



## OPEN

## Fair and unfair punishers coexist in the Ultimatum Game

SUBJECT AREAS:

HUMAN BEHAVIOUR  
EVOLUTIONARY ECOLOGYPablo Brañas-Garza<sup>1</sup>, Antonio M. Espín<sup>2</sup>, Filippos Exadaktylos<sup>3</sup> & Benedikt Herrmann<sup>4</sup>Received  
14 April 2014Accepted  
4 July 2014Published  
12 August 2014Correspondence and  
requests for materials  
should be addressed to  
B.H. (benedikt.  
herrmann@gmail.com)

<sup>1</sup>Business School, Middlesex University London, London NW4 4BT, UK, <sup>2</sup>GLOBE, Departamento de Teoría e Historia Económica, Universidad de Granada, Campus de la Cartuja s/n, 18071 Granada, Spain, <sup>3</sup>Istanbul Bilgi University, BELIS, Murat Sertel Center for Advanced Economic Studies, 34060 Eyup Istanbul, Turkey, <sup>4</sup>School of Economics, University of Nottingham, University Park Nottingham, Nottingham, NG7 2RD, UK.

**In the Ultimatum Game, a proposer suggests how to split a sum of money with a responder. If the responder rejects the proposal, both players get nothing. Rejection of unfair offers is regarded as a form of punishment implemented by fair-minded individuals, who are willing to impose the cooperation norm at a personal cost. However, recent research using other experimental frameworks has observed non-negligible levels of antisocial punishment by competitive, spiteful individuals, which can eventually undermine cooperation. Using two large-scale experiments, this note explores the nature of Ultimatum Game punishers by analyzing their behavior in a Dictator Game. In both studies, the coexistence of two entirely different sub-populations is confirmed: prosocial punishers on the one hand, who behave fairly as dictators, and spiteful (antisocial) punishers on the other, who are totally unfair. The finding has important implications regarding the evolution of cooperation and the behavioral underpinnings of stable social systems.**

A wealth of interdisciplinary research has attested to the pivotal role of fairness norms in explaining cooperation amongst unrelated individuals<sup>1–6</sup>. The Ultimatum Game (UG)<sup>7</sup> has been one of the most prolific set-ups for unraveling the nature of human fairness over the last years<sup>1,3,5,8–18</sup>. In this game, one player (the proposer) proposes a way to split a sum of money with another player (the responder). If the responder accepts the offer, both players are paid accordingly; if she rejects the offer, neither player is paid. The rejection of positive, albeit low offers is considered to be an expression of costly punishment of unfair behavior able to enforce the social norm of fairness<sup>2,3</sup>. Thus, the prevailing view is that individuals' (prosocial) preferences for fairness motivate the rejection of low offers in the UG. Such an argument has been used to support the strong reciprocity model of the evolution of human cooperation<sup>4,19–21</sup>. However, recent evidence now challenges this interpretation<sup>22</sup>.

Indeed, less prosocial motivations may be behind UG rejections. Rejections are equally compatible with competitive, spiteful motives<sup>23–26</sup>. A spiteful responder who is concerned with her own relative standing prefers a zero-zero outcome over one that leaves her below the proposer. Thus, she will reject any offer below the equal split, just like an individual concerned with the fairness norm. This implies that the mere observation of UG behavior is insufficient to determine the motivation of rejections. A crucial factor that should not be overlooked points directly to the punisher<sup>27,28</sup>: does she herself comply with the norm?

Recent research on cooperation games reveals that costly punishment is not only used by cooperators but also by non-cooperators who punish (other) non-cooperators<sup>29</sup>, cooperators<sup>30,31</sup> or both<sup>28</sup>. These punishment patterns, which cannot be reconciled with fairness motives, have been traced back to competitive, spiteful individuals aiming to increase their own relative payoff. Such findings are particularly important insofar as theoretical and empirical evidence demonstrates that the commons can be destroyed by the presence of “spiteful (unfair) punishers”, which can ultimately turn a sanctioning institution into a detrimental force for public cooperation<sup>(30–34; see<sup>35</sup> for a recent overview)</sup>.

