



## Research

**Cite this article:** Schroeder KB, Nettle D, McElreath R. 2015 Interactions between personality and institutions in cooperative behaviour in humans. *Phil. Trans. R. Soc. B* **370**: 20150011.  
<http://dx.doi.org/10.1098/rstb.2015.0011>

Accepted: 2 July 2015

One contribution of 13 to a theme issue 'Solving the puzzle of collective action through inter-individual differences: evidence from primates and humans'.

**Subject Areas:**  
behaviour

**Keywords:**  
punishment, public goods game, extraversion, neuroticism, agreeableness

**Author for correspondence:**

K. B. Schroeder  
e-mail: [kari.britt.schroeder@gmail.com](mailto:kari.britt.schroeder@gmail.com)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2015.0011> or via <http://rstb.royalsocietypublishing.org>.

## Interactions between personality and institutions in cooperative behaviour in humans

K. B. Schroeder<sup>1,2</sup>, D. Nettle<sup>1</sup> and R. McElreath<sup>2</sup>

<sup>1</sup>Centre for Behaviour and Evolution, Newcastle University, Framlington Place, Newcastle upon Tyne NE24HH, UK

<sup>2</sup>Department of Anthropology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

Laboratory attempts to identify relationships between personality and cooperative behaviour in humans have generated inconsistent results. This may partially stem from different practices in psychology and economics laboratories, with both hypothetical players and incentives typical only in the former. Another possible cause is insufficient consideration of the contexts within which social dilemmas occur. Real social dilemmas are often governed by institutions that change the payoff structure via rewards and punishments. However, such 'strong situations' will not necessarily suppress the effects of personality. On the contrary, they may affect some personalities differentially. Extraversion and neuroticism, reflecting variation in reward and punishment sensitivity, should predict modification of cooperative behaviour following changes to the payoff structure. We investigate interactions between personality and a punishment situation via two versions of a public goods game. We find that, even in a strong situation, personality matters and, moreover, it is related to strategic shifts in cooperation. Extraversion is associated with a shift from free-riding to cooperation in the presence of punishment, agreeableness is associated with initially higher contributions regardless of game, and, contrary to our predictions, neuroticism is associated with lower contributions regardless of game. Results should lead to new hypotheses that relate variation in biological functioning to individual differences in cooperative behaviour and that consider three-way interactions among personality, institutional context and sociocultural background.

## 1. Introduction

Social dilemmas are situations in which the best outcome for an individual comes from acting in her immediate self-interest, but the best outcome for the group results when everyone acts in the interest of the group (cooperates) [1]. Thus, social dilemmas present a conflict for the individual between choosing the action that benefits herself and choosing the action that benefits the group. Social dilemmas have garnered immense cross-disciplinary interest for two reasons. First, the remarkably successful spread of humans in recent evolutionary history can be at least partially attributed to our ability to 'solve' social dilemmas, and second, many of the great challenges of our time, such as nuclear proliferation and global warming, are social dilemmas.

Both paths of inquiry lead us to question why some groups of people achieve more cooperative outcomes than others. One of the answers to this question lies with individuals. In the laboratory, social dilemmas have been studied with the prisoner's dilemma (PD), public goods game (PGG) and commons dilemma. Laboratory studies demonstrate considerable individual variation in cooperative behaviour in social dilemmas [2], and the level of cooperation achieved by a group depends upon the individual composition of the group [3–7].

Briefly, we describe these three games as typically presented in the studies reviewed in this paper; readers who desire a more comprehensive background are referred elsewhere [8,9]. In the PD, each of two players privately chooses to either cooperate or defect. The highest payoff is achieved by either player when

she defects and her partner cooperates, but the payoff to both players is higher when both cooperate than when both defect. The PGG is an extension of the PD to groups larger than two. Individuals receive an allotment and can contribute some or all of it to the public good; the units in the public good are then multiplied (by a number greater than one and less than the number of players) and divided equally among all players. A player's payoff from the public good is in addition to whatever portion of the allotment a player may have retained. In the commons dilemma, players decide how much of a finite commonly held resource to use. While a player can achieve the highest immediate payoff via unrestrained use of the resource, overexploitation leads to the collapse of the resource and decreased payoffs for all players.

Individual heterogeneity in cooperative behaviour in social dilemmas may be stable over a period of at least months [10,11], if not years [12]. Behaviour is also relatively consistent across different economic games, including not just 'strict' social dilemmas, but other games, such as the trust game [11,13] (but see [14]). Furthermore, cooperative behaviour in social dilemmas in the laboratory is predictive of real-world cooperative behaviour [5,12,15–17] and is potentially heritable as well, as inferred from heritability of behaviour in the trust and dictator games [18] and positive correlations in game play among social dilemmas and the trust and dictator games [11,13] (note that Peysakhovich *et al.* [11] but not Carlsson *et al.* [13] found a strong positive correlation between behaviour in the dictator game and a social dilemma—PGG and PD, respectively).

Given the important role that within-group variation may play in the emergence and maintenance of cooperation, the proximate origins of this variation have received increasing attention in the past few years [19]. A number of researchers have sought a relationship between personality, as assessed with self-report questionnaires, and cooperation in experimental social dilemmas. The majority of these have implemented the five-factor model (FFM). The FFM, or 'big five' (extraversion, agreeableness, conscientiousness, openness to experience and neuroticism), is the most widely used schema for describing personality [20]. While there is broad agreement about the number of factors underlying personality, a number of different, closely related conceptualizations of these factors exist (see [20,21]). Generally, though, extraversion may be viewed as a tendency towards positive affect and sociability, agreeableness as empathy and motivation to cooperate, conscientiousness as self-discipline and forethought, openness as intellectuality and curiosity, and neuroticism as tendency towards negative affect and emotional instability.

Attempts to associate personality with behaviour in social dilemmas in the laboratory have met with conflicting results. Of the five dimensions, agreeableness, the 'prosocial personality' [22], would appear to be the most obvious candidate for predicting cooperative behaviour. Indeed, individuals high in agreeableness were more likely to be conditional cooperators in a 'one-shot' PGG (i.e. only one round of the game is played) [23] and more likely to cooperate unconditionally in a commons dilemma when the resource was severely threatened [24]. They were also more likely to cooperate in a one-shot PD, but only when the payoff structure was such that cooperation, rather than defection, was the optimal strategy [25], and in the first round of an iterated PD (IPD; i.e. the game is played for repeated rounds, as opposed to a one-shot game) [26]. However, neither Kurzban & Houser [27] nor Hilbig *et al.* [28] found a robust association between agreeableness and contributions in a

PGG. While Kurzban & Houser [27] used a circular PGG—that is, some participants knew the aggregate contribution to the public good before making their decisions—Hilbig *et al.* [28] conducted a PGG both without and with punishment. (A variant of the PGG involves punishment. Punishment is typically decentralized and thus involves the opportunity for each player to pay to reduce the payoffs of other players in their group.) Likewise, neither Pothos *et al.* [25] nor Hirsh & Peterson [29] observed a relationship between agreeableness and cooperation in an iterated or one-shot PD, respectively.

