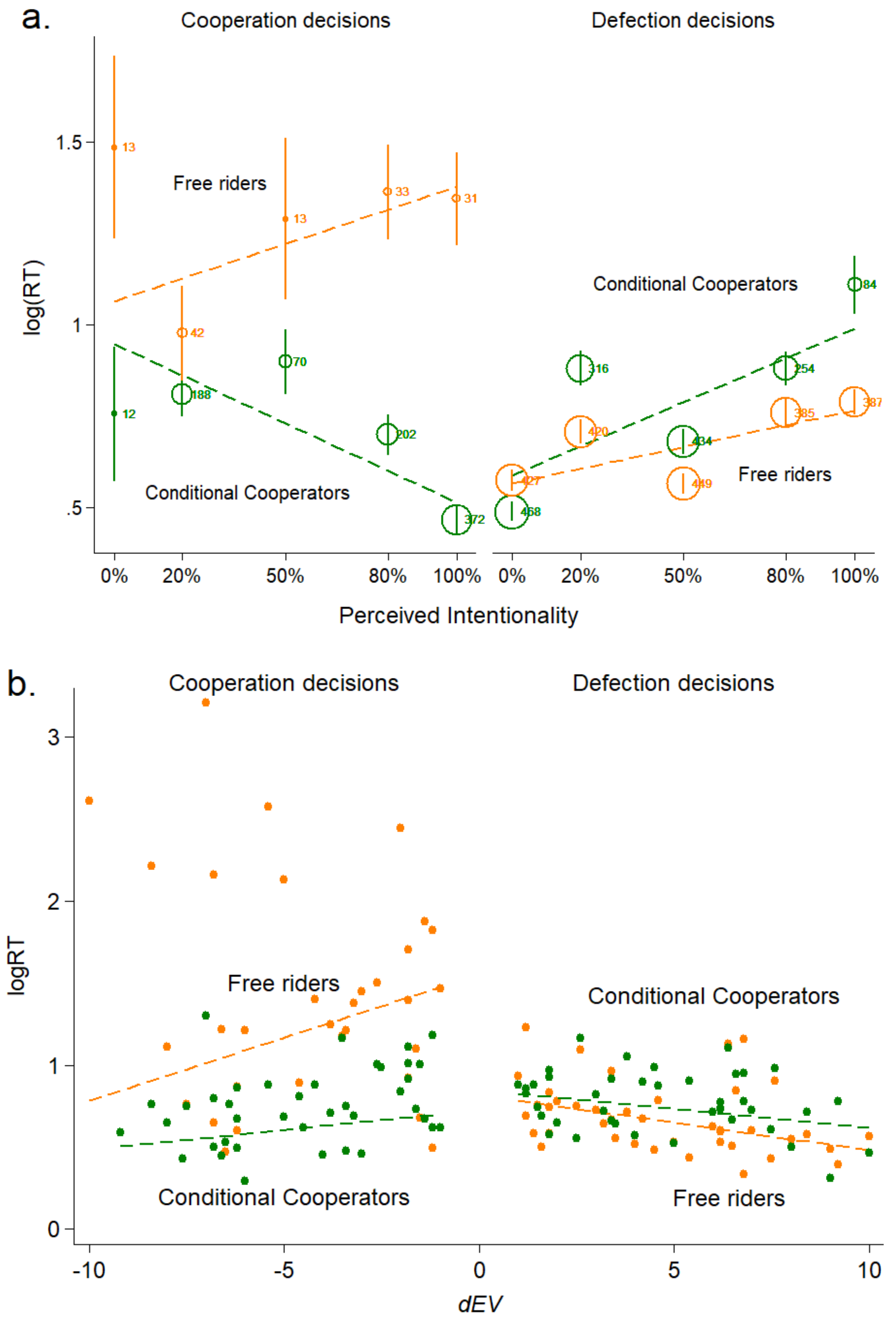

Figure 4. Individual cooperation attitudes (responsiveness to PI) as moderator of the effect of **(a)** *PI* and **(b)** *dEV* on *log(RT)*. The Figure depicts mean observed log(RT), as well as the Model 3 (Table S1) predictions (dashed lines) separately for conditional cooperators (green) and free riders (orange). In Figure 4a, the position of circles represents the observed mean log(RT) for cooperation and defection decisions as a function of PI, and their size represents number of observations. The trajectories are different for conditional cooperators and free riders only in the case of cooperation decisions. For conditional cooperators, the higher the PI the faster they take the decisions, while the opposite pattern is observed among free riders. In Figure 4b, we observe that the effect of *dEV* is drastically diminished once the effect of cooperation profiles is considered.

As a robustness test with respect to the measure of individual cooperation propensities used, we have replicated the analysis using a different proxy for cooperation tendencies: The probability of cooperation across the rounds for each individual indexed as the number of times out of 100 the individual choses to cooperate. Results remained essentially the same (see Tables S2 and S3 and Figures S2 and S3).

Discussion

In summary, the results of the study suggest that cooperation *per se* is not always the intuitive and fast choice in a social dilemma situation (cooperation choices were slower on average). Rather, we find that the intuitiveness of choices is determined by Perceived Intentionality (see Figures 3 and 4a). Indeed, regression analysis confirmed that PI is the principal factor in determining the intuitiveness of choices even after controlling for other confounding factors and crucially for conflict of choices (Models 1-3). Instead, we see that the effect of choice conflict turned out to be substantially reduced when the social environment (PI) and individual differences were jointly considered (Figure 4b, Models 2 and 3). This is important because it implies that RTs did not merely reflect choice complexity. That is, even though RTs is a *correlational* behavioral measure and therefore cannot be an undisputable indicator of causality, it can be a useful tool for inferring cognitive processes provided that social environment within which the choices are made and the individual differences are considered. Additionally, we posit that previously contradictory results may be better understood by employing the present PI framework.

