Personality and Individual Differences 51 (2011) 210–215

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Understanding cooperation in the Prisoner's Dilemma game

Emmanuel M. Pothos [a,*], Gavin Perry [a], Philip J. Corr [b], Mervin R. Matthew [c], Jerome R. Busemeyer [c]

[a] Department of Psychology, Swansea University, UK
[b] School of Social Work and Psychology, and the Centre for Behavioural and Experimental Social Science (CBESS), University of East Anglia, UK
[c] Department of Psychological and Brain Sciences, Indiana University, USA

ARTICLE INFO

Article history:
Received 24 March 2010
Received in revised form 27 April 2010
Accepted 3 May 2010
Available online 31 May 2010

Keywords:
Prisoner's Dilemma game
Personality
Cooperation
Disjunction effect
Reward Responsiveness
Agreeableness

ABSTRACT

In the standard one-shot Prisoner's Dilemma game, participants often choose to cooperate, when the optimal strategy is to defect. This puzzling finding has attracted considerable interest both in terms of possible computational frameworks for modeling human judgment, but also regarding the more general debate of human altruism. In this research, we ask how much of human behavior in this task can be explained by a putative bias for cooperative behavior and whether this, in turn, is influenced by personality. We compared performance on the standard task with performance on a matched neutral one; we manipulated the optimal strategy (defect or cooperate); and we manipulated the amount of payoff for responding optimally. Results revealed little evidence for a bias for cooperative behavior, but significant associations with the personality factors of Behavioural Activation System (BAS) Reward Responsivity and Agreeableness were found. These findings are discussed in terms of the attempt to explain judgment in one-shot, Prisoner's Dilemma tasks with statistical or probabilistic models.

1. Introduction

The role of personality in explaining heterogeneity of decision making in the Prisoner's Dilemma game is under-explored. The potential for personality to play a role is suggested by the nature of the game, in which, typically, two players receive different levels of reward/punishment based on their decision either to cooperate (C) or defect (D). Indeed, there have been several reports of how particular personality traits, such as Reward Responsiveness, can predict cooperative behavior in games analogous to Prisoner's Dilemma (e.g., Scheres & Sanfey, 2006; Skatova & Ferguson, 2011). A major aim of this article is to investigate the role of personality in this important economic game.

In the standard game, the actions of both players can be denoted thus: DD, DC, CD, and CC, where the first index corresponds to the first player's action. A typical payoff pattern for the first player is DD = £10, DC = £25, CD = £5, and CC = £20. In a 'one-shot' game, if both players choose C then they both get a reasonably high payoff of £20; but if the first player, in a spirit of mutual benefit, decides to C and the second player to D, then the first player will only get £5, the lowest possible payoff. Let us put ourselves in the shoes of the first player. If he thinks the second player will C, then the greatest payoff is obtained by choosing to D and, likewise, if he thinks the second player will D. In fact, the decision to D is a Nash equilibrium in the standard one-shot Prisoner's Dilemma game: it is not possible to improve upon this strategy by a unilateral decision. Thus, D is always the optimal strategy.

But human decision making does not always conform to this optimal decision strategy! Shafir and Tversky (1992; Busemeyer, Matthew, & Wang, 2006; Croson, 1999; Li & Taplin, 2002; Tversky & Shafir, 1992) created a well-known modification to the Prisoner's Dilemma game: in some trials, participants were told what the other player was doing. Unsurprisingly, when participants were told that the other person decided to D, then their probability to D was 97%; and when they were told that the other person decided to C, then their probability of D was 84%. However, in trials (within participants design) when participants were not told what the other person did, the probability to D dropped to 63%. This pattern of responding is a puzzle for optimal decision theorists. In fact, it is an extremely significant finding in relation to theories of human judgment based on classical probability. The 'sure thing principle' states that $P(A) = P(A \wedge X) + P(A \wedge notX)$, where $P(A)$ is the probability of event A, etc. This is a fundamental principle of classical probability theory. But, participants' behavior in Shafir and Tversky's (1992) experiment is inconsistent with the sure thing principle and so with classical probability theory (this is sometimes called the disjunction effect). Why is this significant? Because classical probability theory makes sense (in the same way that $2+2=4$ makes sense) and it has been extensively employed in modeling human judgment.