A potential relationship between extraversion and cooperation has also been posited. Hirsh & Peterson [29] suggest that extraversion may be associated with cooperation via two routes: a more rewarding subjective experience of cooperation [30], and a tendency towards positive affect, resulting in an increased expectation for partners' cooperation. The latter suggestion, however, runs contrary to evidence that positive affect may actually engender less cooperation. Van Lange *et al.* [31] and Boone *et al.* [32] suggest that sensation-seeking, a construct related to extraversion, should be associated with cooperation in situations with opportunity for repeated interactions, because attempting to initiate cooperation requires risk, and those high in sensation-seeking are drawn to risky situations.

Contrary to these hypotheses, Koole *et al.* [24] suggest that extraversion is negatively associated with cooperation in social dilemmas, because individuals low in extraversion are more averse to interpersonal conflict. Rather than the FFM, Skatova & Ferguson [33] consider variation in the behavioural activation and inhibition systems (BAS and BIS, respectively). According to reinforcement sensitivity theory [34], personality arises from variation in the sensitivity of these neurobiological motivational systems. Deyoung & Gray [35] claim that extraversion and neuroticism reflect variation in BAS and BIS functioning, respectively. Skatova & Ferguson [33] propose that individuals high on BAS (high activation of BAS) will contribute less to the public good in a one-shot PGG so as to maximize their own reward. In a situation with opportunity for repeated interactions, however, individuals high on BAS should cooperate so as to reap the rewards from reputation building.

Results from experimental economic games provide evidence both for and against the hypothesis that individuals high in extraversion are more cooperative in social dilemmas. In an IPD, greater cooperation was associated with the enthusiasm aspect of extraversion [29] (a mid-level aspect of extraversion according to the big five aspect scale) as well as sensation-seeking [32]. However, in a commons dilemma, extraversion was associated with faster depletion of a commonly held resource [24]. When the commonly held resource was severely threatened or others were depleting the resource at a slow rate, those low in extraversion were more cooperative [24]. In a similar vein, high BAS was associated with lower contributions to the public good [33] in a one-shot PGG and defection in a one-shot PD [25], adding potential support to a hypothesized negative relationship between extraversion and cooperation. At the same time, neither a positive nor a negative robust relationship was found between extraversion and cooperation in PGGs without [23,27,28] or with punishment [28] or in a one-shot PD [36].

Like extraversion, neuroticism has been posited to be both positively and negatively related to cooperation. On the one hand, Hirsh & Peterson [29] suggest anxiety about the repercussions of defection may lead to those high in neuroticism to cooperate. Similarly, Skatova & Ferguson [33] posit that high

BIS (which may underlie neuroticism [35]) could lead to cooperation by inhibiting reward-motivated behaviour (i.e. free-riding) in the presence of punishment [33]. Also, negative emotions may in fact motivate cooperation [31]; this may be most readily apparent with guilt [37].

Alternatively, Lu & Argyle [30] suggest that neuroticism is negatively related to cooperation because it entails a lack of concern about the well-being of others. Citing an association between high neuroticism and risk-aversion, Lönnqvist *et al.* [36] also suggest that neuroticism is negatively related to cooperation because of the risk of being defected upon. Ashton *et al.* [38] propose that neuroticism (low emotional stability) is a negative predictor of cooperation, because tendency towards negative affect would lead to a lack of forgiveness and therefore decay in reciprocal altruism.

As with extraversion, empirical results have been mixed. In support of a positive relationship between neuroticism and cooperation, neuroticism was associated with greater contributions in a circular PGG [27] and lower rates of defection in an IPD [29]. High BIS was negatively associated with free-riding (zero contributions) in a one-shot PGG where the subject was first informed that the other (imaginary) players had made high contributions [33]. Contradicting these results, neuroticism was associated with defection in a one-shot PD [36]; this association was partially mediated by risk aversion. It was also associated with decreased contributions in a one-shot PGG with punishment—i.e. when players were informed that other players would be given the opportunity to fine them based on their contributions [28]. To further confuse matters, Koole *et al.* [24] did not find a reliable effect of neuroticism on resource depletion in a collective resource dilemma, and Volk *et al.* [23] did not find a robust relationship between neuroticism (emotional stability) and contributions in a one-shot PGG. Similarly, neither Zettler *et al.* [39] nor Pothos *et al.* [25] observed a robust effect of neuroticism or BIS on cooperation in a one-shot PD.

There has been less discussion as to potential relationships between the other two dimensions in the FFM, conscientiousness and openness, and cooperation, perhaps, because these dimensions are considered less relevant to interpersonal interactions [24]. However, Lönnqvist *et al.* [36] found that openness was associated with cooperation in a one-shot PD; they attributed this to a relationship between openness and moral reasoning. Kurzban & Houser [27] observed larger contributions to the public good among individuals low in conscientiousness but did not have an explicit hypothesis with respect to the relationship between conscientiousness and cooperation.

Thus, far from an emerging consensus as to the effects of personality on cooperative behaviour in social dilemmas, the picture is one of failed replications and outright conflicting results. Why is there such inconsistency in these results? One possibility is publication bias and a lack of a relationship between personality and cooperative behaviour; it may be that noise in measurement is being published. However, there are two other possible causes of inconsistent results, which we explore here, that do not preclude a true relationship between personality and cooperation. The first is methodological. The experiments noted above have been conducted primarily by experimental economists and psychologists, who have their own distinct experimental traditions. Two practices that are universally employed in experimental economics but not always in psychology are (i) use of real (usually monetary) incentives and (ii) no deception (of the participant) [40].

Inconsistent implementation of these practices may lead to conflicting experimental results both within psychology and between economics and psychology.

