Dark Triad personality traits vary across countries and predict antisocial behavior

Paul Deutchman[a] and Nichola Raihani[b*]

[a] Department of Psychology, Skidmore College, Saratoga Springs, NY, United States, 12866

[b] Department of Experimental Psychology, 26 Bedford Way, University College London, London, United Kingdom WC1H 0AP.

*Corresponding author: nicholaraihani@gmail.com

Abstract

Humans commonly punish exploitative group members but punishment is also frequently targeted at cooperative individuals. The proclivity for 'antisocial punishment' varies widely across societies, although the reasons for this variation remain unclear. Here, we identify personality factors associated with antisocial punishment, using a joy-of-destruction game with participants from India and the USA. This game allows players to harm one another, by destroying the partner's earnings, without any strategic incentive for doing so. High Dark Triad scores, implying the presence of personality traits underlying selfish and aggressive behavior, predicted destruction in this game. Participants from India scored higher on the Dark Triad scale than players from the USA, and were more likely than US-based participants to destroy the partner's endowment. These data suggest that Dark Triad personality traits could be a proximate explanation for antisocial behavior.

Keywords: spite; antisocial behavior; cross-cultural; Dark Triad; personality

Introduction

Punishment involves paying a cost to harm another individual (Clutton-Brock & Parker, 1995; Raihani, Thornton & Bshary, 2012). Punishment is typically aimed at cheating individuals (Bshary & Grutter 2002; Yamagishi, 1986; Fehr & Gächter, 2002) and can deter the target from cheating in future interactions (e.g. Bshary & Grutter, 2002; Raihani, Grutter & Bshary, 2010). Nevertheless, humans also sometimes punish cooperative or prosocial individuals, a phenomenon known as antisocial punishment (Falk, Fehr & Fischbacher, 2005; Herrmann, Thöni & Gächter, 2008; Bryson, Mitchell, Powers & Sylwester, 2014; Sylwester, Herrmann & Bryson,

2013; Bone, McAuliffe & Raihani, 2016; Prediger, Vollan & Herrmann, 2014). The tendency to invest in antisocial punishment varies across countries (Herrmann et al., 2008; Bryson et al., 2014; Sylwester et al., 2013; Bone et al., 2016), but the reasons for this variation are currently unclear.

One suggestion is that antisocial punishment reflects competition for status: although the punisher incurs a cost to punish, the target usually suffers more, allowing the punisher to elevate their status relative to the target (Bryson et al., 2014; Sylwester et al., 2013). Competition for status should be particularly pronounced where the scale of competition is more local than global (Gardner & West, 2004). Under global competition, an individual can benefit from helping in-group members if this increases the success of their group relative to other groups in the population. Under local competition, it is assumed that one group member's success comes at the expense of other group members, which in turn selects against within-group altruism and can favor the evolution of spite (Gardner & West 2004; Lehmann, Feldman & Rousset, 2009). Resource scarcity can exacerbate local competition: if individuals compete with neighbors for the same scarce resources, then behaviors aimed at harming neighbors can come under positive selection (Briones, Montaña & Ezcurra, 1998; Grossman & Mendoza, 2003; Lehmann et al., 2009). Indeed, in human societies, antisocial punishment and other spiteful behaviors are more common in resource-scarce environments (Prediger et al., 2014) or in low-GDP countries (Herrmann et al., 2008; Bryson et al., 2014; Sylwester et al., 2013; Bone et al., 2016) than in more benign settings. Furthermore, antisocial punishment is vastly reduced when the fee-to-fine ratio is ≥ 1 (Falk et al., 2005) – presumably because antisocial punishment does not allow the punisher to elevate their relative payoffs when this is the case.

