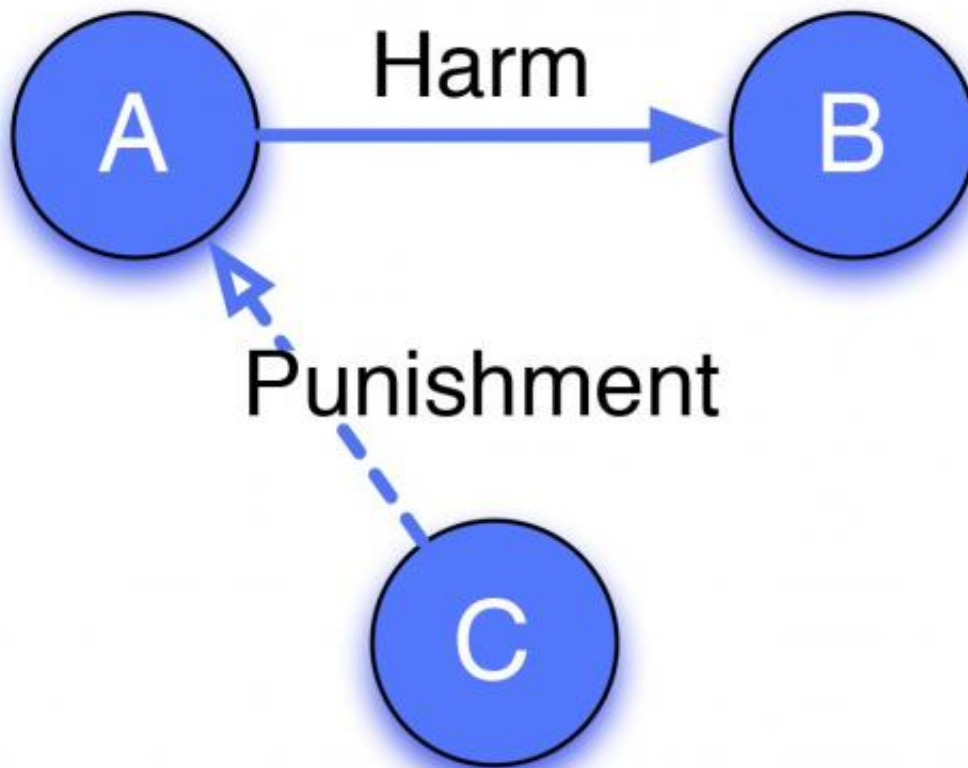# Game theory used to explain evolution of 'third party punishment'

December 10 2013

A harms B, but is punished by C, an uninvolved third party. Credit: Patrick Roos, Michele Gelfand, Dana Nau and Ryan Carr

You're shopping for holiday gifts when you spot someone pocketing a nice pair of leather gloves. What do you do?

A new study by University of Maryland researchers appearing this week in the *Proceedings of the Royal Society B* predicts that whether you alert a manager to the theft or decide to do nothing may depend on whether you're shopping in a local store where you know the owners or in a city far from home.

The study, "High strength-of-ties and low mobility enable the evolution of third-party punishment," suggests that the stronger a community's social ties and the longer most people stay within the community, the more likely it is that otherwise uninvolved third parties will step forward to punish their neighbors.

Psychology Professor Michele Gelfand, an expert in cross-cultural social organizational psychology, teamed with Computer Science and Institute for Systems Research Professor Dana Nau and two of his former Ph.D. students, Ryan Carr and Postdoctoral Researcher Patrick Roos, on the interdisciplinary project. The team began with sociological and psychological hypotheses of how behavior might evolve, then tested them with evolutionary game theoretic computer models.

Gelfand studies how culture influences conflict, negotiation, justice and revenge. In earlier research she has looked at how cultural norms, such as the concept of honor, can cause conflicts among individuals to spread to wider social groups. But some cultures have evolved a way to short-circuit that process: third-party punishment.