Inconsistent use of real incentives tied to game behaviour is particularly troubling with respect to the above studies. Cooperation entails paying an immediate cost; if the behaviour being studied is completely 'free' to the actor, then the behaviour is not cooperation. Financial incentives (and corollary costs) are preferred by economists because, unlike other incentives, they can be easily measured and compared across studies, and it is assumed that there is no satiation [40]. Strong evidence that people often do not behave as economically rational beings in social dilemmas with real financial incentives [41] suggests that decisions are often based on multiple goals in addition to profit maximization, such as behaving in ways that appear socially desirable or improving the welfare of others. The assumption that the importance of these goals differs according to personality is implicit in attempts to link behaviour in social dilemmas to personality. If the stakes in an experiment are hypothetical, then the individual does not actually have to choose among these potentially conflicting goals, and we do not have a reliable estimate of the importance of non-profit-maximizing goals to the participant.

Indeed, Lönnqvist *et al.* [36] demonstrate that use of financial incentives can lead to different inferences with respect to the relationship between personality and cooperation in a social dilemma. They had subjects play either an incentivized or unincentivized one-shot PD with multiple choice (subjects chose how much of an endowment to transfer to another player, which was then doubled). As expected, subjects who played the incentivized version transferred less (about 25%) than those who played the hypothetical version. Only in the incentivized version were robust correlations between personality and behaviour found.

The above-noted social dilemma studies varied in their use of performance-based financial incentives. Performance-based monetary incentives were implemented by Volk *et al.* [23], Kurzban & Houser [27] and Skatova & Ferguson [33] in the PGG and by Lönnqvist *et al.* [36] and Kagel & McGee [26] in the PD. However, they were not used by Hilbig *et al.* [28] in one-shot PGG without and with punishment nor by Pothos *et al.* [25], Hirsh & Peterson [29] or Zettler *et al.* [39] in the PD. Boone *et al.* [32] and Koole *et al.* [24] motivated players with prizes awarded to a few participants with the highest payoffs in PD and commons resource games, respectively (Koole *et al.* [24] used a lottery, with odds determined by points accumulated in the commons game), but players could not attach an exact monetary value to the decisions with which they were faced.

The second possible reason for inconsistent results that we consider is of broader theoretical concern and has generated significant debate within psychology: that is, the relative importance of personality and situation in shaping consistent behaviour (the 'person–situation debate'; see Kenrick & Funder [42]). While there is now general agreement that personality is *not* necessarily manifest in constant behaviour across situations—that is, that behaviour arises from the interaction of both personality and situation, Funder [43] maintains that this has not been borne out in practice. At the same time, there is ongoing debate as to whether 'strong situations', i.e. situations in which external influences constrain optimal behaviour, can in fact generate uniform behaviour and diminish associations between personality and behaviour [44,45].

Many of the above-reviewed papers have discussed only main effects of personality, rather than an interaction between person and situation. There are some exceptions to this generalization. Some studies [24,28,33] mention the role of both personality and situation or environment in producing cooperative behaviour. Koole *et al.* [24] examine how the relationships between extraversion and agreeableness and resource use vary according to the behaviour of other players and state of the environment (resource depletion) and Hilbig *et al.* [28] investigate an interaction between honesty–humility, a proposed sixth dimension in the HEXACO model of personality, and (hypothetical) punishment in a PGG. While not explicitly considering an interaction between personality and situation, other studies consider the role of personality in cooperation when it is conditional versus absolute [23,27,32,33], when the probability the other player will cooperate is known [39], or when the ‘shadow of the future’ is present [32].

However, far richer situations must be considered if we are to use laboratory studies to make inferences about the relationship between personality and cooperation in the diverse realities that exist outside of the laboratory. Many, if not most, cooperative opportunities, such as paying taxes or braking for pedestrians, take place under the auspices of institutions that evolved or were designed to maintain cooperation. (We consider institutions as the set of rules that organize behaviour in repeated, structured interactions [46]). Institutions can alter the payoff structure in a social dilemma, via prescription of punishment (often fines) for non-cooperation, such that the disparity between the individual costs and benefits of cooperation are reduced. That punishment can be effectively used to maintain cooperation in social dilemmas is evident from studies both inside and outside the laboratory [41,47–51].

Punishment presents what can be considered a ‘strong situation’, and, clearly, the general possibility of punishment can have a large effect upon average behaviour. However, it would be erroneous to assume (i) a lack of significant individual variation in cooperative behaviour in the presence of punishment and (ii) that the explanation for cooperation in the presence of punishment lies entirely outside of the individual and can be understood without consideration of psychology [52].

Two dimensions of personality in the FFM, extraversion and neuroticism, should be especially pertinent to potential behaviour modification in the presence of punishment for non-cooperation [33]. Researchers have proposed that the core features of extraversion and neuroticism are sensitivity to reward and punishment, respectively [35,53–55]. In view of this, we propose that the relationships of these traits to cooperative behaviour will depend upon the institution under which the social action is taking place—specifically, the implementation of punishments (and rewards) that alter the payoff that can be achieved by free-riding.

Here, we test this with two versions of a PGG. In the no punishment game, individuals decide how much to contribute to the public good over repeated rounds (remaining in the same group) and have no recourse but to alter their own contributions in the face of free-riding. In the punishment game, however, after each round individuals may pay to reduce the income of other group members and thus punish free-riding.

Theory predicts that individuals high in extraversion will tend to free-ride in the absence of punishment (no punishment

game), relative to those low in extraversion, driven by the monetary rewards of non-cooperation, but be more cooperative when sanctions are present and free-riding is no longer monetarily rewarding (punishment game). Individuals high in neuroticism will not differ from those low in neuroticism in the absence of punishment (no punishment game). Faced with the threat of punishment in the punishment game, however, sensitivity to punishment will lead those high in neuroticism to be more cooperative than those low in neuroticism. An interaction between agreeableness and institution is not predicted, and thus an association between agreeableness and higher contributions in both the no punishment and punishment games is expected.

## 2. Methods

### (a) Participants

One hundred and eighty-four subjects (77 males, mean age 20.8 years and primarily students) took part in 1 of 11 experimental sessions in computer clusters on the Newcastle University Campus. Subjects received either a show-up fee or course credit (latter option for psychology students only). A show-up fee of £3 was increased to £5 for the last six sessions to motivate participation.

The number of participants per session ranged from 8 to 28. Participants were spaced such that there was either an empty computer or wall immediately adjacent to both sides of each participant. They were instructed not to communicate with each other in any way. A purpose-built website was used to communicate all instructions to participants, administer questionnaires and conduct the PGG.

### (b) Personality assessment

Prior to introduction to the PGG, personality was assessed via self-report with the freely available 120-item version of the International Personality Item Pool version of the NEO-PI-R [56].