Despite considerable focus on the adaptive significance of spiteful behaviors in humans, far less attention has been devoted to the proximate psychological mechanisms that underpin spiteful decisions. One candidate proximate psychological force that might predict variation in antisocial behavior is personality. Personality traits describe consistent individual differences in behavior that can reflect adaptive responses to differences in state or ecological conditions (Dingemanse & Wolf, 2010). Here, we focus on three broad personality traits as predictors for spiteful behavior: Machiavellianism, sub-clinical psychopathy, and narcissism. Together, these personality traits comprise the Dark Triad (Paulhus & Williams, 2002), which predisposes individuals towards antisocial and disagreeable behaviors (Jonason, Strosser, Kroll, Duineveld & Baruffi, 2015; Jonason, Duineveld & Middleton, 2015; Jonason & Krause, 2013), and are also associated with low self-control and impulsivity (Jonason & Tost, 2010), risky decision making (Jonason, Koenig & Tost, 2010), and a preference for short-term mating strategies (Jonason, Li, Webster & Schmitt, 2009). Results based on self-report measures have suggested that Dark Triad traits are linked to childhood unpredictability; findings which have been interpreted as evidence for adaptive variation along this personality dimension (Jonason, Icho & Ireland, 2016).

Materials and Methods

This project was approved by the UCL Research Ethics Board (project 3710/001). We recruited 600 participants (300 each from India and the USA) online using Amazon Mechanical Turk (hereafter 'MTurk', http//:www.mturk.com). We restricted our sample to these countries as the vast majority of MTurk workers hail from USA and India, respectively (Martin et al., 2014). Our

intended sample size of 75 participants per country for each of the four experimental conditions (described below) was based on the sample sizes used in previous similar studies (Abbink & Sadrieh, 2009; Abbink & Herrmann, 2011). Of these 300 recruits, data from 22 were ultimately excluded because participants either failed a comprehension question, did not answer all of the survey questions, or took part in the task more than once. Thus, we were left with data from 578 participants (280 from the United States and 298 from India).

Participants first earned an endowment ($2.20 or $1.80, see below) by completing a word search, before taking part in a one-shot, joy-of-destruction game with a partner (allocated ex-post, following Rand, 2012; Raihani, Mace & Lamba, 2013). Finally, all participants completed the Short Dark Triad (SD3) questionnaire (Jones & Paulhus, 2014). The SD3 comprises 27 questions, nine for each Dark Triad trait. Answers are given on a 5-point Likert scale ranging from 1-"Diagree Strongly" to 5-"Agree strongly". We calculated mean scores for each of the three personality traits measured, and then averaged these to obtain a composite Dark Triad score for each participant (following Jonason et al., 2010). One shortcoming of the SD3 questionnaire is that it incorporates only 5 reversals (questions where strongly agreeing indicates lower, rather than higher Dark Triad traits). This is potentially problematic if participants from different cultures are more likely to acquiesce with statements or are more likely to select the most extreme responses on the scale, as has been previously suggested (Smith 2004). In our study, the fact that India is more collectivist than the USA (Hofstede 2001) could produce a greater acquiescence bias among Indian subjects or could result in more extreme responses from the Indian subjects (c.f. Smith 2004) – potentially therefore producing a spurious finding of higher Dark Triad scores among Indian subjects. To check that acquiescence bias or tendency to select

extreme answers would not systematically bias the Dark Triad scores of Indian and US participants, respectively, we (i) compared the average scores for the reversed statements and (ii) compared the number of extreme responses (subjects selecting 1 or 5 on the scale) for the 27 questions for the Indian and American players.

The joy-of-destruction game is a two-player task, where both players earn an endowment and can then simultaneously decide whether to destroy a portion of their partner's money (Abbink & Sadrieh, 2009). The joy-of-destruction game lacks incentives for punishment: choosing to destroy is therefore a proxy for spite and is likely to stem from similar proximate motives as those involved with antisocial punishment. In the current study, participants were informed that they could destroy $0.00 to $1 of the partner's earned income, in $0.20 increments. Participants were randomly allocated to one of four conditions, using a 2x2 experimental design: Costly / No Cost of destruction, and Hidden / Open destruction. The Costly and No Cost conditions differed in whether destroying the partner's endowment was costly to the perpetrator (pay $0.10 to reduce their partner's bonus by any amount) or could be done for free. Based on past literature, which suggests punishment depends on the cost-to-impact ratio (Anderson & Putterman, 2006; McCullough, Kurzban & Tabak, 2013) we predicted that participants would be more likely to destroy the partner's bonus in the No Cost condition than in the Costly condition. The Hidden / Open destruction conditions manipulated the extent to which a participant was able to infer that their earnings had been destroyed by the partner rather than by chance. In the Open condition, participants were informed that their destruction decision would be revealed to their partner after both players made their decisions. Participants in the Hidden condition were informed that, in addition to any destruction they chose, a portion of their partner's endowment might also be