The disjunction effect indicates that, if there is some uncertainty – however logically irrelevant – naïve observers are more

* Corresponding author. Address: Department of Psychology, Swansea SA2 8PP, UK. Tel.: +44 2920569104.
E-mail address: e.m.pothos@swansea.ac.uk (E.M. Pothos).

0191-8869/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2010.05.002

willing to cooperate. This raises an interesting question: do people have a bias for cooperative behavior? Fehr and Fischbacher (2003, p. 785) introduced a *Nature* review on altruism with the statement: 'Human societies represent a huge anomaly in the animal world. They are based on a detailed division of labor and cooperation between genetically unrelated individuals in large groups.' Regarding Prisoner's Dilemma, similar ideas have been explored in the economics literature. For example, Cooper, DeJong, Forsythe, and Ross (1996, p. 14), states, 'The strongest evidence that altruism is at least part of the explanation for observed cooperative play comes from these one-shot PD games. If players are egoists, cooperation will not be observed in one-shot PD games.'

There have been demonstrations of cooperative behavior which is adaptive in the sense that the cooperating agent can gain in the long term. For example, in a repeated Prisoner's Dilemma game, cooperating can lead to the greatest payoff, if the players reciprocate (Kreps, Milgrom, Roberts, & Wilson, 1982). Likewise, repeated encounters which allow the formation of reputation for cooperating can be adaptive (reputation formation is like tit-for-tat; Fehr & Fischbacher, 2003). Situations which punish selfish behavior or reward cooperative behavior also reveal a preference for cooperative behavior (Fehr & Fischbacher, 2004).

Our purpose is to examine whether the disjunction effect can be understood in terms of a bias to cooperate. We compare a standard Prisoner's Dilemma game with a task with identical payoff matrices but presented as a card game. As in the latter task participant responses do not involve C/D actions, a disjunction effect could not occur from a bias for cooperative behavior (such a bias would favor C responses). Also, in half the trials the optimal strategy was to D but in the other half the optimal strategy was to C. In the latter case, a cooperative bias should increase the frequency of optimal responses. Moreover, in half the trials, the optimal behavior led to a higher payoff than the other half. If the disjunction effect arises from a bias to C, then we would expect this bias to be more pronounced in the low payoff trials. Finally, we examine the propensity to C with two individual differences questionnaires which might reveal participant characteristics relevant to cooperative behavior or the desire to maximize profits. Some researchers have reported links between strategic 'giving' and Reward Responsiveness in Public Good Games and variants (Scheres & Sanfey, 2006; Skatova & Ferguson, 2011) and others that cooperative behavior is enhanced by reciprocation (Bardsley & Sausgruber, 2005). An interesting aspect of the current task is that cooperative behavior carries no financial reward (e.g., through reciprocation).

2. Method

2.1. Participants

All 113 participants were undergraduate students at Swansea University and took part for course credit. The experiment had one between participants factor (whether the game was the standard Prisoner's Dilemma game or a card game) and two within participants factors (the information given regarding the other player's action and whether the payoff matrix was considered 'high' or 'low variability').

2.2. Materials

2.2.1. Prisoner's Dilemma games

The stimuli consisted of a series of payoff matrices, individually presented to participants, with instructions to respond so as to maximize payoff. There were two versions of the stimuli, one in which they corresponded to payoff matrices in a Prisoner's Dilemma game and another as payoff matrices in a single-player card game. The possible participant actions in the Prisoner's Dilemma game were D or C (as is standard) and in the card game to play a red card or a black card.

The Prisoner's Dilemma task and the card task differed only in instructions. The underlying payoff matrices were identical. The payoff matrices were organized into two groups of 50 distinct matrices each, which we label high variability and low variability. The prototype matrix for the high variability matrices was: DD = £45, CD = £15, DC = £80, CC = £50. These matrices were designed to encourage participants to respond optimally, as there is a large payoff for doing so. The 50 high variability matrices were created from the prototype by adding to each payoff a random number between 0 and 30.