This note investigates experimentally to what extent fair and unfair punishers coexist in the UG. To disentangle “prosocial” and “antisocial” types of punishment, this paper combines the rejection behavior of subjects in the UG with their decisions as dictators in a Dictator Game (DG)<sup>36</sup>. The DG is identical to the UG except that the second player is now passive, that is, she cannot reject the offer. As a result, generous offers by dictators are genuinely prosocial. A prosocial individual concerned about fairness will split the pie equally in the DG *and* reject unequal offers in the UG. A spite-driven, antisocial individual, however, will still reject unequal offers but transfer nothing in the DG (hence being totally unfair) in order to achieve the highest payoff differential. Note that pure selfishness also predicts a zero-transfer in the DG but never the rejection of any positive offers in the UG.



Therefore, the rejection of unequal but positive offers combined with zero-transfers in the DG is an unequivocal symptom of competitive spite.

We report data from two large-scale experimental studies. Study 1 ( $n = 754$ ) is a survey-experiment employing a representative sample of a city's adult population which was carried out at the participants' households. The pie to be split was €20 in each game. For UG responses, the strategy method was used in which the responder states whether she accepts/rejects any possible offer beforehand. Study 2 ( $n = 623$ ) is a replication of Study 1 in the laboratory employing university students (freshmen) as subjects (see Methods).

## Results

Figure 1 breaks down the sample into three groups according to participants' decisions in the DG: "unfair" refers to participants who offer zero in the DG, "fair" refers to those who make an equal split, while those who make an offer in between the two are labeled "remaining". For each group, the figure displays the percentage of responders who reject offers below the equal split in the UG. Study 1 [2] is captured by the left [right] panel.

The data clearly demonstrate that it is not only the "fair" but also the (totally) "unfair" dictators who reject unequal offers significantly more often than the "remaining" group (Probit model controlling for order effects in decisions; fair vs. remaining:  $p < 0.001$  in Study 1 and 2; unfair vs. remaining:  $p = 0.005$  in Study 1,  $p < 0.001$  in Study 2; see model 1 in Table S1 for Study 1 and Table S2 for Study 2 in the Supplementary Information (SI)). What is more, both groups are similarly likely to reject an unequal offer ( $p = 0.123$  in Study 1,  $p = 0.356$  in Study 2). As analyzed in more detail in the SI (see Figure S3), there is a statistically significant U-shaped, non-linear relationship between the two variables in both samples (all  $ps < 0.001$ ) when using the offers in the DG as a continuous explanatory variable (rather than comparing between the three DG groups). Furthermore, having decided first as dictator or as responder does not affect the reported relationship (no significant main or interaction order effects are observed in any study: all  $ps > 0.16$ ; see SI).

Thus, fair and unfair punishers coexist in the UG. In addition, in both samples fair dictators are more numerous than unfair ones (see the numbers on the top of the bars in Figure 1; the percentage of fair dictators is significantly higher than the percentage of unfair dictators according to a two-tailed binomial test:  $p < 0.001$  in both studies). This implies that fairness-based punishment is more frequent in both samples – which, nevertheless, should not necessarily be the case in samples taken from other populations/societies<sup>30</sup> (also, as discussed in the SI, methodological factors might influence these