### (c) Public goods game

The PGG structure used closely follows that of Herrmann *et al.* [51]. In the no punishment game, each player received 20 tokens per round and privately decided how many tokens to contribute to the group fund. The public good (the sum of tokens contributed by all group members to the project) was multiplied by 1.6 and then divided equally among all players. Thus, each player’s income for a given round consisted of the tokens she had retained plus 0.4 times the public good.

After reading instructions for the no punishment game, participants had to correctly answer a set of questions designed to assess their understanding before proceeding. Participants were told only that they would be introduced to a different version of the game after playing the current game for 10 rounds. Following the contribution stage of each round, each player was shown the contribution and income of all players in her group. Player identity could not be tracked from round to round.

Participants were then introduced to the punishment game, which was also played for 10 rounds. They had to correctly answer questions designed to assess their understanding of the new version of the game before proceeding. The punishment game differed from the no punishment game in that after players’ contributions and incomes for a given round were revealed, each player could assign up to 10 negative tokens to each other player. Each negative token cost the giver one token and the recipient three tokens. Each player then saw a summary screen that included the number and cost of negative tokens given and received and income adjusted for the cost of negative tokens.

**Table 1.** Estimated fixed effects coefficients and variance components for best candidate model for contributions to the public good. Lag refers to the previous round. *Lag contribution* is the lagged contribution of ego. *Lag mean contribution others* is the lagged mean contribution of the group, excluding ego. *Lag punishment received* is the lagged number of negative tokens ego received. *First round* is the initial round of either game (i.e. round 1 or round 11). Parentheses contain standard errors of the estimates. Ninety-five per cent confidence intervals for the coefficients are quadratic estimates.

	estimate	2.5%	97.5%
fixed effects			
intercept	0.55 (0.29)	−0.01	1.10
punishment game	1.45 (0.16)	1.13	1.77
first round	7.85 (0.31)	7.25	8.45
round	−0.05 (0.03)	−0.10	0.00
lag contribution	0.33 (0.02)	0.29	0.36
lag mean contribution others	0.51 (0.02)	0.47	0.54
lag punishment received	0.10 (0.04)	0.02	0.18
male	0.62 (0.31)	0.01	1.23
agreeableness	0.21 (0.15)	−0.08	0.50
extraversion	−0.52 (0.17)	−0.85	−0.19
neuroticism	−0.34 (0.17)	−0.66	−0.01
agreeableness × first round	0.45 (0.22)	0.01	0.88
extraversion × punishment game	0.37 (0.14)	0.10	0.63
neuroticism × lag punishment received	0.07 (0.04)	−0.01	0.14
variance components			
participant	1.64	n.a.	n.a.
residual	3.99	n.a.	n.a.

Immediately after completing the game, participants were paid their earnings and show-up fee in cash (one token = £0.015).

#### (d) Statistical analyses

We analysed the data and created all figures in the R statistical and computing environment [57] with the following packages: *rethinking* [58], *lme4* [59,60] and *bbml* [61].

To investigate the effects of the personality dimensions agreeableness, extraversion and neuroticism on the number of tokens contributed to the group fund, we first constructed a base Gaussian regression model with varying intercepts for individuals and fixed effects for the individual's sex (*male*) and the following game variables: *punishment game* (binary; whether punishment game), *first round* (first round of either game), *round*, *lag contribution* (lagged contribution of ego; lagged refers to the previous round), *lag mean contribution others* (lagged mean contribution of group members, excluding ego) and *lag punishment received* (lagged number of negative tokens received). Previous studies have shown these to be important predictors of contributions [41,62]. Ordered logit and binomial models were also considered but resulted in little change in predictions (see [63] for results); thus, for all subsequent analyses, we used the Gaussian model.

We then added the personality dimensions (*agreeableness*, *extraversion* and *neuroticism*) to the base model; these variables were standardized with mean of zero and standard deviation (s.d.) of one. Following our hypotheses, we created models that included interactions, singly and in combination, between each personality variable and the variables *punishment game* and *lag punishment received*. This was done in turn with *agreeableness*, *extraversion* and then *neuroticism*; the candidate model with the highest Akaike weight [61,64] was selected as the base model upon which subsequent models were built. We also modelled an interaction between *agreeableness* and *first round* based on

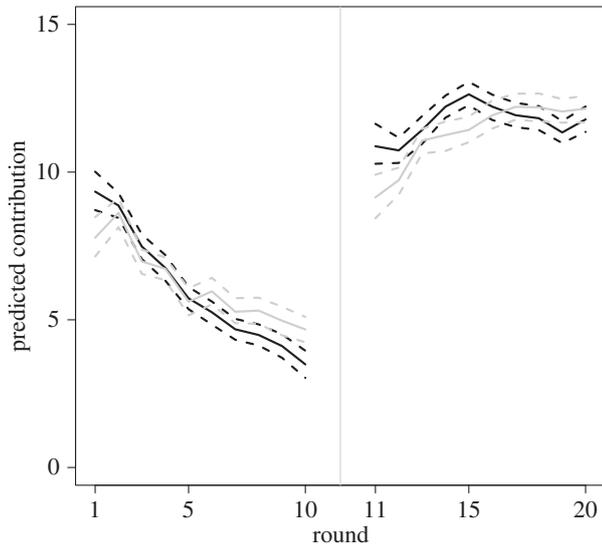
the results of Kagel & McGee [26], who observed an association between agreeableness and cooperation in the first round only of an IPD. The resulting model (henceforth, 'best candidate model') provides a far better fit to the data than does the base model that does not include personality (Akaike weight for base model < 0.001). Estimated fixed effect coefficients and variance components for the best candidate model are presented in table 1. Predicted contributions for players with high or low (+1 s.d., −1 s.d.) scores in the relevant personality dimension were generated from samples from the posterior density of the best candidate model, assuming a multivariate normal density. This was done 100 times for each game, across all 10 rounds.

We also checked to ensure that the different compensations received by participants (a show-up fee of £3, a show-up fee of £5 or course credit) did not impact our inferences with respect to personality. We did this by adding a categorical variable for participant compensation to the best candidate model. Estimated fixed effect coefficients and variance components are presented in the electronic supplementary material, table S1.