destroyed by chance, thereby allowing them to 'hide' their destruction behind this random process. The chance destruction was \$0.00, \$0.20, \$0.40, \$0.60, or \$0.80; with participants being equally likely to be assigned to any of these conditions. To account for the additional destruction, participants allocated to the Hidden condition were endowed with \$2.20 for completing the word search, compared with \$1.80 for the participants in the Open condition. Previous research has shown that players are less likely to destroy the partner's endowment in the Open condition (Abbink & Sadrieh, 2009) and we expected to replicate that effect here.

First, we asked whether composite Dark Triad score varied across US-based and India-based respondents, using a two-sample t-test. Then, we explored the factors influencing destruction decisions, using a general linear model (GLM). We had insufficient data points to explore the factors affecting the different amounts that participants chose to destroy. Instead, we created a dummy binary variable, 'destroyed', where 0 = did not destroy the partner's bonus and 1 = destroyed some or all of the partner's bonus. This variable was set as the response term in a GLM with binomial error structure. We included cost of destroying (Cost / No Cost), visibility of destruction (Open / Hidden) and country of origin (India / USA) as explanatory terms. We also wanted to include the composite Dark Triad score but, since this varied significantly across the two countries (see results), we instead used the residual variance in Dark Triad score (hereafter 'residual Dark Triad') that was not explained by country of origin. To obtain this residual measure, we ran a linear model, with composite Dark Triad score as the response term and country of origin as the explanatory term and extracted the residuals from this model. In addition to the main terms, we also considered the following interactions: Cost x Country, Cost x

Visibility, Country x Visibility, Residual Dark Triad x Cost, and Residual Dark Triad x Visibility.

Statistical Methods

To obtain parameter estimates for the terms we included in the GLM, we used a model selection approach with model averaging (Burnham & Anderson, 2002), as described in (Grueber, Nakagawa, Laws & Jamieson, 2011). Under this philosophy, a candidate set of models are defined a priori and are then compared to one another using the Akaike Information Criterion, corrected for small sample sizes (AICc, Hurvich & Tsai, 1989). Lower AICc values indicate a greater degree of support for the candidate model. All input parameters were standardized according to Gelman (2008) so that effect sizes associated with different parameters can be interpreted on the same scale. We specified a global model, containing all fixed effects and interactions specified above, and then used the MuMIn (Bartoń, 2009) package to find the subset of models that were within 2AICc units of the best model (the model with the lowest AICc value). This subset of models is referred to as the top model set. To obtain weighted parameter estimates and associated confidence intervals, we averaged parameter estimates across the top model set. Model averaging therefore takes into account uncertainty as to the true 'best' model. Here, we report the full model averaged estimates, which yield conservative effect sizes for parameters that are not included in all of the top models. All data and R code are available (Raihani & Deutchman, 2016).

Results

Of the 578 participants, 87 (15.1 %) chose to destroy some or all of the partner's bonus. Of these 87, 34 (39.0 %) decided to destroy all the partner's bonus, while the remainder destroyed a smaller portion of the partner's bonus. The mean score for Machiavellianism was 3.35 ± 0.03, for Narcissism was 2.85 ± 0.03, and for Psychopathy was 2.30 ± 0.03. Overall, all three personality measures were highly correlated (Mach-Narc: $r = 0.45$, Narc-Psych: $r = 0.41$, Mach-Psych: $r = 0.52$; all $P < 0.001$) so we combined the scores into a composite Dark Triad score for each participant. This was done by standardizing the scores associated with each trait, then averaging to obtain a composite measure.