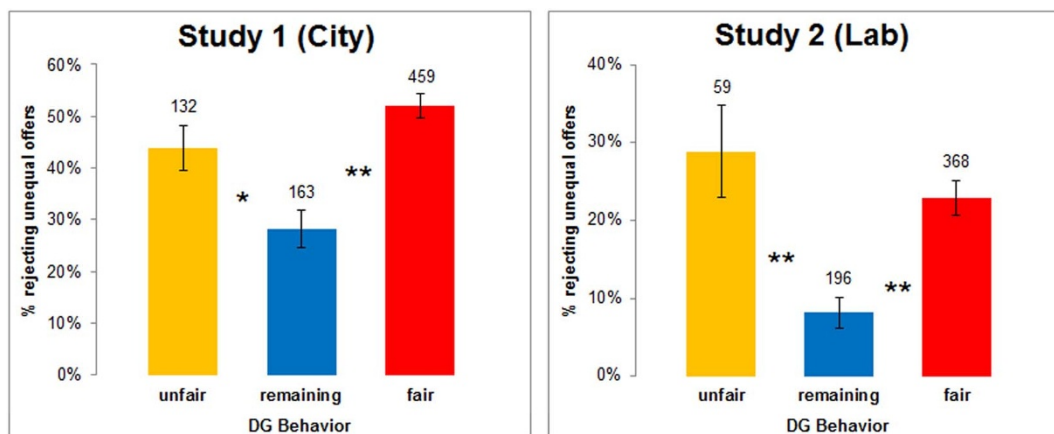
proportions). Indeed, among the UG responders who reject unequal offers in Study 1 [2], 17% [15%] are unfair dictators while 70% [72%] are fair dictators (these percentages are also significantly different according to a two-tailed binomial test:  $p < 0.001$  in both studies). Importantly, note that the relationship between DG offers and UG rejections holds even in the presence of differences between the two samples. In particular, in Study 1 the proportion of unfair dictators as well as the likelihood of rejecting unequal UG offers is higher compared to Study 2 (in both cases, two-tailed Fisher's exact test yields  $p < 0.001$ ).

## Discussion

The results show that punishment decisions in the UG are indistinguishably "prosocial" and "antisocial"; it is not one or the other, but both kinds of human behavior that shape the outcomes of the UG. Such a finding has important implications in interpreting previous results and designing future research.

One prominent example lies in the realm of behavioral and social neuroscience, where data from rejections in the UG have been extensively used to investigate the neurobiological basis of costly punishment. This research has implicated the brain areas responsible both for negative emotional processes (e.g., the anterior insula) and for executive control (e.g., the dorsolateral prefrontal cortex) in rejection behavior. Yet, there is much debate on the exact role of executive control. Some studies appear to indicate that executive control must be exerted to override the emotional impulse to punish unfairness at personal cost<sup>5,37</sup> whereas others suggest that it is the selfish impulse to accept an unfair offer which must be overridden in order to impose fairness through rejection, thus implying that punishment is an act of self-control<sup>8,38</sup>. Recently, more studies have shed light on these apparently contradictory observations<sup>13–16</sup> but the debate is far from closed. The results presented in this note indicate that there is a non-negligible fraction of rejections that are rooted not in normative, fairness-based judgments but instead in competitive, spiteful desires. It would in fact be hard to claim that a common neural mechanism underlies these extremely different natures of rejection behavior. Instead, one of the two might be overrepresented in some databases – which might have been due to the small sample sizes typically featured in brain studies – and, as suggested in<sup>29</sup>, this could explain part of the above controversy. Note also that the proportion of prosocial and antisocial punishers may vary dramatically across societies<sup>30,35,39</sup>.

Thus, in order to unravel the neurobiological basis of costly punishment, researchers should carefully investigate not only which behavior gets punished but also *who* is the punishing individual.



**Figure 1 | Willingness to reject unequal offers by DG groups.** Left [right] panel for Study 1 [2]. The horizontal axis depicts behavior in the DG: unfair (offer 0%), remaining (offer between 0 and 50%), fair (offer 50%). The numbers on top of the bars denote the total number of observations in each group. The vertical axis represents the percentage of individuals ( $\pm$  SE) who reject offers below 50% in the UG, i.e. whose minimum acceptable offer is the equal split (mean percentage: 45.49% in Study 1, 18.78% in Study 2). \*  $p = 0.005$ , \*\*  $p < 0.001$ .



However, this cannot be addressed using the standard UG as the only information researchers obtain from responders is whether they accept or reject a given proposal (or a number of them). Other experimental settings, or the combination of UG rejections with subjects' behavior in other frameworks, should be employed.

