

Patient and impatient punishers of free-riders

Antonio M. Espín^{1,*}, Pablo Brañas-Garza³, Benedikt Herrmann⁴
and Juan F. Gamella²

¹*Departamento de Teoría e Historia Económica, and* ²*Departamento de Antropología Social,
Universidad de Granada, Campus de la Cartuja s/n, 18071 Granada, Spain*

³*Department of Economics and International Development, Middlesex University Business School,
London NW4 4BT, UK*

⁴*Institute for Health and Consumer Protection, Joint Research Centre, European Commission, 21027 Ispra, Italy*

Costly punishment of cheaters who contribute little or nothing to a cooperating group has been extensively studied, as an effective means to enforce cooperation. The prevailing view is that individuals use punishment to retaliate against transgressions of moral standards such as fairness or equity. However, there is much debate regarding the psychological underpinnings of costly punishment. Some authors suggest that costly punishment must be a product of humans' capacity for reasoning, self-control and long-term planning, whereas others argue that it is the result of an impulsive, present-oriented emotional drive. Here, we explore the inter-temporal preferences of punishers in a multilateral cooperation game and show that both interpretations might be right, as we can identify two different types of punishment: punishment of free-riders by cooperators, which is predicted by patience (future orientation); and free-riders' punishment of other free-riders, which is predicted by impatience (present orientation). Therefore, the picture is more complex as punishment by free-riders probably comes not from a reaction against a moral transgression, but instead from a competitive, spiteful drive. Thus, punishment grounded on morals may be related to lasting or delayed psychological incentives, whereas punishment triggered by competitive desires may be linked to short-run aspirations. These results indicate that the individual's time horizon is relevant for the type of social behaviour she opts for. Integrating such differences in inter-temporal preferences and the social behaviour of agents might help to achieve a better understanding of how human cooperation and punishment behaviour has evolved.

Keywords: cooperation; competition; altruistic punishment; delay discounting; self-control

1. INTRODUCTION

Altruistic (costly) punishment refers to the readiness of humans to punish cheating group members at their own cost, even in one-shot interactions when no clear future returns are available. Such costly sanctions are a powerful instrument for protecting cooperation against exploitation by cheaters, and therefore help to sustain high cooperation levels [1–8]—a fact that puzzles scientists across the behavioural and biological sciences.

Despite increasing research interest, the mechanisms involved in costly punishment are poorly understood. Costly punishment of free-riders is supposed to be spurred by a moralistic drive to impose norms of fairness [2,3,5,7,9–15]. But what if the punishing individual is also a free-rider? Free-riders' punishment is unlikely to be driven by the same moral sentiments. More likely, the punishment by a free-rider could serve a competitive desire to achieve a higher pay-off than the other group members even at the punisher's own absolute cost [7,16–19]. Falk *et al.* [7] described the different nature of punishment by free-riders versus the punishment by cooperators. Punishment by free-riders is very sensitive to the relative cost of punishment: when no improvement of

relative standing is possible, free-riders no longer punish. By contrast, punishment by cooperators is barely influenced by the cost of punishment, as if cooperators were ready to teach cheaters a lesson at any cost, even if this means losing relative standing within the group.

This potentially fundamental difference in motivation must be kept in mind when investigating the possible drivers of punishment decisions. Moralistic punishment of norm violations is currently interpreted as either a product of humans' capacity for reasoning, self-control and long-term planning [9,10] or, at the opposite extreme, as a result of an impulsive, present-oriented emotional drive [11–15]. However, within the debate on the psychological roots of punishment, the possibility that some punishers (i.e. free-riders or norm-violators) may be guided by non-moralistic motives has not been deeply explored. This study focuses on the link between the punisher's inter-temporal preferences and the type of costly punishment she opts for, and explores whether the two antagonistic forces behind punishment may be partially predicted by this individual characteristic.