Participants from India and the USA, respectively, showed systematic differences in Dark Triad personality scores (two-sample t-test, $t = 11.8$, $df = 521$, $p < 0.001$), with India-based participants having higher composite Dark Triad scores than their US-based counterparts (mean ± sem: 0.34 ± 0.04 and -0.37 ± 0.05, respectively; Fig. 1). Higher Dark Triad scores among Indian players were not due to a systematic tendency to acquiesce with the statements: when looking at the mean scores for the five reversed statements (where lower scores predict Dark Triad traits), India-based players scored 3.23 ± 0.04 compared with 3.53 ± 0.04 for US-based players, indicating that the responses for the reversed statements matched the overall pattern of responses rather than an acquiescence bias (2-sample t-test: $t = 3.52$, $df = 587$, $p < 0.001$). There was also no tendency for Indian players to show an extreme response bias – in fact US-based players more often selected the extreme responses (mean number of extreme responses *India:* 12.0 ± 0.32; *USA:* 13.1 ± 0.33; 2-sample t-test: $t = 2.29$, $df = 588.9$, $p = 0.02$).

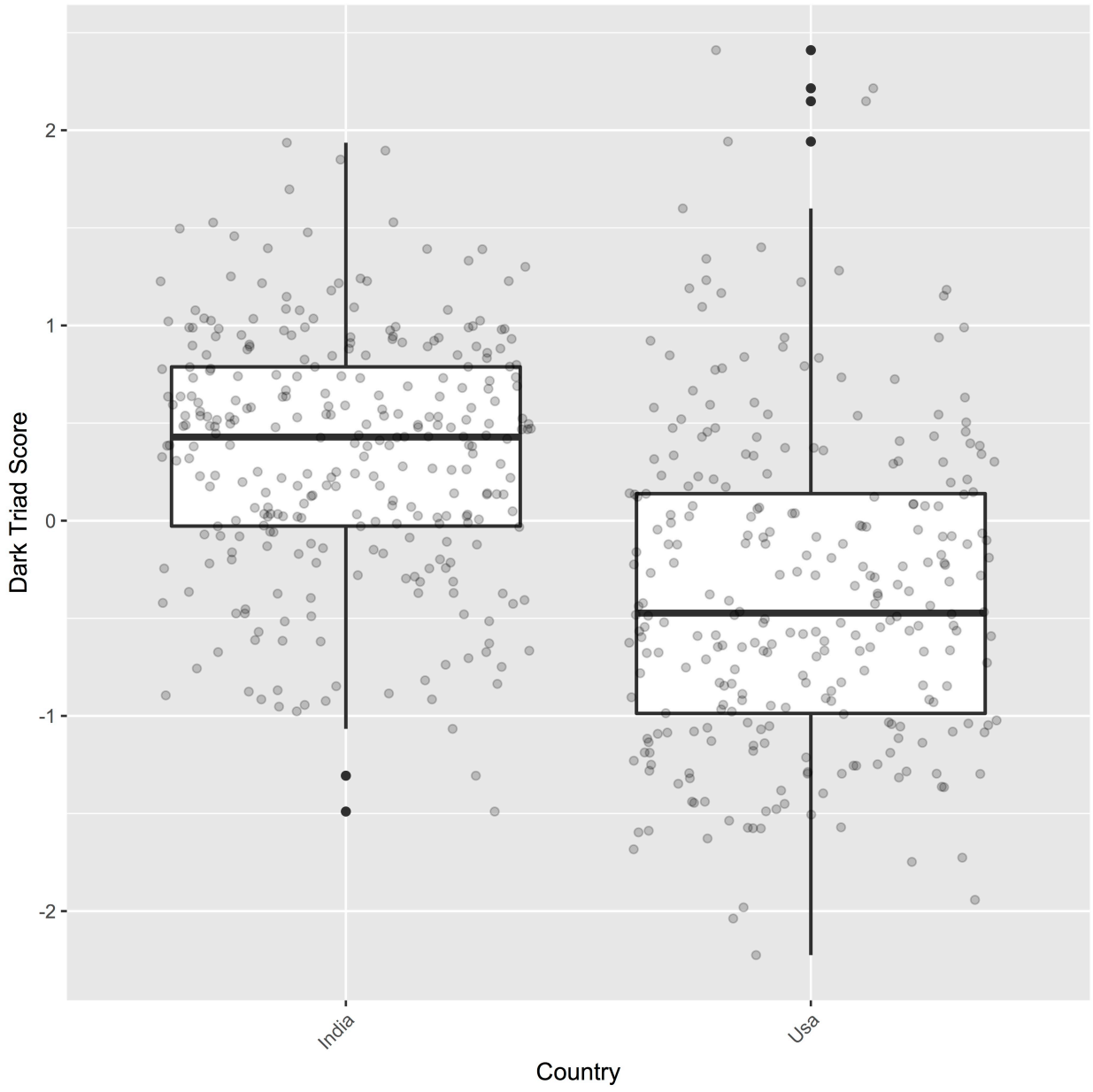

Figure 1. Boxplots of composite Dark Triad scores for participants from India and the USA. The box denotes the interquartile range, with the median for each country indicated by a solid horizontal black line. Vertical lines indicate scores that fall outside the middle 50 %. Grey dots are raw data points and black dots are outliers.