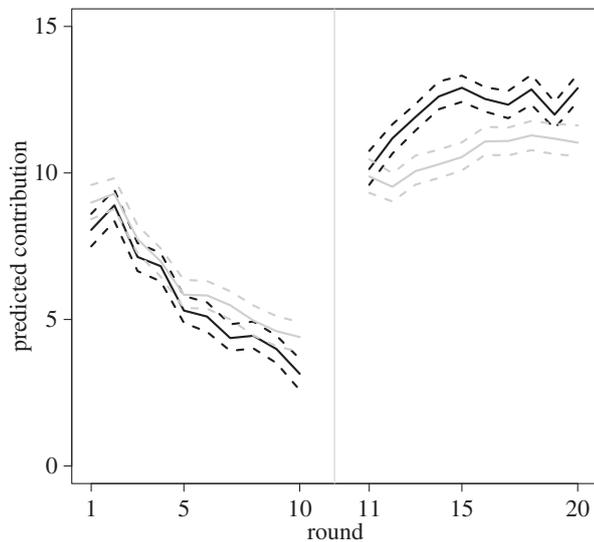
## 3. Results

### (a) Game

Players used the option to fine other players in their group during the punishment game, assigning a mean of 1.55 (s.d. = 3.17) negative tokens per round. Results from Poisson regression indicate that number of negative tokens received (i.e. punishment) was lower for larger contributions ( $\beta = -0.093$ ; 95% CI −0.099, −0.088). Mimicking previous results [41,47], fines directed towards free-riders led to increased contributions to the public good relative to the no punishment game (table 1; figures 1–3).



**Figure 1.** Predicted contribution, across rounds, in the no punishment (rounds 1–10) and punishment (rounds 11–20) games for players high (+1 s.d. or greater; black) and low (–1 s.d. or less; light grey) in agreeableness. Dotted lines illustrate 95% CIs.



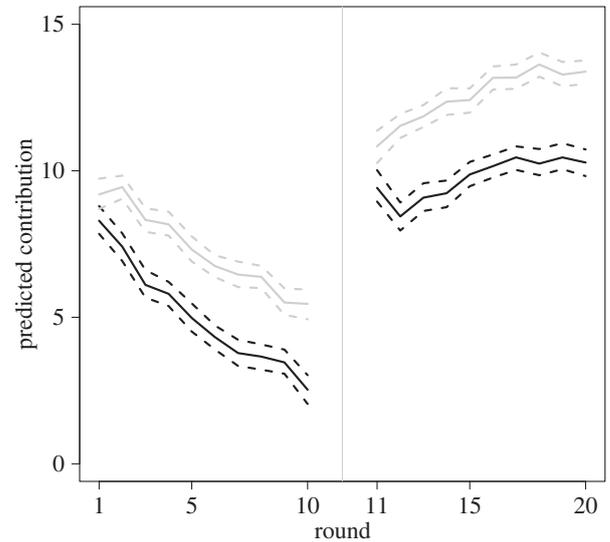
**Figure 2.** Predicted contribution, across rounds, in the no punishment (rounds 1–10) and punishment (rounds 11–20) games for players high (+1 s.d. or greater; black) and low (–1 s.d. or less; light grey) in extraversion. Dotted lines illustrate 95% CIs.

### (b) Agreeableness

As expected, the effect of agreeableness on contributions to the public good does not depend on institution. The best candidate model does not include an interaction between *agreeableness* and *punishment game* but does, however, include an interaction between *agreeableness* and *first round* (table 1). Plotted predictions reveal that an increase in agreeableness is associated with higher contributions in both the no punishment and punishment games, but that this is limited to the first round of each game (figure 1).

### (c) Extraversion

As predicted, the effect of extraversion on contributions to the public good depends on the institution. The best candidate model includes an interaction between *extraversion* and



**Figure 3.** Predicted contribution, across rounds, in the no punishment (rounds 1–10) and punishment (rounds 11–20) games for players high (+1 s.d. or greater; black) and low (–1 s.d. or less; light grey) in neuroticism. Dotted lines illustrate 95% CIs.

*punishment game* (table 1). While the main effect of *extraversion* on *contribution* is robustly negative, the interaction of *extraversion* with *punishment game* is positive. Plotted predictions show that players high in extraversion tended to free-ride in the no punishment game (figure 2). However, in the punishment game, they contributed more than those low in extraversion.

### (d) Neuroticism

Contrary to expectations, the effect of neuroticism on contributions to the public good was not contingent upon institution. The best candidate model includes an interaction between *neuroticism* and *lag punishment received* but not an interaction between *neuroticism* and *punishment game* (table 1). The plotted predictions reveal that players high in neuroticism contributed less in both the no punishment and punishment games.

## 4. Discussion

Initial attempts to assess relationships between personality (primarily using the FFM) and cooperative behaviour in experimental social dilemmas have been disappointing due to a lack of concurrence among studies. One of the possible reasons for these inconsistent results may be that researchers have largely focused on main effects of personality, rather than considering the interaction between personality and situation. Outside of the laboratory, cooperative actions are shaped by monetary and social rewards and punishment; many of these actions are governed by institutions that exist to maintain cooperation. Reputations are at stake, martyrs are promised virgins, and people are shot by firing squad for army desertion. Clearly, these strong situations can have an immense effect on average behaviour, but this does not preclude variation in cooperative behaviour via an interaction between personality and the situation.

As a first step, we considered an interaction between personality and a PGG in the laboratory that included the possibility for monetary punishment. Our focus was on punishment because it is often used to coerce cooperative behaviour (indeed, one of the oldest preserved texts, the Code of Hammurabi, details

punishments for various non-cooperative offenses, such as theft and fraud). Based upon prior research relating extraversion and neuroticism to sensitivity to reward and punishment [35,53–55], respectively, we expected these two FFM dimensions to be particularly relevant to understanding changes in behaviour when the payoff, or reward, from free-riding is altered via administration of punishment.

As predicted, the effect of extraversion on contributions to the public good was dependent upon the presence of a punishment institution. Individuals high in extraversion contributed less in the no punishment game than those low in extraversion. When cooperation, rather than free-riding, became rewarding (in the punishment game), individuals high in extraversion contributed more. Because we did not counter-balance the order of the no punishment and punishment games, we cannot exclude the possibility that the game-specific change in behaviour associated with extraversion is due to increased experience with the PGG. This is unlikely, however, Fehr & Gächter [41] showed the main effect of punishment on behaviour does not rest on game order, and given the general decrease in contributions over rounds in the no punishment game, it is difficult to conceive of a reason that individuals high in extraversion would contribute substantially more than those low in extraversion if they played the no punishment game again.