The relationship between inter-temporal preferences and punishment behaviour has so far been investigated only with the ultimatum game (UG). The UG is based on a stake that has to be shared between two individuals according to the proposal of one of them (proposer), which the second player (responder) can accept or reject. If the responder rejects, both players get nothing. Rejection

* Author for correspondence (kanton@ugr.es).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2012.2043> or via <http://rspb.royalsocietypublishing.org>.

of unfair offers is considered an act of costly fairness enforcement. In this game, impatient (present-oriented) individuals are more prone to reject low offers [11]. This result seems to back other researchers' interpretation that costly punishment is driven by impulsive emotions [12–15]. According to this view, an 'irrational' impulse would lead the punisher to disregard the future consequences of punishing norm violations. However, the standard UG does not allow disentangling whether observed behaviour is driven by competition on relative outcome—envy, in psychological terms—or by moralistic reactions against unfairness, because both natures of punishment would result in the rejection of low offers (that is, the same observable behaviour) [20]. In fact, some challenging neural evidence points to the involvement of self-control and long-term planning in rejection decisions [9,10].

We analysed the connection between inter-temporal preferences and the nature of punishment by cooperators and free-riders, using a one-shot public good game with punishment (PGP). The PGP makes it easier to disentangle different types of punishers by analysing their behaviour in the cooperation stage prior to punishment. Therefore, it allows determining whether the punishing individuals are in compliance with the norm or not—a dichotomy that has been found to have critical implications for cooperation and its evolution [4,8,16,21–23].

We used a one-shot procedure in order to elicit individuals' behavioural norms when punishing [5]. In our PGP, four anonymous players endowed with €10 first decided how much money to contribute to a common group pot. The sum of contributions in the pot was then multiplied by two and shared evenly among the four group members, which incentivized free-riding on others' cooperation. Therefore, although the socially efficient outcome in this game is full cooperation, the Nash equilibrium based on narrowly defined selfish rationality predicts full defection. The results of the contribution stage were then made public, and participants were allowed to reduce other group members' earnings at their own cost (punishment stage). Participants were allowed to spend up to €3 to reduce other group members' earnings, with each euro spent reducing the target player's earnings by €3. This 1:3 ratio allows punishment to be implemented with competitive as well as moralistic goals. However, a selfish individual would never make use of punishment in our one-shot anonymous setting. We also asked participants how much punishment they expected to receive from the other group members (see §4).

Individuals' manner of discounting delayed outcomes (i.e. their inter-temporal preferences or impatience) is a stable personal attribute [24] that unambiguously influences many fields of human behaviour [25]. High delay discounting (DD), measuring the willingness to prefer smaller rewards to larger but more delayed rewards, has been related to different scales of impulsivity and to lessened self-control [26] (however, see [27] for neural evidence suggesting that self-control and the evaluation of delayed rewards might respond to different psychological processes). As DD can predict inter-temporal decisions [25], it constitutes a helpful method for disentangling whether individuals perceive a given behavioural strategy as linked to early or delayed psychological incentives (see below for a discussion on an alternative interpretation).

We obtained DD functions for each participant through a standard task [28] computing their discounting parameter k from the hyperbolic characterization [29]. The parameter k represents the steepness of the discount function. The higher an individual's k , the more she discounts delays, and therefore the higher her impatience.

We ran field experiments with 160 participants (mean age 46.8 years; 64% females) from all walks of life in southern Spain. By means of non-laboratory experiments, we expected to attain higher heterogeneity among individuals' discount rates [30]. We indeed found important differences in DD among participants (see the electronic supplementary material, figure S4). The average k in our sample was 0.759 (± 0.034 , s.e.m.) in annual terms and related negatively to different income variables as in other field studies [31], but it was unrelated to individuals' contributions to the public good (see the electronic supplementary material, table S1). This lack of a relationship between DD and contributions might result from the incentives to strategically cooperate introduced by punishment (i.e. potential free-riders cooperate in order not to be punished), because others have found that DD and contributions are negatively correlated in one-shot public good games without punishment [32] (see the electronic supplementary material).