The 50 low variability matrices were created so that there would be less (financial) incentive to respond optimally. Their prototype was DD = £30, CD = £25, DC = £70, CC = £65. If there is a bias for cooperative behavior, one would imagine that it would affect participants' performance more so with the low variability matrices. The low variability matrices were created by adding a random number between 0 and 5 to each of the values in the prototype.

The 50 matrices in each of the low and high variability groups were further divided in two ways. First, the 50 matrices in each group were divided into groups of 25 matrices each. In the first group, the optimal strategy was to D (or to play a black card). In the second group, the payoffs were reversed so that the optimal strategy became to C (or to play a red card). Next, the 25 matrices in the high variability group when the optimal strategy was to D, were divided into a group of 15 matrices, in which the participant was not told the action of the other player (or the card drawn from the deck), five matrices in which the other player decided to D (or a black card was drawn) and five matrices in which the other player decided to C (or a red card was drawn). The same approach was adopted in all groups of payoff matrices obtained by crossing high/low variability with optimal strategy.

2.2.2. Personality measures

Two measures of individual differences were employed. The BIS/BAS scales (Carver & White, 1994) measure various aspects of a person's motivational systems with respect to appetitive behavior and avoidance of undesirable situations. There are three Behavioural Activation System (BAS) scales, (drive, fun seeking, and Reward Responsiveness) and one BIS scale. Regarding Prisoner's Dilemma, one possibility is that Reward Responsiveness would predict the extent to which a participant responds in a way to maximize his payoff.

The second measure of individual differences was a shorter (50-item) version of the Trait Self-Description Inventory (SSDI; Collis & Elshaw, 1998; modified by Adam Perkins and Philip Corr) which measures the Big-5 factors. The Big-5 questionnaire allows a characterization of a person along five dimensions: Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to Experience. Regarding Prisoner's Dilemma, participants high on Agreeableness would be expected to be more likely to C even when the optimal strategy is defection.

2.2.3. Procedure

In the standard version of the task, the Prisoner's Dilemma game was explained with the help of a payoff matrix. Participants were told that they would see several payoff matrices, each of which corresponding to playing a Prisoner's Dilemma game with a different player. Participants were asked to decide whether to C or D so as to win more money and they were told that in some cases they would know what the other player did before making their own choices. Finally, participants were asked to *imagine* that they would be playing against real players and that they would be winning real money (cf. Rilling et al., 2002).

We refer to the neutral version of the task as the card version. The instructions for the card version were matched as closely as possible with the instructions for the standard Prisoner's Dilemma game. Participants were told that they would play a series of card games for one player and that in each game they would have to decide whether to play a red or a black card. Likewise, a red or a black card would be drawn from a pack of red and black cards. The instructions then showed a corresponding payoff matrix. One obtains the standard Prisoner's Dilemma game by substituting red with C and black with D.

Table 1
Evidence for a disjunction effect across all conditions of the experiment.

	Low variability	High variability
Prisoner's Dilemma game; optimal strategy is D		
Pr(D|Known C)	.872	.904
Pr(D|Known D)	.936	.944
Pr(D|Unknown)	.853	.901
Card game; optimal strategy is D (play black)		
Pr(D|Known C)	.890	.857
Pr(D|Known D)	.935	.930
Pr(D|Unknown)	.856	.884
Prisoner's Dilemma game; optimal strategy is C		
Pr(D|Known C)	.088	.084
Pr(D|Known D)	.129	.120
Pr(D|Unknown)	.103	.087
Card game; optimal strategy is C (play red)		
Pr(D|Known C)	.050	.090
Pr(D|Known D)	.080	.052
Pr(D|Unknown)	.080	.033

A trial started with a blank screen for 1 s, followed by the payoff matrix (together with possible information about the other player's action). For the first 5 s after the payoff matrix had appeared, the participant could not respond. After the 5 s, a caption appeared saying 'Defect or Cooperate?' (or 'Black or Red?'); this was participants' cue that they could respond from that point onwards. Participants were allowed in total 30 s to respond. If no response had been made during that time, the next trial was shown (this happened in 0.16% of all trials). Participants indicated their responses by pressing C or D to indicate cooperation or defection or R and B to indicate red or black card.