Moreover, an association between extraversion and strategic cooperation—i.e. cooperation only in the presence of punishment—is in line with theoretical predictions and is corroborated by another result. Strategic cooperation may also be revealed when comparing giving behaviour in the dictator game and the ultimatum game; in the latter, the second player can reject the proposal of the first player, resulting in zero payoff for both players. Scheres & Sanfey [65] found that high BAS (high activation of BAS may manifest as extraversion [35]) was associated with giving less in the dictator game and more in the ultimatum game. On the other hand, Hilbig *et al.* [28] did not observe an association between extraversion and a strategic shift in cooperation when participants played a one-shot PGG with and without punishment. However, in that study, both other players and monetary incentives were hypothetical, and in a recent meta-analysis of studies that used reward and punishment to promote cooperation, Balliet *et al.* [50] found that the effect of both rewards and punishment on cooperation were significantly smaller when incentives were only hypothetical.

In this study, we considered the reward of non-cooperation. Another possibility would be an association between extraversion and changes in cooperative behaviour in the presence of rewards for cooperation, rather than punishment for non-cooperation. The results of Campbell *et al.* [66] may be interpreted as support for this; they found that an association between extraversion and leadership in small groups charged with a task emerged only in the presence of extrinsic rewards for leadership.

While we have interpreted the observed association between extraversion and strategic cooperation within the framework of extraversion as reward sensitivity, our results are also consistent with another hypothesis. Ashton *et al.* [67] contend that the core feature of extraversion is a tendency to engage positive social attention. The monetary fines in the punishment game may also serve as a cue to threatened social standing, thus motivating individuals high in extraversion to cooperate. While individuals high in extraversion did contribute less in the no punishment game, which is more consistent

with extraversion as reward sensitivity rather than social attention-seeking, Ashton *et al.* [67] consider reward sensitivity as causally related to extraversion. Thus, the observed association between extraversion and lower contributions in the no punishment game would also be reconcilable with their hypothesis.

Also in line with prediction, individuals high in agreeableness contributed more to the public good, independent of game. However, this increase in cooperation was restricted to the first of 10 rounds in each game. This result is in concordance with that of Kagel & McGee [26], who found that agreeableness was associated with cooperation in only the first round of an IPD. That increased cooperation is limited to the first round is consistent with individuals high in agreeableness being motivated to cooperate, but suppressing or diverting that motivation when faced with evidence that they will be exploited in a particular situation [68]. Given this interpretation, it is interesting that the change in institution was apparently enough to reset expectations of others' cooperative behaviour (as players knew they were remaining in the same group) such that this initial desire to cooperate was manifest again.

Contrary to prediction, neuroticism was not associated with an increase in contributions to the public good in the punishment game. Neuroticism was associated with a small increase in contribution for each negative token received on the previous round, but the plotted predictions (figure 3) reveal that this was swamped by the overall effect of neuroticism on contributions, which was large and negative. Hilbig *et al.* [28] also observed a decrease in contributions to the public good associated with neuroticism in a one-shot PGG with (hypothetical) punishment, but not in a one-shot PGG without punishment. One possibility is that explicitly social, rather than monetary, punishment would be a more effective deterrent for individuals high in neuroticism [21].

Our observation that the effect of extraversion on contributions to the public good depends upon punishment demonstrates the importance of incorporating the interaction of person and situation in a social dilemma. We suggest that in studying how personality affects behaviour in social dilemmas, researchers consider the myriad institutions and associated rewards and punishments that influence cooperative behaviour. We also suggest that they be very explicit to both participants and colleagues about the situations that they are attempting to mimic in the laboratory and we contend that there is no such thing as a 'frameless' social dilemma in the laboratory. Studying the interaction of personality and framing in social dilemmas should also be fruitful. For example, people contribute more to the public good with a 'community frame' than a neutral frame [69]. Might there be an interaction between agreeableness and such labelling in a social dilemma?

However, we do not intend to leave the reader with the impression that studies of cooperation and personality should be limited to laboratory social dilemmas and to monetary incentives. Our issue is primarily with hypothetical, rather than non-monetary incentives. Indeed, teasing apart the role that social repercussions might play in punishment situations will require more creative, non-monetary incentives.

We believe that an interesting direction for future research lies in further consideration and comparisons of three-way interactions: that is, the person, the situation in the laboratory, and the sociocultural background of the person. Divergent institutional experiences should lead individuals with a given personality trait to adopt different adaptive

strategies over the course of development. Thus, the sociocultural background of participants will influence how they expect the institution in the laboratory to operate and how they expect other players in the laboratory to operate under that institution.

For example, in our study, we found that extraversion was associated with increased contributions in the punishment game. We attribute this to cooperation being more rewarding than free-riding in the punishment game, because other players paid to fine free-riders in the punishment game. We can infer that players expected free-riders to be fined, given that this increase in contributions is apparent from the first round (figure 1). However, if the same study was conducted in a culture where players did not preferentially punish free-riders [51], we might not expect to observe an association between extraversion and increased contributions in the punishment game. Similarly, we might not expect to observe an association between extraversion and increased contributions if punishment in the game was centralized but participants lived in a country with a weak rule of law.

Finally, the FFM is a descriptive model of personality. It is useful in part because its widespread implementation makes

comparison among studies straightforward. If more consistent associations can be drawn between personality and behaviour in social dilemmas, then new hypotheses can be generated that incorporate the potential biological processes that give rise to personality. For example, is the association between extraversion and strategic shifts in cooperative behaviour explained by variation in the brain's reward system? Following that, the next step in hypothesis generation can return to the three-way interaction mentioned above, combining biological functioning with the situation in the laboratory and diverse sociocultural backgrounds.

**Ethics.** This study was approved by the Newcastle University Faculty of Medical Sciences Ethics Committee.

**Data accessibility.** The dataset supporting this article has been uploaded as part of the electronic supplementary material.

**Authors' contributions.** K.B.S. conceived the study; K.B.S., D.N. and R.M. designed the study; K.B.S. conducted the study; R.M. and K.B.S. created the website used for the study and analysed the data; K.B.S. drafted the manuscript.