Consider for example the contradictory results of Lotito et al. (2013) and Lohse et al. (2017) in the context of a Public Goods Game, where the former study reported shorter and the latter longer RTs for cooperation decisions. Apart from other design differences, importantly in Lotito et al. (2013) before playing the game, participants engaged in a group activity (either team work, cheap talk communication, mutual introduction or a combination). It is well-documented that these kinds of social interactions directly affect not only cooperative behaviour, but also beliefs about cooperative behaviour of others (see the seminal work by Charness & Dufwenberg, 2006). It is therefore plausible to assume that the beliefs about cooperation, i.e. PI, were higher in Lotito et al. (2013). Then, according to our framework, this difference has affected the intuitiveness of cooperative behaviour. Thus, we might postulate that in both cases, the shorter RTs where associated with the most intuitive behaviour given the social context, namely cooperation in the case of Lotito et al. (2013) and defection in the case of Lohse et al. (2017).

Similarly, in both Cappelen et al. (2016) and Piovesan & Wengström (2009) participants played a Dictator Game both as proposers and as responders. The former study, testing participants from Denmark, reported that the most pro-social choices were faster than the defection ones, whereas the latter, testing participants from Spain, reported the opposite pattern. However, interpersonal trust is well known to vary from country to country, and it happens to be higher amongst Danes than amongst Spaniards (Knack & Keefer, 1997), and thus the former might reasonable hold a higher PI about others than the latter. Therefore, and given that in both cases the design allowed trustworthiness to affect decisions (participants played both roles and that was common knowledge), it is plausible to assume that the fastest choices in both cases were the most intuitive ones: cooperative for the Danes and selfish for the Spaniards.

Lastly, the same framework can be used to interpret studies employing cognitive depletion methodologies. For example, Halali et al. (2014) found that cognitively depleted participants returned higher amounts in response to trusting individuals in a Trust game. In other words, cooperation was the intuitive behaviour whenever the individuals were holding a high PI for the matched trustors, but not otherwise.