All 50 high variability matrices were presented in one block and all 50 low variability matrices in another block. For each participant block order and trial order within blocks was randomized.

3. Results

For simplicity, we describe results in terms of C or D regardless of task version and focus on the probability to D, Pr(D), regardless of strategy. Outliers were examined for Pr(D), regardless of task version. For three participants, Pr(D) was more than 2.5 SD below the mean Pr(D) so their results were eliminated. This left us with 50 participants who did the Prisoner's Dilemma game and 60 who did the card game.

We first examined the evidence for a disjunction effect. This was done with paired-samples *t*-tests comparing Pr(D|Unknown), Pr(D|Known C), and Pr(D|Known D) (Table 1). Regarding the Prisoner's Dilemma game, when the optimal strategy was D, for the high variability matrices, significant differences were observed only in the case of the pair Pr(D|Known D) – Pr(D|Unknown)

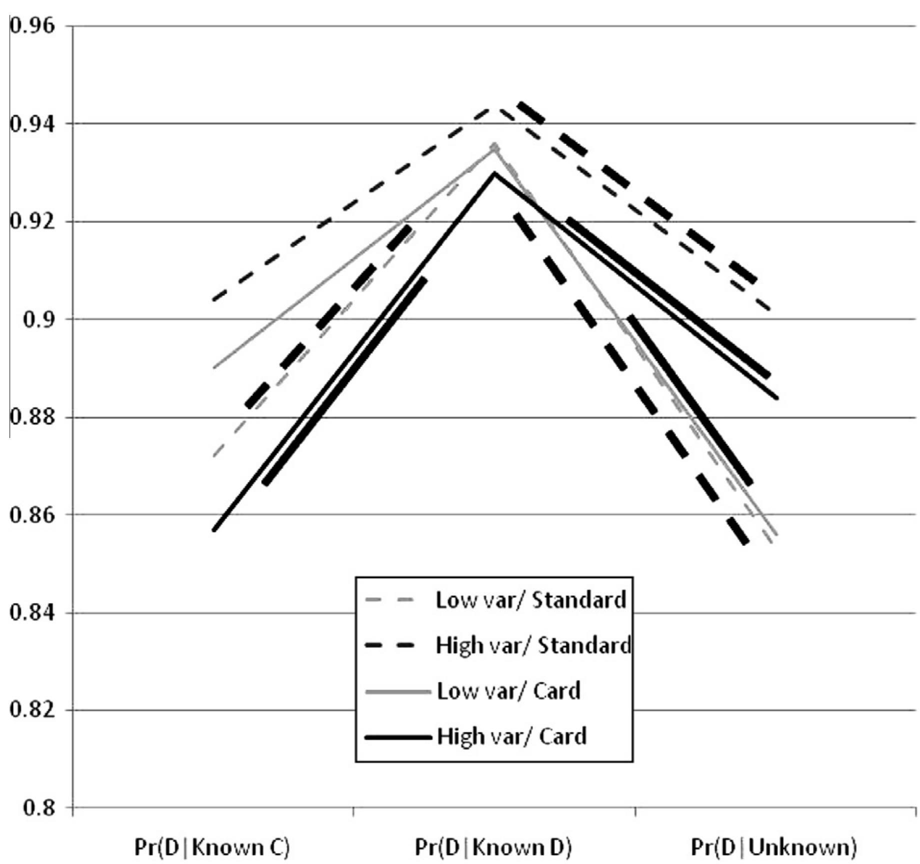

Fig. 1. The basic results for when the optimal strategy was D. Where duplicate lines appear this indicates that the corresponding difference was either significant or very nearly significant.

($t(49) = 2.79$, $p = .008$). For the low variability matrices, significantly different pairs were Pr(D|Known C) – Pr(D|Known D) ($t(49) = -1.91$, $p = .062$) and Pr(D|Known D) – Pr(D|Unknown) ($t(49) = 2.34$, $p = .023$). When the optimal strategy was C, the only marginally significant difference was identified for low variability matrices in the case of Pr(D|Known C) – Pr(D|Known D) ($t(49) = -1.83$, $p = .074$).

