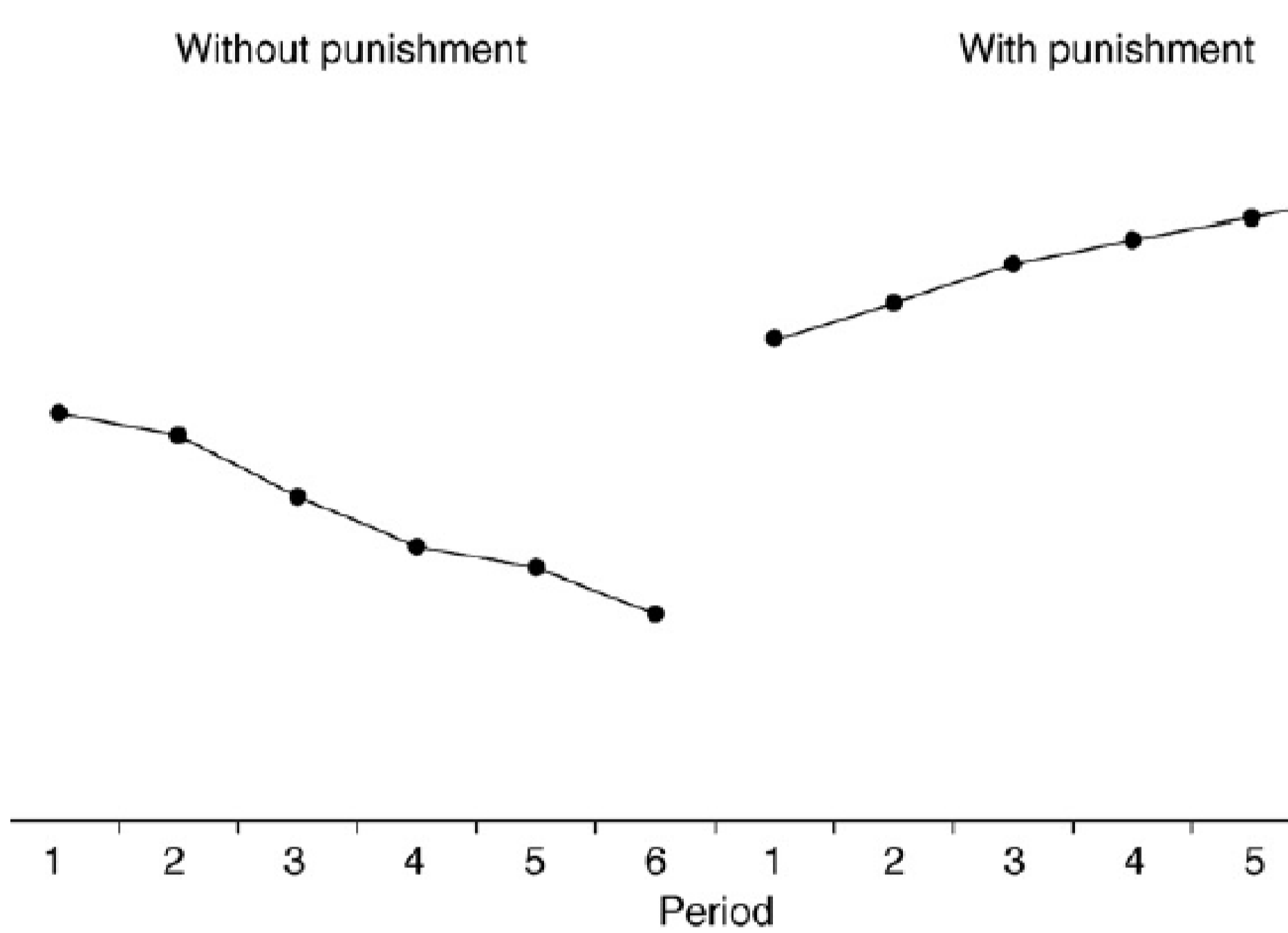


Sanctioning free riders is a public good

Empirical and theoretical research on mutual-aid games indicates that the threat of punishment can curb free riding in human groups engaged in joint enterprises. Since punishment is often costly, however, this raises an issue of second-order free riding: indeed, the sanctioning system itself is a public good that can be exploited.