The PI framework is also in line with results and conclusions from other recent studies. Nishi et al. (2016) re-analysed four datasets from previously published studies and found that instead of cooperative choices, it is in fact reciprocal choices that are the fastest in repeated Public Goods Games, where participants were informed a priori about the choices of the other members in previous rounds. Therefore, a high (low) PI leads cooperation (defection) decisions to be the fastest choices. They suggested that this was due to the fact that in all four studies the repeated interactions were among members within a stable group formation. Thus, they concluded that it is the social environment of repeated interactions that defines which behaviours are intuitive; in cooperative environments cooperation is intuitive and in non-cooperative environments defection is intuitive. Here, we show that strictly stable group formation is not a necessary condition for the social environment to determine which behaviours are intuitive. Whereas PI alone can in fact determine when cooperation is intuitive, and importantly after controlling for the role of discriminability as well as individual differences regarding cooperation, RTs then can be used as a marker of intuition/deliberation.

Recently, two more studies explicitly pointed to the importance of individual differences vs. conflict of choices in social dilemmas and can also be endorsed in the PI framework. Nielsen et al. (2014) employed a Public Goods Game under the strategy method, i.e. where each participant was required to make a contribution decision for each potential level of contribution of their team members. They found that the fastest choices were those of the conditional cooperators and in contrast the slowest choices were made by the free riders. This finding is indicative that the level of difficulty of the choice task cannot fully account for choice RTs, since the “always free-ride” strategy should be cognitively the easiest. Similarly, Yamagishi et al. (2017) suggested that pro-social individuals are “nervous co-operators”, i.e. individuals who are willing to act pro-socially as long as they expect the same from the others, but they are ready to defect when they fear that they will be exploited by others. They found that cooperation choices were significantly faster, while longer RTs for cooperation choices reflected cooperators’ “social risk aversion”. Interestingly, Yamagishi et al. (2017) found this effect to be minimal among selfish individuals. In our design, we experimentally manipulated PI, i.e. the “risk of being exploited” and we showed that it clearly defines the intuitiveness of cooperation among cooperators, and similarly, that is more predictive among cooperators than among free riders. (Figure 4a). Yamagishi and colleagues concluded that the effect of conflict of choices is, if anything, small, since the increase in RTs of cooperation choices was more prominent among the most pro-social individuals which indicates that the strength of preference does not play a very important role since for those individuals the conflict should be the smallest (i.e., they have strong preferences in favour of cooperation). This result is in line with the diminished overall effect of discriminably among cooperators that we observed in our dataset after controlling for individual cooperation profiles (Figure 4b).

Finally, a note is in order with regard to the Social Heuristics Hypothesis (SHH) (Rand et al., 2014), a previous attempt to reconcile contradictory results. Under the SHH, participants bring the successful strategies of their everyday life in the lab. They apply them intuitively in economic games, even when they are not advantageous, unless their deliberate system takes over, in which case they move closer to more profit-maximization choices. Indeed, SHH has been successful in explaining a large part of the intuitive vs. deliberate behaviour literature (Rand, 2016). The PI framework proposed here is similar insofar as it points to the importance of the social environment. It importantly diverges however, in suggesting that the actors before making a choice, they evaluate the environment in which they act each time. In other words, while SHH suggests context independence of intuitive choices, here we show that intuition depends on the PI, and not only to the general context of the situation.

References

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Barnes, R. D., Ickes, W., & Kidd, R. F. (1979). Effects of the Perceived Intentionality and Stability of Another's Dependency on Helping Behavior. *Personality and Social Psychology Bulletin*, *5*(3), 367–372. https://doi.org/10.1177/014616727900500320

Basu, K. (1994). The traveler's dilemma: Paradoxes of rationality in game theory. *The American Economic Review*, *84*(2), 391–395. https://doi.org/10.2307/2117865

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., … Wollbrant, C. E. (2017). Registered Replication Report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, *489*(3), 427–430. https://doi.org/10.1177/1745691617693624

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., & Wengström, E. (2016). Fairness is intuitive. *Experimental Economics*, *19*(4), 727–740. https://doi.org/10.1007/s10683-015-9463-y