The results for the card version of the task were similar. When the optimal strategy was D, for the high variability matrices, there were significant differences involving the pair Pr(D|Known C) – Pr(D|Known D) ($t(59) = -2.29$, $p = .026$) and the pair Pr(D|Known D) – Pr(D|Unknown) ($t(59) = 1.96$, $p = .054$). For the low variability matrices, the only significant difference was in the case of Pr(D|Known D) – Pr(D| Unknown) ($t(59) = 3.32$, $p = .002$). When the optimal strategy was C, the only marginally significant difference was observed for the high variability matrices in the case of the pair Pr(D|Known C) – Pr(D|Known D) ($t(59) = -1.91$, $p = .061$).

These results are easy to explain. When the optimal strategy was D, in all cases, regardless of whether the task was Prisoner's Dilemma or the card game, the Pr(D) was highest for the trials in which it was known that the other player's response was D, and lower when the other player's response was C or unknown (equivalently for the card game). Fig. 1 illustrates the equivalence of the results for all conditions. At face value, Fig. 1 undermines the perception of any differences between either the standard and the card version of the task, or between the low and high variability matrices. For when the optimal strategy was C, the pattern of results was a little more mixed. However, importantly, the same 'hats' were observed for the results regarding Prisoner's Dilemma (see Croson (1999) for similar results).

If participants have a cooperative bias, then there should be more optimal responses when the optimal response is C, in the case of the standard Prisoner's Dilemma game. A mixed design ANOVA with version of the task as a between participants' factor and optimal strategy as a within participants' factor was run. The dependent variable was Pr(optimal response) across all trials. The main effect of optimal strategy was not significant ($F(1108) = 2.80$, $p = .097$), but the interaction was significant ($F(1108) = 3.93$, $p = .05$). However, the interaction was in the opposite direction from that predicted by a cooperativeness hypothesis. Why should there be a larger Pr(optimal response) when the optimal response was to play one card but not the other? This could reflect a sampling bias (e.g., participants might consider a 'better' choice the red card than the black card).

Next, we looked at the disjunction effect directly, restricting the analyses to when the optimal strategy was D (or play a black card). Note first that on average we did not observe a disjunction effect as such, but rather a situation such that Pr(D|Known D) > Pr(D|Known C), Pr(D|Unknown). We defined the disjunction effect as Disj $= \frac{\Pr(D|\text{Known C}) + \Pr(D|\text{Known D})}{2} - \Pr(D|\text{Unknown})$, noting that according to the law of total probability Pr(D|Unknown) has to be convex combination of Pr(D|Known D) and Pr(D|Known C) (i.e., it cannot be lower than either). Using this disjunction effect as a dependent variable, we ran a mixed design ANOVA with version of the task as a between participants variable and high vs. low variability matrices as a within participants variable (Fig. 2). There was a trend for a main effect of high vs. low variability matrices ($F(1108 = 3.16$, $p = .078$). The interaction was not significant ($F(1108) = 0.21$, $p = .647$).

A few observations are in order. First, the overall level of disjunction effect was much lower than that observed by Shafir and