Public good games with punishment

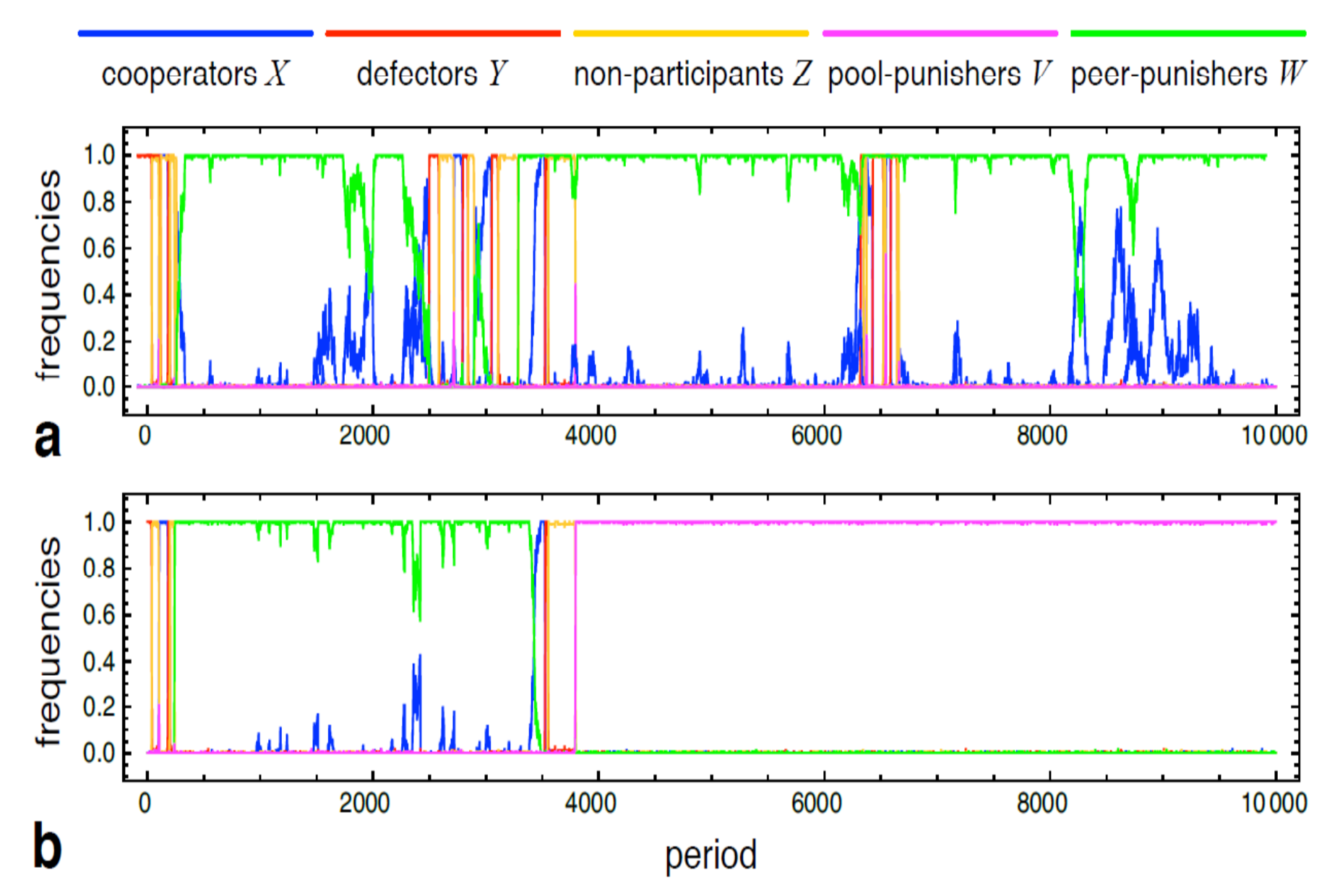
In a typical public good experiment, four players can decide to contribute between 0 and 20 euros to a common pool, knowing that the experimenter will double its content and then divide it equally between all four players. The temptation to free ride leads to declining contributions.

If players, after each round, can impose fines on others, at a cost to themselves, the contributions increase. Sanctioning boosts the public good. But it is itself a public good, and players are tempted by second-order free riding (thus exploiting the willingness of others to punish free riders). (Fehr and Gächter, Nature, 2000)