Cappelletti, D., Güth, W., & Ploner, M. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology*, *32*(6), 940–950. https://doi.org/10.1016/j.joep.2011.08.001

Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, *74*(6), 1579–1601. https://doi.org/10.1111/j.1468-0262.2006.00719.x

Chuah, S.-H., Hoffmann, R., & Larner, J. (2016). Perceived Intentionality in 2x2 Experimental Games. *Bulletin of Economic Research*, *68*(S1), 78–84. https://doi.org/10.1111/boer.12073

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are Social Value Orientations Expressed Automatically? Decision Making in the Dictator Game. *Personality and Social Psychology Bulletin*, *37*(8), 1080–1090. https://doi.org/10.1177/0146167211405996

Dennett, D. C. (1989). *The intentional stance*. MIT Press.

Døssing, F., Piovesan, M., & Wengström, E. (2017). Cognitive load and cooperation. *Review of Behavioral Economics*, *4*(1), 69–81.

Evans, A. M., Dillon, K. D., & Rand, D. G. (2014). Reaction Times and Reflection in Social Dilemmas: Extreme Responses are Fast, But Not Intuitive. *Journal of Experimental Psychology: General*, *144*(5), 951–966. https://doi.org/10.2139/ssrn.2436750

Evans, A. M., & van de Calseyde, P. P. F. M. (2017). The effects of observed decision time on expectations of extremity and cooperation. *Journal of Experimental Social Psychology*, *68*, 50–59. https://doi.org/10.1016/j.jesp.2016.05.009

Fermin, A., Sakagami, M., Kiyonari, T., Li, Y., Matsumoto, Y., & Yamagishi, T. (2016). Representation of economic preferences in the structure and function of the amygdala and prefrontal cortex. *Scientific Reports*, *6*, 20982.

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social Value Orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*(2), 272–284. https://doi.org/10.1016/j.obhdp.2012.07.002

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review*, *100*(1), 541–556.

Gächter, S. (2007). Conditional Cooperation: Behavioral Regularities from the Lab and the Field and Their Policy Implications. In B. S. Frey & Stutzer A. (Eds.), *Psychology and Economics: A Promising New Cross-Disciplinary Field* (pp. 19–50). Cambrindge, MA: MIT Press.

Gächter, S. (2012). Human behaviour: A cooperative instinct. *Nature*, *489*(7416), 374–375. https://doi.org/10.1038/489374a

Halali, E., Bereby-Meyer, Y., & Meiran, N. (2014). Between self-interest and reciprocity: The social bright side of self-control failure. *Journal of Experimental Psychology: General*, *143*(2), 745–754. https://doi.org/10.1037/a0033824

Hauge, K. E., Brekke, K. A., Johansson, L.-O., Johansson-Stenman, O., & Svedsäter, H. (2016). Keeping others in our mind or in our heart? Distribution games under cognitive load. *Experimental Economics*, *19*(3), 562–576. https://doi.org/10.1007/s10683-015-9454-z

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(6), 1727–32. https://doi.org/10.1073/pnas.1417904112

Howell, D. (2016). *Fundamental statistics for the behavioral sciences*. Nelson Education.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Knack, S., & Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics*, *112*(4), 1251–1288. https://doi.org/10.1162/003355300555475

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex. *Science*, *314*(5800).

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Science, 6*, 7455.

Lohse, J., Goeschl, T., & Diederich, J. H. (2017). Giving is a Question of Time: Response Times and Contributions to an Environmental Public Good. *Environmental and Resource Economics*, *67*(3), 455–477. https://doi.org/10.1007/s10640-016-0029-z

Lotito, G., Migheli, M., & Ortona, G. (2013). Is cooperation instinctive? Evidence from the response times in a public goods game. *Journal of Bioeconomics*, *15*(2), 123–133. https://doi.org/10.1007/s10818-012-9141-5

Nielsen, U. H., Tyran, J.-R., & Wengström, E. (2014). Second thoughts on free riding. *Economics Letters*, *122*(2), 136–139. https://doi.org/10.1016/j.econlet.2013.11.021

