Evolution and Human Behavior xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

When repentance and forgiveness are possible, direct reciprocity does not explain one-shot generosity, even under arbitrarily high levels of uncertainty

You hail a cab at the airport and a person in khaki asks you to share, but you decline. At your job interview the next day, it turns out that the person in khaki is your interviewer! This scenario is problematic, to be sure, and in Delton and Krasnow's (2014) (hereafter "commentators") commentary you lose the job. However, suppose instead that you apologize and the interviewer accepts your apology, reviews your qualifications, finds you are the best person for the job and hires you. Once in a long-term employer–employee relationship, you perform admirably and bring many accolades to your department.

Both you and your interviewer are better off than in the commentators' scenario. You are better off having not shared the cab because, not only did you still get the job, you also arrived at your hotel sooner and were more rested than you would have been otherwise. Over many similar situations (most where the potential cab-sharer is not also your future interviewer) you also reap the rewards of not being overly cooperative in one-time interactions. Importantly, if you are repentant, your interviewer is better off accepting your apology. After all, he hired the candidate who brought the most adulation to his department, and he only had to forgive the minor sunk cost of waiting for another cab.

These scenarios illustrate the key difference between Delton, Krasnow, Cosmides, and Tooby's (2011a) model (hereafter "the DKCT model") and my expansion of it. They posit a world where repentance and forgiveness are impossible. A defecting agent, finding itself in a repeated interaction, can never again cooperate with its partner. My paper shows that, when forgiveness and repentance *are* possible, the selective advantage of one-shot cooperation in the DKCT model disappears. Another explanation is needed to explain one-shot cooperation.

The commentators, despite dramatic differences in our findings, argue that the DKCT model actually predicts my results. For evidence, they point to a few parameter combinations where one-shot cooperation fails to evolve in either of our simulations. However, since their main findings are the conditions where generosity in one-shot encounters *does* evolve, these conditions are more appropriate comparisons. In all of these conditions, one-shot generosity disappears when repentant and forgiving strategies are added (Zefferman, 2014, Figure 4 and Supplemental Appendix A).

DKCT not only did not predict my results, they explicitly predicted the opposite of my results. In a response to a commentary where McNally and Tanner (2011) argued, as I do, that DKCT's conclusions would not hold if their model included forgiving and hesitant (a sub-set of repentant) strategies, Delton, Krasnow, Cosmides, and Tooby (2011b) wrote:

> "McNally and Tanner voice the reasonable concern that our results may not generalize widely because we did not model hesitant or forgiving strategies… We concur that the fitness differential between initial cooperation and defection might be smaller in such a model. *Importantly, however, the direction of selection would be the same, just with reduced strength. Hence, the effects would not be eliminated and the generality of our results would be unchanged*" [emphasis mine].

My model demonstrates that this explicit prediction is incorrect. Instead of "complete harmony" between our findings, I show that modeling hesitant and forgiving strategies does not just make the fitness differential smaller, but changes the direction of selection and eliminates one-shot cooperation from any population where it evolves in DKCT's simulations. I also show, analytically, that this result holds when uncertainty is maximized. It thus meets Delton et al.'s (2011b) own criteria for challenging the generality of their results.

The commentators also write that my simulations employ strategies that are "complex and carefully chosen." This would be problematic if selection were not clever enough to discover repentance or forgiveness without my help. However, Axelrod (1997) showed that these behaviors quite readily evolve and out-perform Tit-for-Tat when given the opportunity. I did choose strategies, but only in the sense that, for illustration, I identified *simple* representatives of the infinitely large class of the type of repentant and forgiving strategies that evolved endogenously in Axelrod's simulations. Using standard complexity measures (Rubinstein, 1986), I modeled no strategy more complex than the commonly-modeled Tit-for-Two-Tats (Zefferman, 2014, Supplemental Appendix D).

Finally, the commentators, referencing Fehr and Henrich (2003) and Gintis, Bowles, Boyd, and Fehr (2003), also suggest that my view of reciprocity challenges the consensus. This is puzzling since both papers, far from "assuming reciprocity existed," call on decades of theoretical and empirical research to conclude, as I do, that direct reciprocity in ancient environments does not sufficiently explain one-shot cooperation in modern laboratory experiments. They also argue, as do I, that it evolves through a process of gene-culture coevolution and equilibrium selection on variation in cultural norms between groups. My paper expands this consensus, showing that direct reciprocity still does not explain one-shot generosity, even under arbitrarily high levels of uncertainty.

Despite my critique of the DKCT model, I consider it a positive development. I agree with the commentators that disagreement over the evolution of human cooperation have for too long been mired in a semantic debate creating more confusion than clarity and that DKCT should be commended for translating a previously vague argument into an explicit model. However, I do not see such formalization as the *end* the conversation, but only the beginning. DKCT have done the science of human cooperation a great service by starting this conversation and I thank *Evolution and Human Behavior* for facilitating its continuation here.

1090-5138/© 2014 Elsevier Inc. All rights reserved.

Please cite this article as: Zefferman, M.R., When repentance and forgiveness are possible, direct reciprocity does not explain one-shot generosity, even under arbitrarily..., *Evolution and Human Behavior* (2014), http://dx.doi.org/10.1016/j.evolhumbehav.2014.07.003

ARTICLE IN PRESS

Acknowledgments

This work was performed at NIMBioS, sponsored by the NSF, U.S. Department of Homeland Security, and U.S. Department of Agriculture through NSF Awards EF-0832858 and DBI-1300426, with additional support from the University of Tennessee. I thank Andrew Delton and Max Krasnow for their commentary and Sergey Gavrilets and Emily Zefferman for their comments on earlier drafts of this response.

Matthew R. Zefferman
National Institute for Mathematical and Biological Synthesis, University of Tennessee, USA
Corresponding author.
1122 Volunteer Blvd., Suite 106
University of Tennessee Knoxville
TN 37996–3410.
E-mail address: mrz1@nimbios.org

Available online xxxx

References

Axelrod, R. A. (1997). *The complexity of cooperation.* Princeton, NJ: Princeton University Press.

Delton, A. W., & Krasnow, M. M. (2014). *An independent replication that the evolution of direct reciprocity under uncertainty explains one-shot cooperation: Commentary on Zefferman.*

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011a). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, *108*, 13335–13340.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011b). Reply to McNally and Tanner: Generosity evolves when cooperative decisions must be made under uncertainty. *Proceedings of the National Academy of Sciences*, *108*, E972.

Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 55–82). Cambridge, MA, US: MIT Press.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, *24*(3), 153–172.

McNally, L., & Tanner, C. J. (2011). Flexible strategies, forgiveness, and the evolution of generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, *108*, E971.

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Economic Theory*, *39*, 83–96.

Zefferman, M. R. (2014). Direct reciprocity under uncertainty does not explain one-shot cooperation, but demonstrates the benefits of a norm psychology. *Evolution and Human Behavior*, *35*(5), 358–367.

Please cite this article as: Zefferman, M.R., When repentance and forgiveness are possible, direct reciprocity does not explain one-shot generosity, even under arbitrarily..., *Evolution and Human Behavior* (2014), http://dx.doi.org/10.1016/j.evolhumbehav.2014.07.003