2. RESULTS

Sixty participants (37.5%) used the sanctioning mechanism at least once. The total amount of money reduced through punishment was €496 (from €2585 earned by cooperation), with 124 instances of punishment in total (€124 paid by punishers caused a reduction of €372 to the punished group members). In figure 1*a*, we show how the individual's DD and her deviation from other group members' mean contribution ('deviation' henceforth) impact on her willingness to punish. Individuals contributing more than €1 below the others' mean (i.e. deviation < -1) are included within the 'below average' category, those around the others' mean contribution (deviation between -1 and 1) within 'average' and high contributors (deviation $> +1$) within 'above average' (same classification as used by Gächter & Herrmann [8]). To facilitate visual interpretation, DD is depicted in colours, with k increasing from blue to red. Three categories of DD are constructed, each with one-third of the sample. The probability of punishing, $P(p)$, in the vertical axis represents the fraction of individuals using punishment. That is, $P(p)$ captures the proportion of punishers within each category of figure 1*a*. Evident differences exist between the punishment patterns of the three DD categories. However, because DD and deviation are continuous variables, the proper method to estimate the existing link is through regression analysis, which also allows controlling for other personal characteristics given the field origin of data. That is, the probability of punishing—whether an individual implements punishment or not—is regressed as a function of the punisher's deviation and k (probit regression with robust standard errors clustered at the group level).

Neither the positive effect of the punisher's deviation ($p > 0.5$) nor the negative effect of k ($p > 0.1$) on $P(p)$ reach significance (see the electronic supplementary material, table S2, model 2), but their interaction does ($p < 0.01$; model 4). The predictions of the model are