**Competing interests.** We declare we have no competing interests.

**Funding.** This work was supported by National Science Foundation Award 1003961 to KBS.

## References

- Dawes RM. 1980 Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193. (doi:10.1146/annurev.ps.31.020180.001125)
- Fischbacher U, Gächter S, Fehr E. 2001 Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404. (doi:10.1016/S0165-1765(01)00394-9)
- Kurzban R, Houser D. 2005 Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. *Proc. Natl Acad. Sci. USA* **102**, 1803–1807. (doi:10.1073/pnas.0408759102)
- Gächter S, Thöni C. 2005 Social learning and voluntary cooperation among like-minded people. *J. Eur. Econ. Assoc.* **3**, 303–314. (doi:10.1162/1542476054473323)
- Rustagi D, Engel S, Kosfeld M. 2010 Conditional cooperation and costly monitoring explain success in forest commons management. *Science* **330**, 961–965. (doi:10.1126/science.1193649)
- Ones U, Putterman L. 2007 The ecology of collective action: a public goods and sanctions experiment with controlled group formation. *J. Econ. Behav. Organ.* **62**, 495–521. (doi:10.1016/j.jebo.2005.04.018)
- Gunnthorsdottir A, Houser D, McCabe K. 2007 Disposition, history and contributions in public goods experiments. *J. Econ. Behav. Organ.* **62**, 304–315. (doi:10.1016/j.jebo.2005.03.008)
- Komorita SK, Parks CD. 1996 *Social dilemmas*. Boulder, CO: Westview Press.
- Gintis H. 2009 *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton, NJ: Princeton University Press.
- Volk S, Thöni C, Ruigrok W. 2012 Temporal stability and psychological foundations of cooperation preferences. *J. Econ. Behav. Organ.* **81**, 664–676. (doi:10.1016/j.jebo.2011.10.006)
- Peysakhovich A, Nowak MA, Rand DG. 2014 Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* **5**, 4939. (doi:10.1038/ncomms5939)
- Carlsson F, Johansson-Stenman O, Nam PK. 2014 Social preferences are stable over long periods of time. *J. Public Econ.* **117**, 104–114. (doi:10.1016/j.jpubeco.2014.05.009)
- Yamagishi T *et al.* 2013 Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organ. Behav. Hum. Decis. Process.* **120**, 260–271. (doi:10.1016/j.obhdp.2012.06.002)
- Blanco M, Engelmann D, Normann H. 2011 A within-subject analysis of other-regarding preferences. *Games Econ. Behav.* **72**, 321–338. (doi:10.1016/j.geb.2010.09.008)
- De Oliveira ACM, Eckel C, Croson RTA. 2012 The stability of social preferences in a low-income neighborhood. *South Econ. J.* **79**, 15–45. (doi:10.4284/0038-4038-79.1.15)
- Fehr E, Leibbrandt A. 2011 A field study on cooperativeness and impatience in the tragedy of the commons. *J. Public Econ.* **95**, 1144–1155. (doi:10.1016/j.jpubeco.2011.05.013)
- Algan Y, Benkler Y, Morell MF, Hergueux J. 2013 Cooperation in a peer production economy: experimental evidence from Wikipedia. Working paper, Sciences Po Department of Economics, Paris, France, pp. 1–31.
- Cesarini D, Dawes CT, Johannesson M, Lichtenstein P, Wallace B. 2009 Genetic variation in preferences for giving and risk taking. *Q. J. Econ.* **124**, 809–842. (doi:10.1162/qjec.2009.124.2.809)
- Wischniewski J, Windmann S, Juckel G, Brüne M. 2009 Rules of social exchange: game theory, individual differences and psychopathology. *Neurosci. Biobehav. Rev.* **33**, 305–313. (doi:10.1016/j.neubiorev.2008.09.008)
- Digman JM. 1990 Personality structure: emergence of the five-factor model. *Annu. Rev. Psychol.* **41**, 417–440. (doi:10.1146/annurev.ps.41.020190.002221)
- Denissen JJA, Penke L. 2008 Motivational individual reaction norms underlying the five-factor model of personality: first steps towards a theory-based conceptual framework. *J. Res. Pers.* **42**, 1285–1302. (doi:10.1016/j.jrp.2008.04.002)
- Graziano WG, Eisenberg N. 1997 Agreeableness: a dimension of personality. In *Handbook of personality psychology* (eds R Hogan, JA Johnson, SR Briggs), pp. 795–824. San Diego, CA: Academic Press.
- Volk S, Thöni C, Ruigrok W. 2011 Personality, personal values and cooperation preferences in public goods games: a longitudinal study. *Pers. Individ. Dif.* **50**, 810–815. (doi:10.1016/j.paid.2011.01.001)
- Koole SL, Berg AE, Van D, Hofstee WKB. 2001 On the social nature of personality: effects of extraversion, agreeableness, and feedback about collective resource use on cooperation in a resource dilemma. *Pers. Soc. Psychol. Bull.* **27**, 289–301.
- Pothos EM, Perry G, Corr PJ, Matthew MR, Bussemeyer JR. 2011 Understanding cooperation in the prisoner's dilemma game. *Pers. Individ. Dif.* **51**, 210–215. (doi:10.1016/j.paid.2010.05.002)
- Kagel J, McGee P. 2014 Personality and cooperation in finitely repeated prisoner's dilemma games. *Econ. Lett.* **124**, 274–277. (doi:10.1016/j.econlet.2014.05.034)

