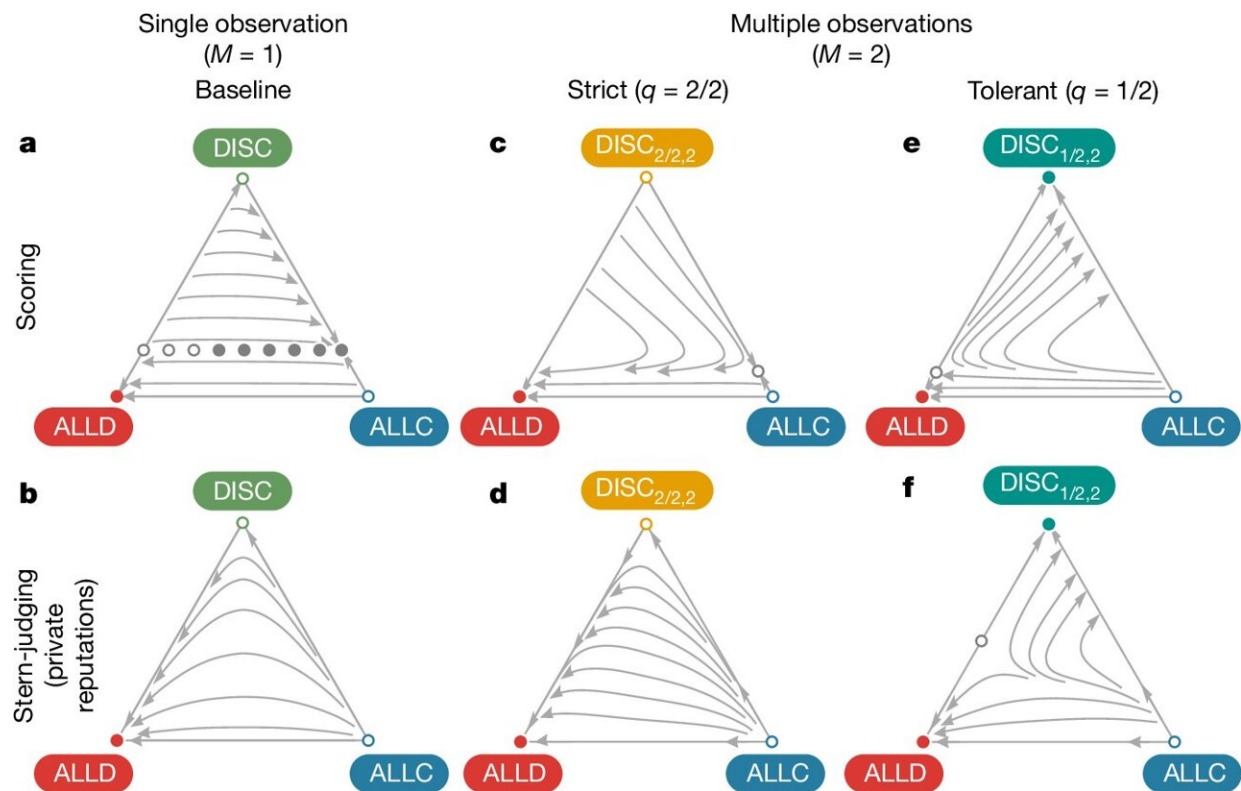# A method of 'look twice, forgive once' can sustain social cooperation

September 26 2024, by Erica Moser

'Look twice, forgive once' solves both the scoring and the punishment dilemmas. Credit: *Nature* (2024). DOI: 10.1038/s41586-024-07977-x

The theory of indirect reciprocity holds that people who earn a good reputation by helping others are more likely to be rewarded by third parties, but widespread cooperation depends on agreement about

reputations.

In most [theoretical models](#) examining how reputations impact people's desire to cooperate with one another, reputations are binary—good or bad—and based on [limited information](#). But there is a lot of information available about people's behavior in today's world, especially with social media.

Biology professors Joshua B. Plotkin of the University of Pennsylvania and Corina Tarnita of Princeton University lead teams that have been collaborating on [theoretical research](#) about [cooperation](#). Sebastián Michel-Mata, a doctoral student in Tarnita's lab, came up with the idea of addressing how to judge someone in an information-rich environment.

"The current theory of indirect reciprocity suggests that reputations can only work in a few societies, those with complex norms of judgment and public institutions that can enforce agreement," Michel-Mata says. But, as an anthropologist, he sees that such societies are the exception and not the rule, and he wondered about the simple idea that reputations are summaries of multiple actions.

"Prior models have typically assumed that a single action determines someone's reputation, but I think there's more nuance to how we assign reputations to people. We often look at multiple actions someone has taken and see if they are mostly good actions or bad actions," says Mari Kawakatsu, a postdoctoral researcher in Plotkin's lab.

Through mathematical modeling, the research team showed that looking at multiple actions and forgiving some bad actions is a method of judging behavior that is sufficient to sustain cooperation, a method they call "look twice, forgive once." Their research is [published](#) in *Nature*.

This builds on previous work Plotkin led about indirect reciprocity. For example, he worked with Kawakatsu and postdoctoral researcher Taylor A. Kessinger on a paper calculating how much gossip is necessary to reach sufficient consensus to sustain cooperation.

Plotkin says of the new paper, "Even if different people in a society subscribe to different norms of judgment, 'look twice, forgive once' still generates sufficient consensus to promote cooperation."

He adds that this method maintains cooperation without gossip or public institutions, which confirms the original hypothesis that Michel-Mata, first author on the paper, had that public institutions are not a prerequisite for reputation-based cooperation. It also offers an important alternative when public institutions exist but erosion of trust in institutions inhibits cooperation.

Kessinger says that, as in the paper about gossip, the game-theoretical model here is a one-shot donation game, also known as a simplified prisoner's dilemma. Each player can choose to help or not help their partner, and players will periodically update their views of each other's reputations by observing each other's interactions with other players, to see if the partner cooperates or "defects" with others. More periodically, players update their strategies.

The idea of indirect reciprocity is "not that I'm nice to Mari because she was nice to me; it's that I'm nice to Mari because she was nice to Josh, and I have a good opinion of Josh," Kessinger says.

In this study, "the basic idea is that if you observed two interactions of somebody and at least one of them was an action that you would consider good, then you cooperate with that player, but otherwise you defect with them."

Kawakatsu says all co-authors were surprised that the "look twice, forgive once" strategy couldn't be displaced by other strategies, such as always cooperating or always defecting, looking at more than two actions from another player, or forgiving a different proportion of "bad actions."

Tarnita says that, perhaps most surprisingly, looking more than twice didn't yield an additional benefit. "Information turned out to be a double-edged sword, so that even when information was freely accessible, individuals did not typically evolve to use all of it," she says.

Michel-Mata notes that the overall simplicity and robustness of their findings indicate that this behavioral strategy might be old in human societies. The authors see potential for anthropologists and behavioral scientists to build on their work.

The Plotkin and Tarnita labs are continuing to collaborate by exploring how people interact in more than one context, such as at work and in their personal lives. "This touches on a range of contemporary social problems," Kessinger says, "where private misbehavior becomes a matter of public record."

Provided by University of Pennsylvania