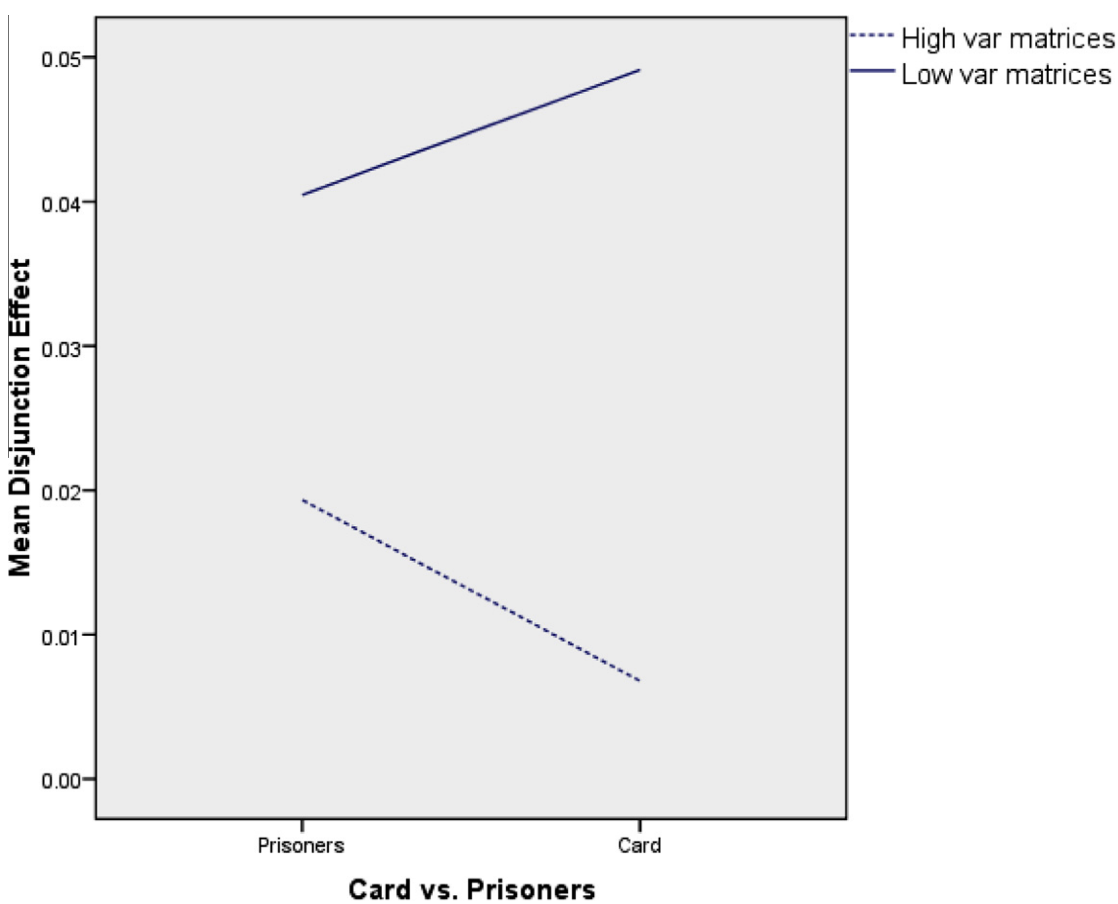

Fig. 2. The disjunction effect for when the optimal strategy is D, separately for the two versions of the task and for high vs. low variability matrices.

Table 2
Personality and Pr(D) correlations. Pr(D) values are averaged across all possibilities for what information the participant received regarding the other player's action and both high and low variability matrices. The column entries show correlation values between Pr(D) and the individual difference indices. *p*-Values are shown for significant or nearly significant correlations.

	Correlation	*p*-Value
Optimal strategy: D		
BAS reward	.33	.018
Neuroticism	−.18	
BIS	.17	
Conscientiousness	−.09	
BAS fun	−.06	
BAS drive	.04	
Openness	−.03	
Agreeableness	−.02	
Extraversion	.01	
Optimal strategy: C		
Agreeableness	−.27	.054
BIS	−.21	
Extraversion	−.21	
BAS drive	.19	
BAS fun	.12	
Openness	.11	
BAS reward	−.05	
Neuroticism	−.04	
Conscientiousness	−.03	

Tversky (1992). However, our results are consistent with other recent similar research (Matthew & Busemeyer, 2010; see also Croson (1999) for a demonstration of a weaker disjunction effect with an 'imagining' procedure similar to the one we adopted). Second, as expected, there was a lower disjunction effect for the high variability matrices. If we consider the disjunction effect to be a measure of sub-optimal performance, then this means that the larger payoff in the higher variability matrices encouraged participants to behave more optimally. However, regarding the putative cooperativeness hypothesis, there was an equivalent disjunction effect in both versions of the task. This finding undermines the hypothesis that the disjunction effect in the standard Prisoner's Dilemma game arises from a bias to C in the unknown condition, since such a bias is irrelevant in the neutral version of the task.