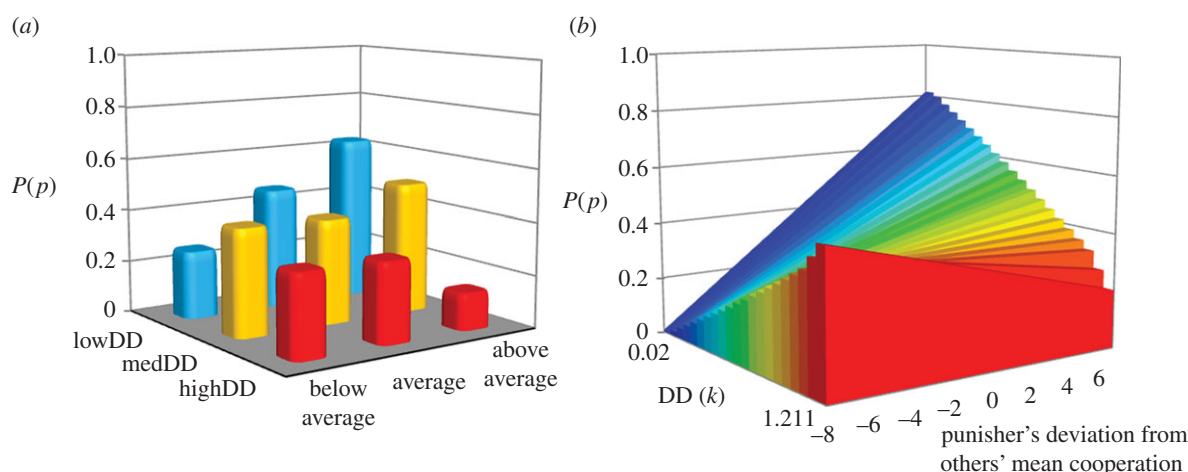


Figure 1. Delay discounting (DD), deviation and willingness to punish. This figure shows the willingness to punish ($P(p)$) as a function of the punisher's DD and relative contribution to the public good. Patient punishers are denoted in blue with impatience increasing towards red. (a) DD is divided into three quantiles, each with one-third of the sample (low, medium and high DD), whereas the punisher's deviation from other group members' mean cooperation is split into three theoretical categories: below average, average and above average. The number of observations in each category (lowDD, medDD, highDD) are: below average (23, 22, 19), average (13, 10, 13) and above average (19, 20, 21). (b) The punisher's DD and deviation are plotted as continuous variables impacting on $P(p)$ from the specification of model 4 of the electronic supplementary material, table S2 (including controls and keeping them at the mean), but using ordinary least squares coefficients (see the electronic supplementary material, table S3). When comparing both panels, it can be observed that the disparities increase as more extreme cases are reached. Although the punisher's deviation theoretically belongs to the interval $[-10, +10]$, in our sample it is reduced to the interval $[-8, +6.667]$. The four extreme values (corners) are (k , deviation, $P(p)$): (0.02, -8 , 0.007), (0.02, $+6.667$, 0.814), (1.211, -8 , 0.489) and (1.211, $+6.667$, 0.204). (a) Empirical data; (b) linear predictions.

shown in figure 1b. It is notable that the strong positive relationship between deviation and $P(p)$ capturing the behaviour of low-DD subjects reverses its slope as DD moves closer to its highest value. Wald tests reveal that DD is negatively related to $P(p)$ for extreme positive deviations (most cooperative individuals; $p < 0.01$), while for extreme negative deviations (strongly free-riding individuals), the sign of this relationship is positive ($p < 0.05$). In sum, punishment from the cooperative side is carried out by patient individuals, but those impatient individuals implement punishment when their own contributions are relatively low.

The next analysis is to explore who receives the punishment by patient cooperators and impatient free-riders. Figure 2 shows the predicted likelihood of punishing another group member depending on the punisher's and target's absolute cooperative levels (i.e. their raw contributions, from 0 to 10). Two different panels for the low and high categories of DD characterized in figure 1a are presented. For this model (see the electronic supplementary material, table S6), we use three observations per subject (one for each partner) with the likelihood of punishing each partner as the dependent variable (robust standard errors are clustered to account for correlation at the individual and group dimensions). The estimate of the interaction effect between the punisher's DD and cooperation is negative and significant ($p < 0.01$, model 4), thus supporting the previous result using the deviation variable. The axis in figure 2 representing the punisher's cooperation shows that low-DD, future-oriented individuals (figure 2a) are more likely to punish the more cooperative they are, whereas high-DD, present-oriented individuals (figure 2b) punish less the more cooperative they are. On the other hand, the target's cooperation always impacts negatively on the likelihood of her being

punished ($p < 0.01$), meaning that lower contributions are more likely to get punished. However, the interaction between the punisher's DD and the target's cooperation is largely insignificant in our model ($p > 0.6$). Hence, although free-riding behaviour is most likely to receive punishment, looking at the behaviour of punishers, it is patient cooperators and impatient free-riders who head the retaliation.

Analyses based on the punishment expected by the subjects reveal that patient and impatient individuals do not have different expectations about what levels of contribution are more likely to get punished (see the electronic supplementary material, table S5). Also, scrutiny of the subjects' expectations on punishment suggests that, in the eyes of impatient free-riders, punishing other free-riders seemed to be adequate when it came to fighting for the relative position (i.e. to beat the rival). This insight is extracted from the fact that impatient free-riders did not expect to receive a sufficient level of punishment to put at risk the pay-off advantage they had over cooperators (see the electronic supplementary material).

3. DISCUSSION

These results indicate that both previous interpretations of costly punishment might be correct if applied to the right subpopulation of punishers. Patience is characteristic of cooperators who decide to punish free-riders. Impatience, however, links to the punishment of free-riders by other free-riders. It has been shown that moralistic punishment benefits the society only in the long run [6]. Therefore, given its link with future orientation, it is possible that this kind of punishment is grounded in far-sighted collective motivations. On the other hand, the punishment implemented with non-moralistic goals by impatient