The decision to destroy some or all of the partner's bonus was predominantly predicted by country: participants from India were more likely than participants from the USA to destroy some or all of the partner's bonus (estimated probability of someone from USA destroying bonus, relative to someone from India: -1.95, CI: -2.57, -1.33; Table 1, Table 2).

Rank	Parameters	df	AICc	w^i
1	**Country + Dark Triad**	**3**	**432.2**	**0.39**
2	Cost + Country + Dark Triad	4	432.8	0.28
3	Country + Dark Triad + Visibility	4	433.8	0.17
4	Country + Dark Triad + Visibility + Visibility:Dark Triad	5	434.0	0.16

Table 1. GLM to investigate the factors affecting propensity to destroy the partner's bonus (1/0). 'Dark Triad' is the residual variance in personality score that is not explained by country of origin – a high score indicates a person who is higher on the Dark Triad personality traits. The table shows the top models (those that are within 2 AICc units of the top model). The best model is highlighted in bold.

Parameter	Estimate	Unconditional SE	Confidence Interval	Relative Importance
Intercept	-2.13	0.16	(-2.45, -1.81)	
Country (India)	-1.95	0.32	(-2.57, -1.33)	1.00
Dark Triad	1.07	0.27	(0.54, 1.61)	1.00
Cost (No Cost)	-0.08	0.19	(-0.45, 0.28)	0.28
Visibility (Hidden)	0.04	0.16	(-0.27, 0.35)	0.33
Visibility*Dark Triad	0.12	0.34	(-0.55, 0.79)	0.16

Table 2. Full-model averaged estimates, unconditional standard errors, confidence intervals and relative importance for the terms included in the top model set in Table 2. For categorical variables, the reference category is indicated in parentheses.

In addition, the presence of Dark Triad traits had an additive positive effect on destroying the partner's bonus, indicated by the effect size associated with the residual variance in Dark Triad score (estimate: 1.07; CI: 0.54, 1.61; Fig. 2). Somewhat surprisingly, there were no meaningful effects of either the cost or the visibility of destruction on participants' destruction decisions (Table1, Table 2).

