Application for 2010 Russell Ackoff Doctoral Student Fellowships for Research on Human Decision Processes and Risk Management

Applicant: Ewa Szymanska
               Psychology Department
               University of Pennsylvania
               3720 Walnut Street, Solomon Lab Bldg.
               Philadelphia, PA 19104
               Email: ewas@psych.upenn.edu

Project Title: Moral impartiality and neglected loyalties in third-party punishment

Faculty Members: Jon Baron and Robert Kurzban

Project Title: Moral impartiality and neglected loyalties in third-party punishment

Since a decision to punish clearly belongs to the family of social behaviors, we would expect social cognition, or a bundle of mechanisms that enable us to navigate our social world (see Fiske & Taylor, 1992), to heavily contribute to the punishment decisions. One pronounced feature of social cognition is human tendency to parcel the social world into categories, such as the generic *us* versus *them*, or more specific, e.g., *family* versus *unrelated coalition partners* versus *strangers*. Hence, past research took considerable interest in the influence of social identities – such as family and friends (Lieberman & Linke, 2007; O'Gorman et al., 2005), coworkers (Carpenter et al., 2004), or the anonymous ingroup versus outgroup members (Goette et al., 2006; Bernhard et al., 2006; Lieberman & Linke, 2007; Gil-White, 2004) – on punitive judgment. Earlier findings from social psychology on ingroup bias and outgroup hostility (Bernhard et al., 2006), demonstrating that subjects favor their ingroup members when allocating the resources in the minimal group experiments (see Brewer, 1979), were taken as a suggestion that people would want to punish out-groups more severely than in-groups. Subsequent research on punishment, however, produced mixed results (Gil-White, 2004; Carpenter et al., 2004).

The inconclusive findings on the effects of social category on punitive judgment might be due in part to flexibility in people's perceptions of group boundaries (Brewer, 1979) and shifting patterns of coalitional alliances (Kurzban, Tooby, & Cosmides, 2001). However, I argue that our understanding of the third-party punishment would benefit most from exploring **moral impartiality** – the defining, yet largely neglected by research, feature of moral justice. Ironically, moral impartiality implies complete blindness to social categories. In other words, a prototypical third-party punisher would apply the objective standards of moral judgment without regard to family ties, friendships, or any other type of allegiance. In that sense, moral impartiality is a direct opposite of favoritism or group-based loyalty. Carpenter et al. (2004) incorporated the concept of moral impartiality in their theory of social reciprocity "in which people punish norm violators indiscriminately," that is they argued that "people subscribe to the simpler heuristic of punishing all norm violators, regardless of group affiliation." However, with few exceptions, research has not systematically explored the distinction between moral impartiality and group favoritism, and most models of moral judgment cannot predict when impartiality will trump loyalty.

From either economic or biological standpoint, it is not clear why an individual would jeopardize personal relationships in order to adhere to an abstract moral rule (DeScioli & Kurzban, 2009). So it is not at all surprising to find that people *do* often departure from a standard of moral impartiality. However, common deviations do not mean that moral impartiality is merely a theoretical concept in normative moral and political models. To the contrary, it seems that to some extend, people across all cultures "advocate neglecting loyalties when making moral judgments" (DeScioli & Kurzban, 2009). Given the incompatibility of the two behavioral rules of moral impartiality and the in-group favoritism (or the family- and friendship-based loyalties), DeScioli and Kurzban (2009) suggest that some important questions to ask about punitive judgment are "why people are *ever* impartial to any degree at all, why they often claim to be impartial, why they praise impartiality, and why humans possess the concept of moral impartiality."

In proposed research, I will search for the instances of moral impartiality, with an ultimate goal of identifying the specifics of the situation that allow **the rule of moral impartiality to trump the rule of group-based loyalty**. Specifically, I will take a closer look at people's decisions to use rules of either impartiality or loyalty to one's friend or relative in the context of escalating disputes between strangers and one's allies. The studies I propose will focus on the group membership of both the victim and the perpetrator and the effect of the presence (or absence) of different types of audiences on punitive judgment. I will begin by constructing *hypothetical scenarios* featuring a conflict between a friend (or a family member) and a stranger, and accompany the descriptions of conflict situations with pictorial representations of different audiences, such as "mutual friends of you and your friend," "friends of a stranger," or "incidental observers," (or alternatively, have the participants put the names of their friends on blank faces). In addition, I will manipulate factors such as severity and type of the moral violation, and degree of situational ambiguity. I will also design scenarios featuring disputes between co-nationals and foreigners, as well as include the cases of military intervention. I predict to observe a lot of loyalty-guided responses across studies and conditions. However, I also expect to identify a unique combination of factors that prompt people to trade friendship vows for a reputation of an impartial punishers. Since impartiality is a widely praised feature of moral justice, understanding the drivers and stoppers of moral impartiality will shed new light on the studies of moral cognition in general.

References

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review, 96,* 217-221.

Brewer, M. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86,* 307-324.

Carpenter P.J., Matthews, P.H., & Ong'ong'a, O. (2004). Why Punish? Social reciprocity and the enforcement of prosocial norms. *Journal of Evolutionary Economics, 14*, 407-429.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112,* 281-299.

Fiske, S.T., & Taylor, S.E. (1992). Social Cognition. McGraw-Hill

Gil-White, F. J. (2004). Ultimatum Game with ethnicity manipulation: Problems faced doing field economic experiments, and their solutions. *Field Methods.*

Goette, L., Huffman, D., Stephan, M. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Institute for the Study of Labor, IZA Discussion paper,* 2020.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98,* 15387-15392.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology, 5*, 289-305.

O'Gorman, R., Wilson, D. S., & Miller, R. B. (2005). Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions. *Evolution and Human Behavior, 26,* 375-387.