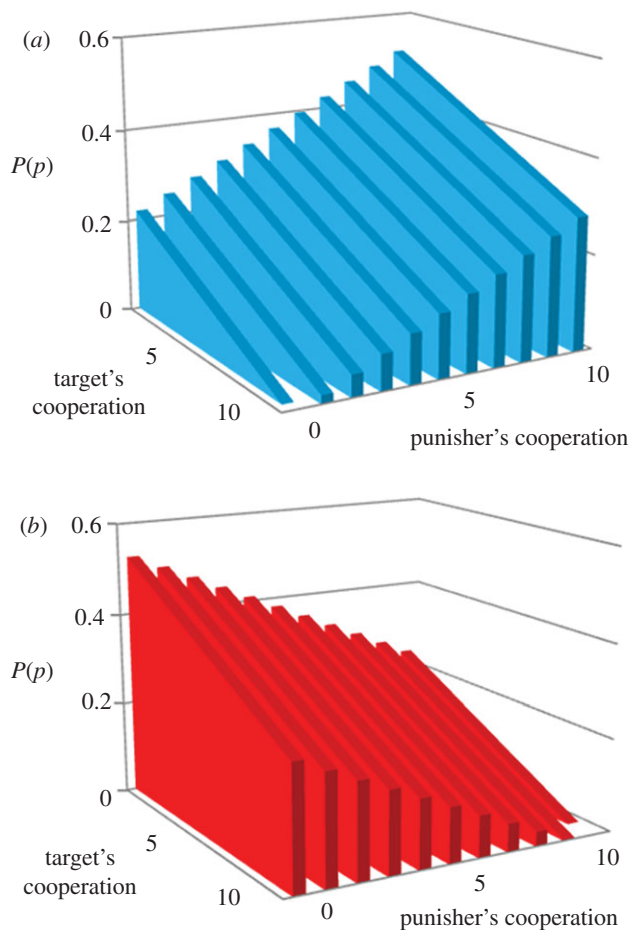


Figure 2. Delay discounting (DD), punisher's and target's cooperation. The figure shows linear predictions from model 4 of the electronic supplementary material, table S6 (including controls and keeping them at the mean) for the two extreme DD categories of figure 1*a*. (a) The predicted likelihood of low-DD (patient) individuals punishing another group member as a function of the contribution level of both the punisher and the target. (b) The same predictions for high-DD (impatient) individuals. For these plots, we use the mean k value within the DD category: 0.2884 and 1.211 for low and high DD, respectively. (a) Punisher is low DD; (b) punisher is high DD.

free-riders seems to be characteristic of aggressive, ultra-competitive behaviour, which has previously been found to be related to present orientation [33].

In the light of recent research on the role of intuition versus reflection in social decision-making [34,35], one might wonder whether the decisions on punishment are also shaped by intuition. Indeed, impatient responses in DD tasks have also been related to individuals' predisposition to follow their intuitions [36]. There might therefore exist an underlying common cognitive process leading individuals to choose smaller rewards that are received sooner (i.e. being impatient in DD tasks) and to behave intuitively without further deliberation. It would be interesting for future research to analyse response times of free-riders and cooperators when punishing in order to unravel whether our results are only due to individuals' inter-temporal preferences or instead driven by a more basic cognitive process [37,38].

From the results, we cannot reject the hypothesis that negative emotions spur moralistic punishment in the PGP

but, if this is the case, these emotions must be founded in more far-sighted, pro-social sentiments than mere self-centred revenge or spite. Given that previous research has found that more impatient responders in the UG are more likely to reject low offers [11], this new evidence also suggests a potential difference between cooperators' punishment in the PGP and responders' rejections in the UG. This possibility should be explored in deeper detail in further research analysing, for instance, whether impatient responders who reject unfair offers are themselves fair or unfair. Indeed, Carpenter [39] found that subjects with a competitive social value orientation [19] rather than 'fairmen' were responsible for most rejections in his experiments.

Our findings indicate that inter-temporal preferences and social behaviour are inter-related with each other in a much more complex fashion than discussed so far. Future research has to elicit the exact role of impulse, habits and reasoning for cooperation and defection [34], as well as for punishment and reward decisions. A better understanding of the role of inter-temporal preferences (and their possible context dependence) for shaping social and anti-social behaviour of agents might be important to refine our understanding of how human cooperation and punishment behaviour has evolved.