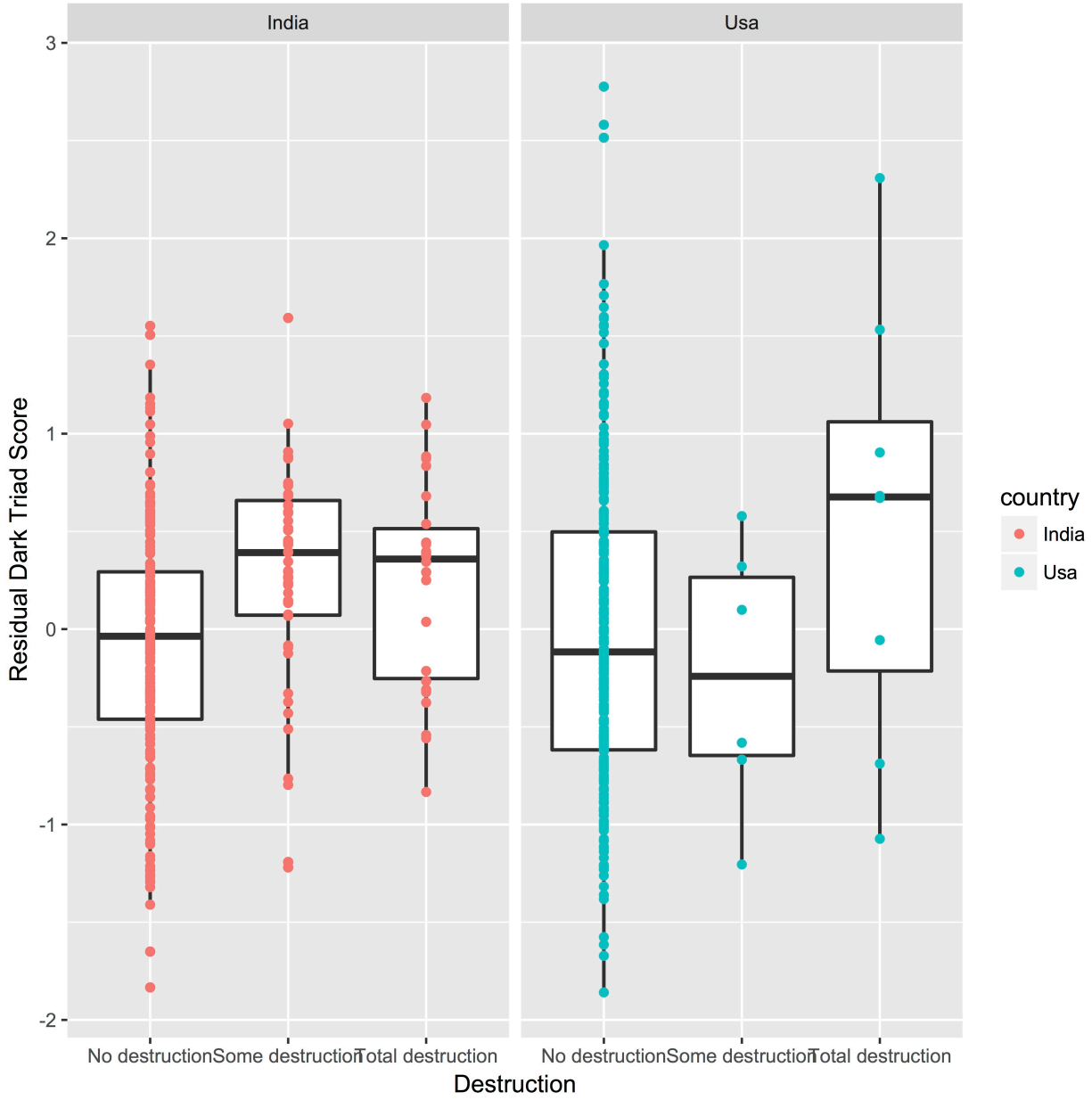

Figure 2. Boxplots showing destruction decisions (no destruction, partial destruction and total destruction) as a function of residual Dark Triad score and country of origin. Residual Dark Triad score described the remaining variance in Dark Triad score, after the variance explained by country is removed. Data from India-based participants are shown on the left panel; data from USA-based participants are shown on the right panel.

Discussion

We showed that Dark Triad personality traits predict antisocial behavior in a joy-of-destruction game. In addition, individuals from India scored higher on Dark Triad traits than individuals from the USA – and were more likely to destroy some or all of their partner's bonus. Dark Triad scores also had an additive positive effect on the tendency to destroy the partner's bonus. This supports previous research showing that Dark Triad traits are related to inter-personal aggression (Jonason et al., 2015) and other work showing that India-based players are more likely to invest in antisocial punishment than US-based counterparts (Bone et al., 2016).

Contrary to our predictions – and to previous findings (Abbink & Sadrieh, 2009) – participants were equally likely to destroy the partner's bonus in the conditions when destruction was open and when it was hidden by chance destruction. This may be because, unlike Abbink & Sadrieh (2009), the joy-of-destruction game used in this study was one-shot and players therefore did not fear retribution from the partner when destroying in the open condition. Another explanation could be that since our participants interacted anonymously via the internet, the hidden and open conditions were largely indistinguishable in terms of potential future consequences. In future, it

would be instructive to conduct an iterated joy-of-destruction game on MTurk to try to disambiguate these two possibilities. We had also predicted that participants would be less likely to destroy the partner's bonus when doing so was costly compared to when it was free but this prediction was unsupported by our data. We found no meaningful effect of the cost condition on participants' destruction decisions. This could be because the cost of destruction was relatively small ($0.10) and, crucially, was always smaller than the impact levied on the target. If destroying the partner's bonus is motivated (even on a subconscious level) by the desire to elevate relative payoffs (Falk et al., 2005; Sylwester et al., 2013; Bryson et al., 2014), then the cost is expected to be relatively unimportant so long as the ratio of cost to impact is < 1. In future, it would be interesting to change the cost of destroying the partner's bonus, such that the cost to impact ratio is ≥ 1 to test this hypothesis. With a cost to impact ratio that is ≥ 1, we would not expect the Dark Triad traits to have any meaningful association with proclivity for antisocial punishment.