The final issue concerns the propensity to C depending on participant individual characteristics. We examined how the four BIS/BAS indices and the five indices of the OCEAN Big Five questionnaire correlated with the average Pr(D) in the standard version of the Prisoner's Dilemma, regardless of what information participants were given about the other player's action or whether the matrices were high/low variability (Table 2). One conclusion is that a different pattern of performance is obtained depending on the optimal strategy. Reward Responsiveness (which measures sensitivity to reward) dominates in the former case. By contrast, when the optimal strategy was C, the highest correlation was observed between Pr(D) and Agreeableness, which is a measure of how much a person wants to be liked by other people.

It is possible that particular traits have a greater influence on behavior when there is more or less information about the other player's action (cf. Skatova & Ferguson, 2011). When the optimal strategy was D, Reward Responsiveness correlated with Pr(D) only when the other player's action was known ($r = .39$, $p = .005$). However, when the optimal strategy was C, Agreeableness was not moderated in an analogous way. This analysis did not reveal any other significant results.

4. Discussion

The explosion of cognitive science research in the last few decades has created a real optimism regarding how much of human behavior can be explained in terms of environment statistics. An example of a formal explanation for Prisoner's Dilemma is Chater, Vlaev, and Grinberg's (2008) suggestion that the structure of the game makes it more likely for participants to both C or D. For example, if the payoff for CC is not much lower compared to the payoff for DC, then this would encourage both participants to C and avoid the risk of a low-payoff DD situation. This idea of agreement between participants' actions is also taken up in Pothos and Busemeyer's (2009) quantum probability model. The classical psychological intuition is that in assessing $P(event)$, where X, not_X are relevant, there are distinct 'thoughts' corresponding to $P(event) \wedge X$ and $P(event \wedge not\_X)$. These 'thoughts' are independent, so that each 'thought' has a simple additive influence in the computation of $P(event)$, as given by the law of total probability. In the quantum approach, in computing $P(event)$, the 'thoughts' for $P(event \wedge X)$ and $P(event \wedge not\_X)$ can interfere with each other and so the law of total probability can be violated.

Could a formal model provide a complete account of Prisoner's Dilemma results? It may be possible (indeed, it seems likely) that cooperative behavior is adaptive at the society level (e.g., Fehr & Fischbacher, 2004) and this may lead to an overall bias for cooperative behavior. However, in the context of a specific task, such as a one-shot Prisoner's Dilemma, a propensity to cooperate would be just that, a bias without any adaptive value (in the sense that it would not translate to payoff). Accordingly, in such tasks, there is a genuine tension between formal explanations (e.g., based on a probabilistic or statistical formulation of the problem) and explanations which rely on a response bias.

We manipulated the thematic content of the task, the amount of payoff for responding optimally, and whether the optimal strategy was C or D. The main findings can be summarized as follows. First, there was a trend for a lower disjunction effect with the high variability matrices, an intuitive finding in the sense that higher payoff encouraged participants to select the appropriate response. Second, it was not the case that there were more optimal responding when the optimal strategy was to C as opposed to when it was to D, contrary to what a hypothesis that participants are biased to C would predict. Third, if the source for the disjunction effect is a bias to C, then we would expect the disjunction effect to disappear in the neutral version of the task, but this was not the case (see Fig. 2). Finally, the higher likelihood to C correlated with theoretically-relevant personality factors from the BIS/BAS questionnaire and the OCEAN Big Five one. One conclusion from these analyses is that Pr(D) depends on different individual characteristics in games in which the optimal strategy is to C as opposed to D.