Unlike police and courts that mete out official punishments, third-party punishment is informal, based on an individual's decision to right a perceived wrong. In some cultures, third-party punishment, when used responsibly, is a useful tool to enforce social norms. Why does it evolve in some places but not others?

Gelfand noted a recent experiment by other researchers, in which U.S.

college students consistently balked at punishing their peers for perceived wrongdoings to others. She wondered whether the result was due to the highly mobile, individualistic nature of U.S. society and the loose social ties of a college campus. Her hypothesis was that in a more traditional culture with strong social ties (where people interact frequently) and low mobility (where people can't easily leave the social group) things might be different. To test her idea, Gelfand turned to Nau, Roos and Carr, computer scientists skilled at evolutionary game theory, a powerful predictive mathematical tool.

"Evolutionary game theory was developed to model the emergence of biological features," says Nau. "But more recently it has been effectively used in sociological and psychological research."

"With evolutionary game theoretic models we can study what types of behaviors are likely to become the most widespread under different conditions," says Roos. "We are trying to understand mathematically how social systems work, and we can explain behavior using these models."

The computer scientists built a mathematical model that incorporated Gelfand's hypotheses. The model results suggest that third-party punishment is much more likely to evolve in contexts of high social and structural constraint because in the long term it benefits the whole community, including the individual who metes out punishment. A lone responsible punisher—that is, someone who steps up to enforce social norms fairly—cannot induce cooperation and actually suffers compared with his or her neighbors. However, if this punisher is joined by another punisher in the neighborhood, together they can induce cooperation and gain a social benefit.

The results suggest when responsible third-party punishment evolves, it does so because the responsible punishers' actions are ultimately not

altruistic. The behavior acts as a signal to others in the neighborhood that non-cooperation will not be tolerated.

The results also show that responsible third-party punishment does not evolve in populations with weak social ties or high mobility. A critical mass of responsible punishers is hard to achieve when there are weak social ties because they cannot give each other enough support. Similarly, in highly mobile societies, fellow responsible punishers move away or non-cooperative agents replace them. In both of these situations it becomes too costly for individuals to take on the burden and risk of being responsible punishers.

Game theory has been used before in social science work, but this is the first time that cross-cultural psychologists and computational game theorists have collaborated to examine the evolution of third-party punishment, says Gelfand. "It's a good example of how psychologists and computer scientists can team up to do something that neither one could have done before."

"There are many other things that can affect this evolution in a community," Nau notes, "for example, the type of government, outside conflicts, and the amount of resources available. Because we now have a collection of intuitions about how third-party punishment works, we believe further research could contribute to a better understanding of why conflicts escalate, or why some societies become steeped in revenge."

How the game theory model works

At each generation, agents interact in a game phase followed by a punishment phase. In the game phase, agents are randomly paired to interact in a classic two-player cooperation game, where they can choose to either cooperate (paying a cost to bestow a benefit of cooperation

upon the other agent) or defect (not paying the cooperation cost, while receiving any potential benefit from the other agent's action).

In the punishment phase, agents get a chance to punish other players, on behalf of themselves or others, at a cost to themselves. They may punish responsibly (only punishing defectors), antisocially (only punishing cooperators), spitefully (punish indiscriminately) or not punish at all.

In the cooperation phase, players may choose to always cooperate, always defect, or be opportunistic, cooperating or defecting when it seems to benefit them the most. In the game, opportunistic agents take the punishment reputation of their neighbors (which they become aware of) into account when deciding whether to cooperate or defect.

More information: Patrick Roos, Michele Gelfand, Dana Nau and Ryan Carr, "High strength-of-ties and low mobility enable the evolution of third party punishment," in *Proceedings of the Royal Society B*, Dec. 11, 2013. Available on publication at rspb.royalsocietypublishing.or … .1098/rspb.2013.2661

Provided by University of Maryland