Some limitations of this study must also be considered. Due to an oversight, we did not collect basic demographic information (e.g. age, gender) about our subjects, so we were unable to determine to what extent the personality scores and associated spiteful behavior were explained by these demographic variables rather than by country of origin. We should also exercise caution in interpreting the scores of the personality scale itself. We used an unmodified version of the Jones & Paulhus (2014) SD3 questionnaire – but this questionnaire unfortunately includes only 5 reversed statements out of the total 27. That means that if there is systematic variation in tendency to agree with statements across countries (c.f. Smith 2004), then the increased Dark Triad scores for Indian players could be a spurious effect of differential acquiescence rather than

reflecting actual personality differences. Our analysis of the scores for the reversed statements somewhat mitigates against this possibility – as Indian players were less likely than US players to acquiesce with the reversed statements. Moreover, it is not the case that country-level differences in Dark Triad scores can be explained by increased tendency for extreme responses among Indian players – in fact we found the opposite pattern in our data, with US-based players showing more extreme responses. Nevertheless, we cannot be certain that cultural differences in responding styles or in the interpretation of the statements themselves did not account for at least some of the variation in the responses.

To sum up, we believe this study provides an important first step in showing how the existence of broad-scale variation in Dark Triad personality traits can explain similarly large-scale patterns in social behavior. Previous work has already outlined how selection can favor spite in resource-scarce environments (Gardner & West, 2004; Lehmann et al., 2009), with empirical data supporting these theoretical predictions (Herrmann et al., 2008; Prediger et al., 2014; Bone et al., 2016). Our data contribute to this picture, by suggesting that cross-cultural variation in antisocial behavior might be associated with cross-cultural variation in personality traits, such as the Dark Triad. A priority for future work will be to challenge these predictions with data from other countries – something that is currently not feasible on MTurk given the predominance of USA and India-based workers.

Acknowledgments

This research was funded by a Royal Society University Research Fellowship to NR.

References

Abbink K., Sadrieh A. (2009) The pleasure of being nasty. Economics Lettters, 105(3), 306–308.

Abbink, K., Herrmann, B. (2011) The moral costs of nastiness. Economic Inquiry 49, 631-633.

Anderson CM., Putterman L. (2006) Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. Games and Economic Behavior, 54(1), 1–24.

Bartoń K. (2009) MuMIn: multi-model inference. R package, version 0.12.2. Available at: http://r-forge.r-project.org/projects/mumin/.

Bone JE., McAuliffe K., Raihani, NJ. (2016) Exploring the motivations for punishment: framing and country-level effects. PLoS ONE, 11(8), e0159769.

Briones O., Montaña C., Ezcurra E. (1998) Competition intensity as a function of resource availability in a semiarid ecosystem. Oecologia 116: 365–372

Bryson JJ., Mitchell J., Powers ST., Sylwester K. (2014) Understanding and addressing cultural variation in antisocial punishment. In: Applied Evolutionary Anthropology (ed M.A. Gibson & D.W. Lawson) Springer Science and Business Media New York.

Bshary R., Grutter AS. (2002) Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. Animal Behaviour, 63, 547–555.

Burnham K., Anderson D. (2002) Information and likelihood theory: a basis for model selection and inference. Model Selection and Multimodel Inference: A Practical Information and Theoretic Approach. New York: Springer.

Clutton-Brock TH., Parker GA. (1995) Punishment in animal societies. Nature, 373, 209-216.

Dingemanse NJ., Wolf M. (2010) Recent models for adaptive personality differences: a review. Philosophical Transactions of the Royal Society of London B Series, 365: 3947–3958

Falk A., Fehr E., Fischbacher, U. (2005) Driving forces behind informal sanctions. Econometrica 73, 2017-2030.