So, overall, there is no support for the hypothesis that the disjunction effect in standard one-shot Prisoner's Dilemma games arises from a bias to C. This is an uplifting conclusion regarding formal approaches to understanding human behavior in Prisoner's Dilemma. Some additional useful conclusions are possible. First, the results from the standard Prisoner's Dilemma game are similar to those with the neutral version (Fig. 1). Therefore, one can ask whether there is something special about Prisoner's Dilemma situations as such, or whether the important aspect of the task is the way in which the payoff matrix is constructed. Second, we did not observe a disjunction effect in the way Shafir and Tversky (1992) did. We found that, when the optimal strategy was D, in the standard Prisoner's Dilemma game, Pr(D|Known D) > Pr(D|Known C), Pr(D|Unknown). However, if participants were trying to maximize their gain, they should be responding with a high Pr(D) when they were told that the other player chose to C as well. Why did we not get this effect? One possibility is that the payoffs were not large enough to entice participants to D. But such an explanation is inconsistent with our other findings which did not support a cooperativeness bias. Another possibility is that there might have been a matching bias in our results (Evans, Newstead, & Byrne, 1991), so that participants were more likely to adopt the response of the other player. Such a possibility is supported by the fact that we observed a similar pattern of results with the card game.

The study of individual differences revealed that, when the optimal strategy was to D, participants high on BAS Reward Responsiveness were more likely to do so, when they knew the action of the other player. This result is consistent with related work with Public Good Games, whereby it has been shown that Reward Responsiveness moderates altruistic contributions (Scheres & Sanfey, 2006; Skatova & Ferguson, 2011). Note, though, that a similar association was not observed when the other player's action was not known or when the optimal strategy was to C, suggesting that task context determines whether a particular trait influences behavior (cf. Bibby & Ferguson, 2011). Also, higher BIS scores are thought to relate to greater suppression of sub-optimal responses (e.g., Gray & McNaughton, 2000), but no such association was observed. Regarding the Big Five, when the optimal strategy was to C, Agreeableness was associated with cooperative behavior, but the lack of a similar association when the optimal strategy was to D is puzzling. Further research is needed to explore these interesting observations.

Acknowledgements

Thanks to Carys John, Hannah Jones, James Gordon, and Dewi Griffiths for help with collecting the data and Ian Deary for his suggestions in relation to individual differences. Gavin Perry was funded by WICN.

References

Bardsley, N., & Sausgruber, R. (2005). Conformity and reciprocity in public good provision. *Journal of Economic Psychology, 26*, 664–681.

Bibby, P. A., & Ferguson, E. (2011). The ability to process emotional information predicts loss aversion. *Personality and Individual Differences, 51*, 263–266.

Busemeyer, J. R., Matthew, M., & Wang, Z. A. (2006). Quantum game theory explanation of disjunction effects. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 131–135). Mahwah, NJ: Erlbaum.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319–333.

Chater, N., Vlaev, I., & Grinberg, M. (2008). A new consequence of Simpson's paradox: Stable cooperation in one-shot prisoner's dilemma from populations of individualistic learners. *Journal of Experimental Psychology: General, 137*, 403–421.

Collis, J. M., & Elshaw, C. C. (1998). *The development of the trait self description inventory: Results of US/UK collaboration*. TTCP/HUM/98-001.

Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior, 12*, 187–218.

Croson, R. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes, 80*, 118–133.

Evans, St. B. T. J., Newstead, S. E., & Byrne, R. J. M. (1991). *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum Associates.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*, 785–791.

Fehr, E., & Fischbacher, U. (2004). Third party punishment and social norms. *Evolution and Human Behavior, 25*, 63–87.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety*. Oxford: Oxford University Press.

Kreps, D., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory, 17*, 245–252.

Li, S., & Taplin, J. (2002). Examining whether there is a disjunction effect in prisoner's dilemma games. *Chinese Journal of Psychology, 44*, 25–46.

Matthew, M. R., & Busemeyer, J. R. (2010). *The disjunction effect on a sliding scale*. Unpublished.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal Society B, 276*, 2171–2178.

Rilling, J. K., Gutman, D., Zeh, G., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron, 35*, 395–405.

Scheres, A., & Sanfey, A. G. (2006). Individual differences and decision making: Drive and reward responsiveness affect strategic bargaining in economic games. *Behavioral and Brain Functions, 2*, 35–43.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology, 24*, 449–474.

Skatova, A., & Ferguson, E. (2011). What makes people cooperate? Individual differences in BAS/BIS predict strategic reciprocation in a public goods game. *Personality and Individual Differences, 51*, 237–241.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*, 305–309.