Additionally, the results are also important from the viewpoint of evolutionary biology and the social sciences. Costly punishment has been shown to be crucial in promoting cooperation<sup>40–43</sup>. Nevertheless, in the presence of spiteful punishers, social efficiency becomes difficult to sustain since spiteful behavior often leads to escalating conflict rather than to lasting cooperation. When sanctions are not used as norm-enforcement devices but instead at the service of dominance- or conflict-seeking behavior, their effects over social stability can be perverse<sup>30–32,44</sup>. If the punisher lacks the legitimacy to teach a moral lesson – because she does not comply with the social norm herself – the punished individual can view punishment as unjustified coercion. This might activate the mechanisms involved in competition with conspecifics instead of those involved in norm compliance, thus paving the way to inefficient, corrupt societies<sup>45,46</sup> rather than to efficient, cooperative ones<sup>47</sup>. In fact, corruption among the responsible for the enforcement of rules is recognized as a major source for the failure of social institutions<sup>48</sup>.

Therefore, special care has to be taken in the interpretation of rejection behavior as a mechanism to enforce the norms implicated in the maintenance of stable social systems. Extending the argument to the field of institutional design, failing to recognize the possible duality of motives behind punishment behavior in bilateral bargaining interactions can lead to less-than-optimal, or even counter-effective incentive mechanisms.

## Methods

The details of the survey-experiment have been reported elsewhere<sup>49</sup>. In both studies, subjects made their decisions in the UG (both roles) and the DG in random order. Subjects' decisions as proposers in the UG are not being used here as a measure of fairness since generous offers might equally be motivated by strategic self-interest (avoidance of rejection) and by other-regarding concerns, thus making them difficult to interpret<sup>50</sup> (indeed, zero offers in the UG are extremely rare). In contrast, the interpretation of subjects' offers in the DG is straightforward because they are not influenced by strategic concerns.

In the DG, subjects had to split a pie of €20 between themselves and another anonymous participant. Subjects decided which share of the €20 (in €2 increments) they wanted to transfer to the other subject. For the role of responder in the UG the strategy method was used<sup>51</sup>. That is, subjects had to state their willingness to accept or reject each of the following proposals (proposer's payoff [€], responder's payoff [€]): (20, 0); (18, 2); (16, 4); (14, 6); (12, 8); (10, 10). After making their decisions, participants in each study were randomly matched and one of every ten was selected for real payment (see Supplementary Information).

For the statistical analyses, we use Probit regressions with the likelihood that a subject rejects any unequal offer (i.e. whether her minimum acceptable offer is the equal split) in the UG as the dependent variable and DG behavior as the explanatory variable. Using the same database, a similar approach was employed in Staffiero et al.<sup>52</sup> for the study of the motivational drives behind the acceptance of zero offers in the UG.

All participants in the experiments reported in the manuscript were informed about the content of the experiment prior to participating. Verbal informed consent was obtained from participants in the city experiment (Study 1) since literacy was not a requirement to participate (this was necessary to obtain a representative sample), and all the instructions were read aloud by the interviewers. Written informed consent was obtained from participants in the lab experiment (Study 2). Anonymity was always preserved (in agreement with Spanish Law 15/1999 on Personal Data Protection) by randomly assigning a numerical code to identify the participants in the system. No association was ever made between their real names/addresses and the results. As is standard in socio-economic experiments, no ethic concerns are involved other than preserving the anonymity of participants. This procedure was checked and approved by the Vice-dean of Research of the School of Economics of the University of Granada; the institution hosting the experiments.

1. Rand, D. G., Tarnita, C. E., Ohtsuki, H. & Nowak, M. A. Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proc. Natl. Acad. Sci. USA* **110**, 2581–2586 (2013).
2. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
3. Henrich, J. et al. Costly punishment across human societies. *Science* **312**, 1767–1770 (2006).