Nishi, A., Christakis, N. A., Rand, D. G., Wyman, E., Herrmann, E., & Nowak, M. (2017). Cooperation, decision time, and culture: Online experiments with American and Indian participants. *PLOS ONE*, *12*(2), e0171252. https://doi.org/10.1371/journal.pone.0171252

Nishi, A., Christakis, N., Evans, A., & O'Malley, A. (2016). Social environment shapes the speed of cooperation. *Scientific Reports, 6*.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York, NY: Cambridge University Press.

Piovesan, M., & Wengström, E. (2009). Fast or fair? A study of response times. *Economics Letters*, *105*(2), 193–196. https://doi.org/10.1016/j.econlet.2009.07.017

Rand, D. G. (2016). Cooperation, Fast and Slow. *Psychological Science*, *27*(9), 1192–1206. https://doi.org/10.1177/0956797616654455

Rand, D. G., Fudenberg, D., & Dreber, A. (2015). It's the thought that counts: The role of intentions in noisy repeated games. *Journal of Economic Behavior & Organization*, *116*, 481–499. https://doi.org/10.1016/j.jebo.2015.05.013

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430. https://doi.org/10.1038/nature11467

Rand, D. G., & Kraft-Todd, G. T. (2014). Reflection does not undermine self-interested prosociality. *Frontiers in Behavioral Neuroscience*, *8*, 300. https://doi.org/10.3389/fnbeh.2014.00300

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Greene, J. D., & Nowak, M. A. (2014). Intuitive Cooperation and the Social Heuristics Hypothesis: Evidence from 15 Time Constraint Studies. *Nature Communications*, *5*. https://doi.org/10.2139/ssrn.2222683

Roch, S. G., Lane, J. A. S., Samuelson, C. D., Allison, S. T., & Dent, J. L. (2000). Cognitive Load and the Equality Heuristic: A Two-Stage Model of Resource Overconsumption in Small Groups. *Organizational Behavior and Human Decision Processes*, *83*(2), 185–212. https://doi.org/10.1006/obhd.2000.2915

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

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, *41*, 77–87. https://doi.org/10.1016/j.joep.2012.08.007

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*(2), 161–188. https://doi.org/10.1037/0096-3445.127.2.161

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2012). Impulse Control and Underlying Functions of the Left DLPFC Mediate Age-Related and Age-Independent Individual Differences in Strategic Social Behavior. *Neuron*, *73*(5), 1040–1051. https://doi.org/10.1016/j.neuron.2011.12.027

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The Sunny Side of Fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*(4), 339–347. https://doi.org/10.1111/j.1467-9280.2008.02091.x

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., … Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, *498*(7452), E1–E2. https://doi.org/10.1038/nature12194

Tomasello, M. (2009). *Why we cooperate*. MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5).

Verkoeijen, P. P. J. L., Bouwmeester, S., & Hegarty, B. (2014). Does Intuition Cause Cooperation? *PLoS ONE*, *9*(5), e96654. https://doi.org/10.1371/journal.pone.0096654

Warneken, F., & Tomasello, M. (2006). Altruistic Helping in Human Infants and Young Chimpanzees. *Science*, *311*(5765).

Yamagishi, T., Matsumoto, Y., Kiyonari, T., Takagishi, H., Li, Y., Kanai, R., & Sakagami, M. (2017). Response time in economic games reflects different types of decision conflict for prosocial and proself individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(24), 6394–6399. https://doi.org/10.1073/pnas.1608877114

Yamagishi, T., Takagishi, H., Fermin, A. de S. R., Kanai, R., Li, Y., & Matsumoto, Y. (2016). Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(20), 5582–7. https://doi.org/10.1073/pnas.1523940113

Zaki, J., & Mitchell, J. P. (2013). Intuitive Prosociality. *Current Directions in Psychological Science*, *22*(6), 466–470. https://doi.org/10.1177/0963721413492764