4. METHODS

One hundred and sixty inhabitants of small, semi-rural populations (1000–7000 inhabitants) in northern Granada (Andalusia, Spain) were invited to take part in experiments designed to elicit their DD and behaviour in a one-shot public good game with punishment. The participants, 103 of whom were female, were aged between 16 and 82 years (mean 46.8 ± 18.5 s.d.). The experiments were conducted in five sessions (32 subjects per session) at five different locations. Adapted standard instructions were read aloud, and several examples were illustrated on a whiteboard to ensure that the participants understood them. An experienced Spanish-speaking experimenter conducted all the sessions with an identical protocol (available in electronic supplementary material). The show-up fee was €5, and a drink and tapas after the experiment.

In the PGP, four anonymous players cooperated by contributing amounts of money from their endowment (€10) to a common pot. The sum of contributions in the pot was multiplied by 2 and evenly shared among the four group members. Hence, the individual returns of each monetary unit inside the pot, whatever their cooperative level, were $\alpha = 2/4$, meaning that contributing one unit had a cost of $1 - \alpha > 0$. Thus, every euro invested in cooperation increased the group's earnings by €2, but cost the investor 50 cents.

The participants cooperated simultaneously and were informed *ex ante* about the possibility of reducing the other group members' pay-offs at a personal cost after the results of the first contribution stage had been revealed. The price of punishment was one-third of the total reduction in income imposed on the punished subject. Reduction through punishment was limited to a maximum of €9 (i.e. three punishment opportunities, without restrictions on their distribution among partners) to rule out negative pay-offs. The subjects also had to report their expectations regarding the punishment they would receive from their partners.

For the statistical analyses, we used the likelihood of punishing and not the intensity of punishment because the

decision to punish and the decision about the amount are intrinsically different [40], and it was our aim to explore what is behind the decision of incurring *any* cost to punish others. Also, the existing limit for the amount of punishment implemented (max. €9) generates dramatically different decisions depending on the distribution of other group members' behaviours, and not only on their mean behaviour. However, the main results remain similar if we use the intensity of punishment as the dependent variable in the regressions (available upon request from the authors).

The discounting task for measuring participants' intertemporal preferences was a simplified version of Harrison *et al.* [28] involving real monetary incentives with a front-end delay procedure (both the sooner and the later reward are delayed). The task consisted of making 20 decisions on whether to receive €150 one month following the experiment or a higher amount (increasing from €151.50 to €225) after six extra months. The decision card contained a table with two columns (options A and B) and 20 rows. In each row, option A offered €150 to be received one month after the experiment, whereas option B offered a higher amount to be received seven months later. Thus, option B in the first row offered €151.50 and option B in the 20th row €225. The participants had to decide between option A and B in each of the 20 rows. The lower amount at which an individual was willing to wait half a year was considered her indifference point (between options A and B). We used the discounting parameter ($k \in [0.02, 1.211]$) from the hyperbolic characterization [29], calculated at the individual's indifference point, because it is the most commonly accepted functional form among behavioural scientists (see the electronic supplementary material, table S7 for analyses based on other discounting functional forms). Data available at <http://dx.doi.org/10.5061/dryad.r7c7p>.

Financial support from the Spanish Ministry of Science and Innovation (ECO2010-17049), the Regional Government of Andalusia (PO7-SEJ-02547) and Fundación Ramón Areces (R+D 2011) is gratefully acknowledged. We also thank J. A. Abril, A. Cortés, J. Martín, E. M. Muñoz, L. A. Palacios, M. Parravano, L. E. Pedauga, A. Quesada, M. Román and J. F. Ruano for their research assistance, as well as M. J. Crockett, S. Gächter, E. Lafuente and D. G. Rand for their helpful comments. Special thanks to the authorities and inhabitants of Benalúa, Darro, Deifontes, Iznalloz and Pedro Martínez. A.M.E., P.B.-G., B.H. and J.F.G. designed research and wrote the paper; A.M.E., P.B.-G. and J.F.G. performed research; A.M.E. and B.H. analysed the data.