Gardner A., West SA. (2004) Spite and the scale of competition. Journal of Evolutionary Biology, 17, 1195–1203

Gelman A. (2008) Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine, 27, 2865-2873.

Grossman HI., Mendoza J. (2003) Scarcity and appropriative competition. European Journal of Political Economy, 19(4), 747–758.

Grueber CE., Nakagawa S., Laws RJ., Jamieson IG. (2011) Multimodel inference in ecology and evolution: challenges and solutions. Journal of Evolutionary Biology, 24, 699–711.

Herrmann B., Thöni C., Gachter S. (2008) Antisocial punishment across societies. Science, 319(5868), 1362–1367.

Hofstede G. (2001) Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Second Edition, Thousand Oaks CA: Sage Publications.

Hurvich CM., Tsai CL. (1989) Regression and time series model selection in small samples. Biometrika, 76(2), 297–307.

Jonason PK., Li NP., Webster GW, Schmitt DP. (2009) The Dark Triad: facilitating short-term mating in men. European Journal of Personality, 23, 5–18.

Jonason PK., Tost J. (2010) I just cannot control myself: the Dark Triad and self-control. Personality and Individual Differences, 49, 611-615.

Jonason PK., Koenig BL., Tost J. (2010) Living a fast life: the Dark Triad and life history theory. Human Nature, 21, 428–442.

Jonason PK., Krause L. (2013) The emotional deficits associated with the Dark Triad traits: cognitive empathy, affective empathy, and alexithymia. Personality and Individual Differences, 55, 532–537.

Jonason PK., Strosser GL., Kroll CH., Duineveld JJ., Baruffi SA. (2015) Valuing myself over others: the Dark Triad traits and moral and social values. Personality and Individual Differences, 81, 102–106.

Jonason PK., Duineveld JJ., Middleton JP. (2015) Pathology, pseudopathology, and the Dark Triad of personality. Personality and Individual Differences, 78, 43–47.

Jonason PK., Icho A., Ireland K. (2016) Resources, harshness, and unpredictability: the socioeconomic conditions associated with the Dark Triad traits. Evolutionary Psychology Jan-March, 1-11.

Jones DN., Paulhus DL. (2014) Introducing the Short Dark Triad (SD3): a brief measure of dark personality traits. Assessment, 21, 27–40.

Lehmann L., Feldman MW., Rousset F. (2009) On the evolution of harming and recognition in finite panmictic and infinite structured populations. Evolution, 63, 2896-2913.

Martin, D., Hanrahan, B.V., O'Neill, J. & gupta, N. (2014) Being a Turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 224-235). ACM.

McCullough ME., Kurzban R., Tabak BA. (2013) Cognitive systems for revenge and forgiveness. Behavioral and Brain Sciences, 36, 1–15

Paulhus DL., Williams K. (2002) The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. Journal of Research in Personality, 36, 556-568.

Prediger S., Vollan B., Herrmann B. (2014) Resource scarcity and antisocial behavior. Journal of Public Economics, 119, 1–9.

Raihani NJ., Grutter AS., Bshary R. (2010) Punishers benefit from third-party punishment in fish. Science 327, 171.

Raihani NJ., Thornton A., Bshary R. (2012) Punishment and cooperation in nature. Trends in Ecology and Evolution, 27, 288–295.

Raihani NJ., Mace R., Lamba S. (2013) The effect of $1, $5 and $10 stakes in an online dictator game. PLoS ONE, 8(8), e73131

Raihani NJ., Deutchman, P. (2016) Joy of Destruction Dark Triad data. Figshare. https://dx.doi.org/10.6084/m9.figshare.4288532.v1

Smith, P. B. (2004) Acquiescent response bias as an aspect of cultural communication style. Journal of Cross-Cultural Psychology, *35*(1), 50–61.

Sylwester K., Herrmann B., Bryson JJ. (2013) Homo Homini Lupus? Explaining antisocial punishment. Journal of Neuroscience and Psychological Economics, 6, 167–188.

Rand DG. (2012) The promise of Mechanical Turk: how online labor markets can help theorists run behavioral experiments. Journal of Theoretical Biology, 299, 172–179.

Yamagishi T. (1986) The provision of a sanctioning system as a public good. Journal of Personality and Social Psychology, 57(1), 110-16.