27. Kurzban R, Houser D. 2001 Individual differences in cooperation in a circular public goods game. *Eur. J. Pers.* **15**, 537–552. (doi:10.1002/per.420)
28. Hilbig BE, Zettler I, Heydasch T. 2012 Personality, punishment and public goods: strategic shifts towards cooperation as a matter of dispositional honesty – humility. *Eur. J. Pers.* **26**, 245–254. (doi:10.1002/per)
29. Hirsh J, Peterson J. 2009 Extraversion, neuroticism, and the prisoner's dilemma. *Pers. Individ. Dif.* **46**, 254–256. (doi:10.1016/j.paid.2008.10.006)
30. Lu L, Argyle M. 1991 Happiness and cooperation. *Pers. Individ. Dif.* **12**, 1019–1030. (doi:10.1016/0191-8869(91)90032-7)
31. Van Lange PAM, Joireman J, Parks CD, Van Dijk E. 2013 The psychology of social dilemmas: a review. *Organ. Behav. Hum. Decis. Process.* **120**, 125–141. (doi:10.1016/j.obhdp.2012.11.003)
32. Boone C, De Brabander B, van Witteloostuijn A. 1999 The impact of personality on behavior in five prisoner's dilemma games. *J. Econ. Psychol.* **20**, 343–377. (doi:10.1016/S0167-4870(99)00012-4)
33. Skatova A, Ferguson E. 2010 What makes people cooperate? Individual differences in BAS/BIS predict strategic reciprocation in a public goods game. *Pers. Individ. Dif.* **51**, 237–241. (doi:10.1016/j.paid.2010.05.013)
34. Gray J. 1970 The psychophysiological basis of introversion–extraversion. *Behav. Res. Ther.* **8**, 249–266. (doi:10.1016/0005-7967(70)90069-0)
35. Deyoung CG, Gray JR. 2009 Personality neuroscience: explaining individual differences in affect, behavior, and cognition. In *The Cambridge handbook of personality psychology* (eds PJ Corr, G Matthews), pp. 323–346. New York, NY: Cambridge University Press.
36. Lönnqvist J-E, Verkasalo M, Walkowitz G. 2011 It pays to pay – big five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Pers. Individ. Dif.* **50**, 300–304. (doi:10.1016/j.paid.2010.10.009)
37. Nelissen RMA, Dijker AJM, deVries NK. 2007 How to turn a hawk into a dove and vice versa: interactions between emotions and goals in a give-some dilemma game. *J. Exp. Soc. Psychol.* **43**, 280–286. (doi:10.1016/j.jesp.2006.01.009)
38. Ashton MC, Paunonen SV, Helmes E, Jackson DN. 1998 Kin altruism, reciprocal altruism, and the big five personality factors. *Evol. Hum. Behav.* **19**, 243–255. (doi:10.1016/S1090-5138(98)00009-9)
39. Zettler I, Hilbig BE, Heydasch T. 2013 Two sides of one coin: honesty–humility and situational factors mutually shape social dilemma decision making. *J. Res. Pers.* **47**, 286–295. (doi:10.1016/j.jrp.2013.01.012)
40. Hertwig R, Ortmann A. 2001 Experimental practices in economics: a methodological challenge for psychologists?. *Behav. Brain Sci.* **24**, 383–403.
41. Fehr E, Gächter S. 2000 Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994. (doi:10.1257/aer.90.4.980)
42. Kenrick DT, Funder DC. 1988 Profiting from controversy: lessons from the person–situation debate. *Am. Psychol.* **43**, 23–34. (doi:10.1037/0003-066X.43.1.23)
43. Funder DC. 2006 Towards a resolution of the personality triad: persons, situations, and behaviors. *J. Res. Pers.* **40**, 21–34. (doi:10.1016/j.jrp.2005.08.003)
44. Mischel W. 1977 The interaction of person and situation. In *Personality at the cross-roads: current issues in interactional psychology* (eds D Magnusson, NS Endler), pp. 333–352. Hillsdale, NJ: Lawrence Erlbaum Associates.
45. Cooper WH, Withey MJ. 2009 The strong situation hypothesis. *Personal. Soc. Psychol. Rev.* **13**, 62–72. (doi:10.1177/1088868308329378)
46. Ostrom E. 2005 *Understanding institutional diversity*. Princeton, NJ: Princeton University Press.
47. Yamagishi T. 1986 The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116. (doi:10.1037/0022-3514.51.1.110)
48. Güreker O, Irlenbusch B, Rockenbach B. 2006 The competitive advantage of sanctioning institutions. *Science* **312**, 108–111. (doi:10.1126/science.1123633)
49. Ostrom E. 1990 *Governing the commons: the evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.
50. Balliet D, Mulder LB, Van Lange PAM. 2011 Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* **137**, 594–615. (doi:10.1037/a0023489)
51. Herrmann B, Thöni C, Gächter S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–1367. (doi:10.1126/science.1153808)
52. Buss DM. 1991 Evolutionary personality psychology. *Annu. Rev. Psychol.* **42**, 459–491. (doi:10.1146/annurev.psych.42.1.459)
53. Lucas RE, Diener E, Grob A, Suh EM, Shao L. 2000 Cross-cultural evidence for the fundamental features of extraversion. *J. Pers. Soc. Psychol.* **79**, 452–468. (doi:10.1037/0022-3514.79.3.452)
54. Carver CS, Sutton SK, Scheier MF. 2000 Action, emotion, and personality: emerging conceptual integration. *Pers. Soc. Psychol. Bull.* **26**, 741–751. (doi:10.1177/0146167200268008)
55. Depue RA, Collins PF. 1999 Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behav. Brain Sci.* **22**, 491–517; discussion 518–69.
56. Johnson J. 2014 Measuring thirty facets of the five factor model with a 120-item public domain inventory: development of the IPIP-NEO-120. *J. Res. Pers.* **51**, 78–89. (doi:10.1016/j.jrp.2014.05.003)
57. R Core Team. 2013 R: a language and environment for statistical computing. See <https://www.R-project.org>.
58. Mcelreath R. 2014 rethinking: statistical rethinking book package. R package version 1.33. See <https://github.com/rmcelreath/rethinking>.
59. Bates D, Maechler M, Bolker B. 2014 lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-7. See <http://CRAN.R-project.org/package=lme4>.
60. Bates D, Maechler M, Bolker B, Walker S. Submitted. Fitting linear mixed-effects models using lme4. See <http://arxiv.org/abs/1406.5823>.
61. Bolker B, R Development Core Team. 2014 bbmle: tools for general maximum likelihood estimation. R package version 1.0.17. See <http://CRAN.R-project.org/package=bbmle>.
62. Ashley R, Ball SB, Eckel CC. 2010 Motives for giving: a reanalysis of two classic public goods experiments. *South. Econ. J.* **77**, 15–26. (doi:10.4284/sej.2010.77.1.15)
63. Schroeder KB, McElreath R, Nettle D. 2013 Variants at serotonin transporter and 2A receptor genes predict cooperative behavior differentially according to presence of punishment. *Proc. Natl Acad. Sci. USA* **110**, 3955–3960. (doi:10.1073/pnas.1216841110)
64. Akaike H. 1974 A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)
65. Scheres A, Sanfey AG. 2006 Individual differences in decision making: drive and reward responsiveness affect strategic bargaining in economic games. *Behav. Brain Funct.* **2**, 35. (doi:10.1186/1744-9081-2-35)
66. Campbell L, Simpson JA, Stewart M, Manning J. 2003 Putting personality in social context: extraversion, emergent leadership, and the availability of rewards. *Pers. Soc. Psychol. Bull.* **29**, 1547–1559. (doi:10.1177/0146167203256920)
67. Ashton MC, Lee K, Paunonen SV. 2002 What is the central feature of extraversion? Social attention versus reward sensitivity. *J. Pers. Soc. Psychol.* **83**, 245–252. (doi:10.1037/0022-3514.83.1.245)
68. Graziano WG, Tobin R. 2013 The cognitive and motivational foundations underlying agreeableness. In *Handbook of cognition and emotion* (eds MD Robinson, ER Watkins, E Harmon-Jones), pp. 347–364. New York, NY: Guilford Press.
69. Rege M, Telle K. 2004 The impact of social approval and framing on cooperation in public good situations. *J. Public Econ.* **88**, 1625–1644. (doi:10.1016/S0047-2727(03)00021-5)