REFERENCES

- 1 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)
- 2 Fehr, E. & Gächter, S. 2002 Altruistic punishment in humans. *Nature* **415**, 137–140. (doi:10.1038/415137a)
- 3 Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. 2007 Egalitarian motives in humans. *Nature* **446**, 794–796. (doi:10.1038/nature05651)
- 4 Herrmann, B., Thöni, C. & Gächter, S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–1367. (doi:10.1126/science.1153808)
- 5 Henrich, J. *et al.* 2006 Costly punishment across human societies. *Science* **312**, 1767–1770. (doi:10.1126/science.1127333)
- 6 Gächter, S., Renner, E. & Sefton, M. 2008 The long-run benefits of punishment. *Science* **322**, 1510. (doi:10.1126/science.1164744)
- 7 Falk, A., Fehr, E. & Fischbacher, U. 2005 Driving forces behind informal sanctions. *Econometrica* **73**, 2017–2030. (doi:10.1111/j.1468-0262.2005.00644.x)
- 8 Gächter, S. & Herrmann, B. 2009 Reciprocity, culture and human cooperation, previous insights and a new cross-cultural experiment. *Phil. Trans. R. Soc. B* **364**, 791–806. (doi:10.1098/rstb.2008.0275)
- 9 Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. 2006 Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314**, 829–832. (doi:10.1126/science.1129156)
- 10 Knoch, D., Gianotti, L. R. R., Baumgartner, T. & Fehr, E. 2010 A neural marker of costly punishment behaviour. *Psychol. Sci.* **21**, 337–342. (doi:10.1177/0956797609360750)
- 11 Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G. & Robbins, T. W. 2010 Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion* **10**, 855–862. (doi:10.1037/a0019861)
- 12 Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D. & Robbins, T. W. 2008 Serotonin modulates behavioural reactions to unfairness. *Science* **320**, 1739. (doi:10.1126/science.1155577)
- 13 Pillutla, M. M. & Murnighan, J. K. 1996 Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organ. Behav. Hum. Dec.* **68**, 208–224. (doi:10.1006/obhd.1996.0100)
- 14 van't Wout, M., Kahn, R. S., Sanfey, A. G. & Aleman, A. 2006 Affective state and decision-making in the ultimatum game. *Exp. Brain Res.* **169**, 564–568. (doi:10.1007/s00221-006-0346-5)
- 15 Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. 2003 The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755–1758. (doi:10.1126/science.1082976)
- 16 Eldakar, O. T., Farrell, D. L. & Wilson, D. S. 2007 Selfish punishment, altruism can be maintained by competition among cheaters. *J. Theor. Biol.* **249**, 198–205. (doi:10.1016/j.jtbi.2007.07.024)
- 17 Shinada, M., Yamagishi, T. & Ohmura, Y. 2004 False friends are worse than bitter enemies: 'altruistic' punishment of in-group members. *Evol. Hum. Behav.* **25**, 379–393. (doi:10.1016/j.evolhumbehav.2004.08.001)
- 18 Charness, G. & Rabin, M. 2002 Understanding social preferences with simple tests. *Q. J. Econ.* **117**, 817–869. (doi:10.1162/003353502760193904)
- 19 Van Lange, P. A. M. 1999 The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *J. Pers. Soc. Psychol.* **77**, 337–349. (doi:10.1037/0022-3514.77.2.337)
- 20 Kirchsteiger, G. 1994 The role of envy in ultimatum games. *J. Econ. Behav. Organ.* **25**, 373–389. (doi:10.1016/0167-2681(94)90106-6)
- 21 Eldakar, O. T. & Wilson, D. S. 2008 Selfishness as second-order altruism. *Proc. Natl Acad. Sci. USA* **105**, 6982–6986. (doi:10.1073/pnas.0712173105)
- 22 Helbing, D., Szolnoki, A., Perc, M. & Szabo, G. 2010 Evolutionary establishment of moral and double moral standards through spatial interactions. *PLoS Comput. Biol.* **6**, e1000758. (doi:10.1371/journal.pcbi.1000758)
- 23 Rand, D. G., Armao IV, J. J., Nakamaru, M. & Ohtsuki, H. 2010 Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* **265**, 624–632. (doi:10.1016/j.jtbi.2010.06.010)
- 24 Kirby, K. N. 2009 One-year temporal stability of delay-discount rates. *Psychon. Bull. Rev.* **16**, 457–462. (doi:10.3758/PBR.16.3.457)