4. Fehr, E. & Gintis, H. Human motivation and social cooperation: Experimental and analytical foundations. *Annu. Rev. Sociol.* **33**, 43–64 (2007).
5. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755–1758 (2003).
6. Sánchez, A. & Cuesta, J. A. Altruism may arise from individual selection. *J. Theor. Biol.* **235**, 233–240 (2005).
7. Güth, W., Schmittberger, R. & Schwarze, B. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367–388 (1982).
8. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314**, 829–832 (2006).
9. Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D. & Robbins, T. W. Serotonin modulates behavioral reactions to unfairness. *Science* **320**, 1739 (2008).
10. Henrich, J. et al. Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484 (2010).
11. Takahashi, H. et al. Honesty mediates the relationship between serotonin and reaction to unfairness. *Proc. Natl. Acad. Sci.* **109**, 4281–4284 (2012).
12. Szolnoki, A., Perc, M. & Szabó, G. Defense mechanisms of empathetic players in the spatial ultimatum game. *Phys. Rev. Lett.* **109**, 078701 (2012).
13. Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C. & Fehr, E. Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat. Neurosci.* **14**, 1468–1474 (2011).
14. Civai, C., Crescentini, C., Rustichini, A. & Rumiati, R. I. Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *Neuroimage* **62**, 102–112 (2012).
15. Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I. & Fink, G. R. Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Soc. Cogn. Affect. Neurosci.* **8**, 424–431 (2013).
16. Crockett, M. J. et al. Serotonin modulates striatal responses to fairness and retaliation in humans. *J. Neurosci.* **33**, 3505–3513 (2013).
17. Artinger, F., Exadaktylos, F., Koppel, H. & Sääksvuori, L. In others' shoes: do individual differences in empathy and theory of mind shape social preferences? *PLoS ONE* **9**, e92844 (2014).
18. Zhong, S., Israel, S., Shalev, I., Xue, H., Ebstein, R. P. & Chew, S. H. Dopamine D4 receptor gene associated with fairness preference in ultimatum game. *PLoS ONE* **5**, e13765 (2010).
19. Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1–25 (2002).
20. Bowles, S., & Gintis, H. *A Cooperative Species: Human Reciprocity And Its Evolution*. Princeton University Press (2011).
21. Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. Egalitarian motives in humans. *Nature* **446**, 794–796 (2007).
22. Yamagishi, T. et al. Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proc. Natl. Acad. Sci. USA* **109**, 20364–20368 (2012).
23. Kirchstetter, G. The role of envy in ultimatum games. *J. Econ. Behav. Organ.* **25**, 373–389 (1994).
24. Levine, D. K. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dynam.* **1**, 593–622 (1998).
25. Falk, A., Fehr, E. & Fischbacher, U. Reasons for conflict: lessons from bargaining experiments. *J. Inst. Theor. Econ.* **159**, 171–187 (2003).
26. Jensen, K. Punishment and spite, the dark side of cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 2635–2650 (2010).
27. Dreber, A. & Rand, D. G. Retaliation and antisocial punishment are overlooked in many theoretical models as well as behavioral experiments. *Behav. Brain Sci.* **35**, 24–24 (2012).
28. Falk, A., Fehr, E. & Fischbacher, U. Driving forces behind informal sanctions. *Econometrica* **73**, 2017–2030 (2005).
29. Espin, A. M., Brañas-Garza, P., Herrmann, B. & Gamella, J. F. Patient and impatient punishers of free-riders. *P. Roy. Soc. Lond. B. Bio.* **279**, 4923–4928 (2012).
30. Herrmann, B., Thöni, C. & Gächter, S. Antisocial punishment across societies. *Science* **319**, 1362–1367 (2008).
31. Rand, D. G. & Nowak, M. A. The evolution of antisocial punishment in optional public goods games. *Nature Commun.* **2**, 434 (2011).
32. Rand, D. G., Armao, IV, J. J., Nakamaru, M. & Ohtsuki, H. Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* **265**, 624–632 (2010).
33. García, J. & Traulsen, A. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *J. Theor. Biol.* **307**, 168–173 (2012).
34. Hilbe, C. & Traulsen, A. Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Sci. Rep.* **2** (2012).
35. Sylwester, K., Herrmann, B. & Bryson, J. J. Homo homini lupus? Explaining antisocial punishment. *J. Neuroscience Psychology Econ.* **6**, 167 (2013).
36. Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. Fairness in simple bargaining experiments. *Games Econ. Behav.* **6**, 347–369 (1994).
37. Tabibnia, G., Satpute, A. B. & Lieberman, M. D. The sunny side of fairness preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* **19**, 339–347 (2008).
38. Knoch, D., Gianotti, L. R., Baumgartner, T. & Fehr, E. A neural marker of costly punishment behavior. *Psychol. Sci.* **21**, 337–342 (2010).



