Costly reputation maintenance behavior is a building block of human cooperation

Reputation plays a crucial role in human cooperation. This is because people tend to cooperate with someone with a good reputation. For example, if Anne knows that Bob helped Cathy before, and Bob is now in need, Anne may be inclined to help Bob. Such a tendency allows cooperation to spread in a society: Anne helps Bob who helped Cathy who helped Dan... This is called indirect reciprocity.

Once the indirect reciprocity system has been established in a society, everyone has an incentive to help others: it gives you a good reputation and guarantees that others will help you. By contrast, if you fail to help someone in need, you may lose your good reputation. At first glance, these incentives appear to reinforce indirect reciprocity. Nevertheless, in reality, they do more bad than good. Why? Because they make “help anyone irrespective of their reputation” the best strategy—it allows you to maintain a good reputation. Even though you may find it unreasonable, if you withhold help from a bad person, you will lose your good reputation—at least temporarily.

This problem can be avoided if people can distinguish justifiable uncooperative behavior (i.e., not helping a bad person) from unjustifiable uncooperative behavior (i.e., not helping a good person). However, this requires knowing detailed information about others’ past behavior. For Anne to decide whether to help Bob, she needs to know not only whether Bob helped Cathy but also whether Cathy helped Dan. According to game theoretic analyses, the availability of such higher-order information stabilizes indirect reciprocity. However, it is controversial as to whether people really take second-order information into account—whether Anne really takes into account Cathy’s reputation in deciding whether to help Bob. In experimental studies on indirect reciprocity, participants (playing Anne) were given first-order information (Bob’s behavior toward Cathy) along with second-order information (Cathy’s behavior toward Dan). In some experiments, participants did use second-order information, while in others, they relied only on first-order information.

In the above example, if Bob had decided to not help Cathy, all he could do is await Anne’s judgment. However, in reality, doesn’t Bob have some say regarding his reputation? Think about the following scenario. Bob did not spend $5 to help Cathy. From Anne’s perspective, Bob’s motivation is not clear. It may be justifiable (e.g., because Cathy did not help Dan) or unjustifiable (e.g., Bob is stingy). However, what if Bob donated the unspent $5 to some charitable organization? It would clearly show that he is not stingy. Alternatively, although less realistic, Bob could burn his $5 bill. Either way, Bob’s abandoning $5 serves as a signal of his justifiable intention.

Our game theoretic analysis showed that this intention signaling strategy is effective in stabilizing indirect reciprocity. Notice that the intention signaling strategy does not require second-order information: Anne only needs to know how Bob behaved. In our laboratory experiment, we gave

our participants (playing Bob) an additional option—to abandon an unspent resource. Although we did not provide any functional explanations about the option (e.g., that it serves as a signal of benign intention), our participants used this option more frequently after justified defection (i.e., not giving their resource to a previously uncooperative partner) than after unjustified defection (i.e., not giving their resource to a previously cooperative partner). Moreover, participants (playing Anne) tended to give the resource to signalers but not non-signalers, who kept their resource and did not use the signal option. Taken together, costly reputation maintenance behavior—a signal of benign intention—seems to play a pivotal role in the stabilization of reputation-based cooperation in human societies.

Yohsuke Ohtsubo
Kobe University, Graduate School of Humanities, Department of Psychology, Kobe, Japan

Publication

The price of being seen to be just: an intention signalling strategy for indirect reciprocity.
Tanaka H, Ohtsuki H, Ohtsubo Y
Proc Biol Sci. 2016 Jul 27