- 25 Chabris, C. F., Laibson, D. I., Morris, C. L., Schuldt, J. P. & Taubinsky, D. 2008 Individual laboratory-measured discount rates predict field behaviour. *J. Risk Uncertainty* **37**, 237–269. (doi:10.1007/s11166-008-9053-x)
- 26 Kirby, K. N. & Finch, J. C. 2010 The hierarchical structure of self-reported impulsivity. *Pers. Individ. Differ.* **48**, 704–713. (doi:10.1016/j.paid.2010.01.019)
- 27 Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E. & Weber, E. U. 2010 Lateral prefrontal cortex and self-control in inter-temporal choice. *Nat. Neurosci.* **13**, 538–539. (doi:10.1038/nn.2516)
- 28 Harrison, G. W., Lau, M. I. & Williams, B. 2002 Estimating individual discount rates in Denmark: a field experiment. *Am. Econ. Rev.* **92**, 1606–1617. (doi:10.1257/000282802762024674)
- 29 Mazur, J. E. 1987 The effect of delay and intervening events on reinforcement value. In *Quantitative analyses of behaviour*, vol. 5 (eds M. L. Commons, J. E. Mazur, J. A. Nevin & H. Rachlin), pp. 55–73. Hillsdale, NJ: Erlbaum.
- 30 Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. 2010 Preference heterogeneity in experiments: comparing the field and laboratory. *J. Econ. Behav. Organ.* **73**, 209–224. (doi:10.1016/j.jebo.2009.09.006)
- 31 Tanaka, T., Camerer, C. F. & Nguyen, Q. 2010 Risk and time preferences: linking experimental and household survey data from Vietnam. *Am. Econ. Rev.* **100**, 557–571. (doi:10.1257/aer.100.1.557)
- 32 Curry, O. S., Price, M. E. & Price, J. G. 2008 Patience is a virtue: cooperative people have lower discount rates. *Pers. Individ. Differ.* **44**, 778–783. (doi:10.1016/j.paid.2007.09.023)
- 33 Nelson, R. J. & Trainor, B. C. 2007 Neural mechanisms of aggression. *Nat. Rev. Neurosci.* **8**, 536–546. (doi:10.1038/nrn2174)
- 34 Rand, D. G., Greene, J. D. & Nowak, M. A. 2012 Spontaneous giving and calculated greed. *Nature* **489**, 427–430. (doi:10.1038/nature11467)
- 35 Schulz, J. F., Fischbacher, U., Thöni, C. & Utikal, V. In press. Affect and fairness: dictator games under cognitive load. *J. Econ. Psychol.*
- 36 Frederick, S. 2005 Cognitive reflection and decision making. *J. Econ. Perspect.* **19**, 25–42. (doi:10.1257/089533005775196732)
- 37 Rubinstein, A. 2007 Instinctive and cognitive reasoning: a study of response times. *Econ. J.* **117**, 1243–1259. (doi:10.1111/j.1468-0297.2007.02081.x)
- 38 Haidt, J. 2001 The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814–834. (doi:10.1037/0033-295X.108.4.814)
- 39 Carpenter, J. 2003 Is fairness used instrumentally? Evidence from sequential bargaining. *J. Econ. Psychol.* **24**, 467–489.
- 40 Carpenter, J. & Matthews, P. H. 2009 What norms trigger punishment? *Exp. Econ.* **12**, 272–288. (doi:10.1007/s10683-009-9214-z)