39. Kocher, M., Martinsson, P. & Visser, M. Social background, cooperative behavior, and norm enforcement. *J. Econ. Behav. Organ.* **81**, 341–354 (2012).
40. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
41. Gächter, S., Renner, E. & Sefton, M. The long-run benefits of punishment. *Science* **322**, 1510–1510 (2008).
42. Mathew, S. & Boyd, R. Punishment sustains large-scale cooperation in prestate warfare. *Proc. Natl. Acad. Sci. USA* **108**, 11375–11380 (2011).
43. Egas, M. & Riedl, A. The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B* **275**, 871–878 (2008).
44. Gächter, S. & Herrmann, B. The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. *Eur. Econ. Rev.* **55**, 193–210 (2011).
45. Duéñez-Guzmán, E. A. & Sadedin, S. Evolving righteousness in a corrupt world. *PLoS ONE* **7**, e44432 (2012).
46. Abdallah, S., Sayed, R., Rahwan, I., LeVeck, B. L., Cebrian, M., Rutherford, A. & Fowler, J. H. Corruption drives the emergence of civil society. *J. R. Soc. Interface* **11**, 20131044 (2014).
47. Baldassarri, D. & Grossman, G. Centralized sanctioning and legitimate authority promote cooperation in humans. *Proc. Natl. Acad. Sci. USA* **108**, 11023–11027 (2011).
48. World Bank. *World Development Report 2004: Making Services Work For Poor People*. Oxford University Press, Oxford (2004).
49. Exadaktylos, F., Espín, A. M. & Brañas-Garza, P. Experimental subjects are not different. *Sci. Rep.* **3**, 1213 (2013).
50. Prasnika, V. & Roth, A. E. Considerations of fairness and strategy: experimental data from sequential games. *Q. J. Econ.* **107**, 865–888 (1992).
51. Mitzkewitz, M. & Nagel, R. Experimental results on ultimatum games with incomplete information. *Int. J. Game Theory* **22**, 171–198 (1993).
52. Staffiero, G., Exadaktylos, F. & Espín, A. M. Accepting zero in the ultimatum game does not reflect selfish preferences. *Econ. Lett.* **121**, 236–238 (2013).

## Acknowledgments

This paper has benefitted from the comments and suggestions of Simon Gächter and David G. Rand and participants of IMEBE 2013 (Madrid) and seminars at the Economics Department of the University of Malaga and Mural Sertel Center (Istanbul). FE acknowledges the post-doctorate fellowship granted by the Scientific and Technological Research Council of Turkey (TUBITAK). Financial support from the Spanish Ministry of Science and Innovation (ECO2010-17049), the Government of Andalusia Project for Excellence in Research (P07.SEJ.02547) and the Ramón Areces Foundation (R + D 2011) is gratefully acknowledged.

## Author contributions

P.B.G., A.M.E., F.E. and B.H. contributed equally to all parts of the research.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Brañas-Garza, P., Espín, A.M., Exadaktylos, F. & Herrmann, B. Fair and unfair punishers coexist in the Ultimatum Game. *Sci. Rep.* **4**, 6025; DOI:10.1038/srep06025 (2014).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>