Abstract

Punishment of free-riders is a key to understanding large scale human cooperation, and thus the issue of evolution of punitive behavior has been a focus of recent studies of human cooperation. One of the topics for discussion in this literature is whether punishers receive personal benefits that surpass the cost for their punitive behavior, on which no clear conclusion has been reached yet. We suspected that the lack of consistent findings on this issue reflects diversity of game situations and reward types used in the past research. We thus conducted a scenario experiment in which we manipulated game types and reward types, and found that punishers were trusted more than non-punishers, and yet, they were chosen less frequently than non-punishers as recipients of rewards. Adaptive advantages of punishers are suggested to be in their likelihood of being chosen as coalition partners, rather than in the reward they receive from beneficiaries of their punitive behavior.

Keywords: punishment, cooperation, reputation.