Pool vs. peer punishment

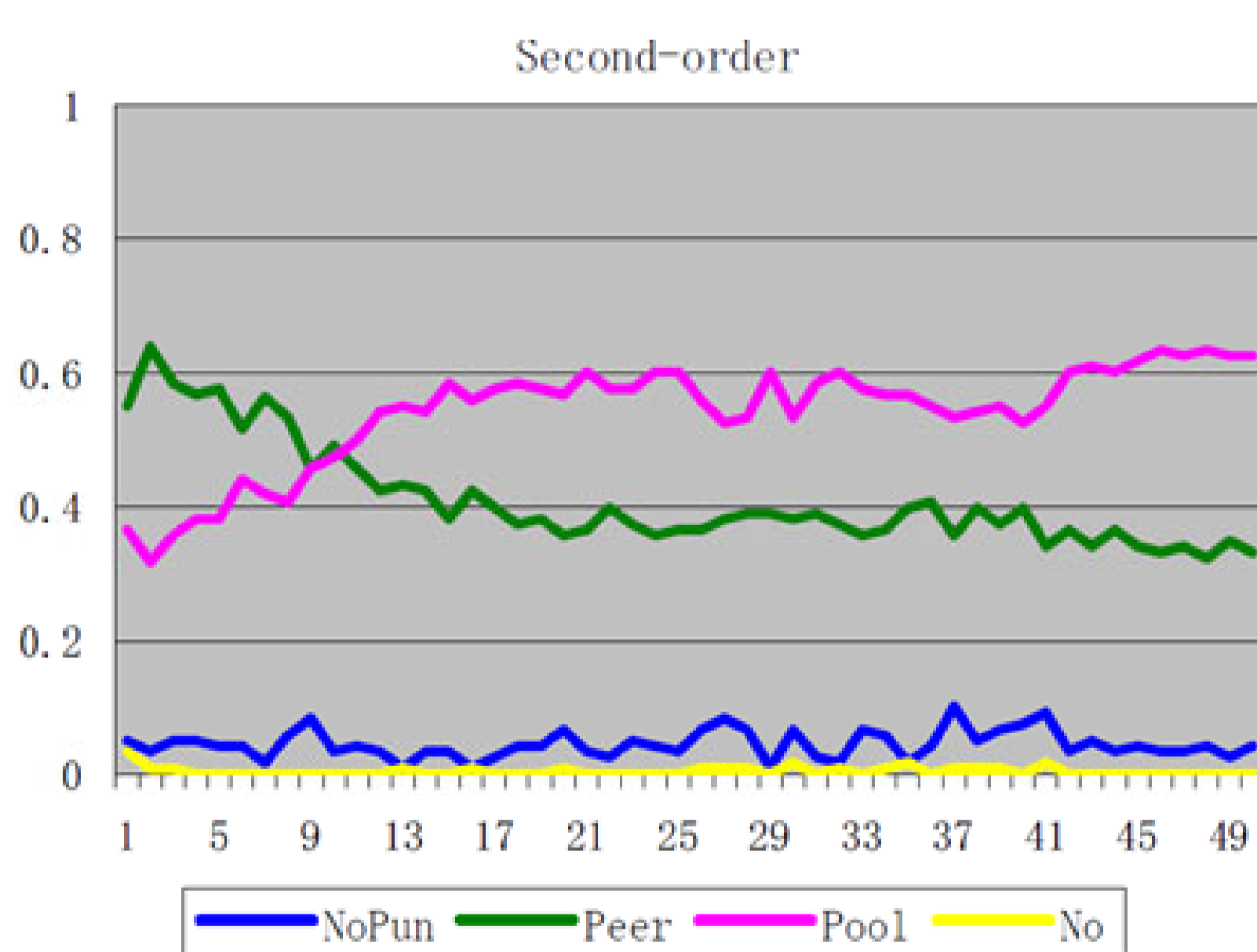
Peer punishment, as described above, can be viewed as self-justice. A more institutionalized form of sanctioning is pool punishment: players pay into a punishment pool before engaging in the public good game, and hence before knowing whether free riding will occur. This is costly even in the absence of free riders (whereas self-justice, in that case, is not).

Individual-based computer simulations show that pool punishment dominates if second-order free riders are punished (lower figure), but peer punishment dominates if they are not (upper figure). Both forms of sanctioning only emerge if players have also the option not to participate in the interaction. (Sigmund et al, Nature 2010)



In the game lab

Economic experiments confirm the theoretical results. If second-order punishment (of those who do not contribute to the sanctioning institution) is allowed, players switch increasingly from peer to pool punishment (see figure). If second-order free riders are not punished, peer punishment prevails, but leads to a less stable cooperative regime.



Social learning leads to a social contract

Most investigations, so far, have considered peer punishment: players can impose fines on those who exploited them, at a cost to themselves. Only a minority of studies have considered so-called pool punishment. In such scenarios, players contribute to a punishment pool before engaging in the public good game, without as yet knowing who the free riders are. Theoretical and experimental investigations show that peer punishment is more efficient than pool punishment but that pool punishment is more stable. Social learning leads to pool punishment if sanctions are also imposed on second-order free riders but to peer punishment if they are not.