Dynamic effects of enforcement on cooperation

Roberto Galbiati, Emeric Henry, Nicolas Jacquemet

Dynamic effects of enforcement on cooperation*

Roberto Galbiati† Emeric Henry‡ Nicolas Jacquemet§

November 2018

Abstract

In situations where social payoffs are not aligned with private incentives, enforcement with fines can be a way to sustain cooperation. In this paper we show, by the means of a lab experiment, that past fines can have an effect on current behavior even when no longer in force. We document two mechanisms: a) past fines affect directly individuals' future propensity to cooperate; b) when fines for non cooperation are in place in the past, individuals experience higher levels of cooperation from partners and, consistent with indirect reciprocity motives, are in turn nicer towards others once these fines have been removed. This second mechanism is empirically prevalent and, in contrast with the first, induces a snowball effect of past enforcement. Our results can inform the design of costly enforcement policies.

Keywords: Laws, social values, cooperation, learning, spillovers, persistence of institutions, repeated games, experiments.

A natural way to realign private and social incentives in social dilemmas is to impose fines on non cooperation: for instance, societies impose fines for smoking in public places, for the use of plastic bags or for littering. Policies based on fines are however costly to setup and enforce. In this paper, based on the results of a lab experiment, we show that the short term cost of temporarily enforcing fines can have longer term benefits, because of dynamic linkages in cooperation.

*Published in Proceedings of the National Academy of Sciences, vol. 115 (49), pp.12425-12428. The supplementary material is available online: `http://www.pnas.org/highwire/filestream/837492/field_highwire_adjunct_files/0/pnas.1813502115.sapp.pdf`. We thank Bethany Kirkpatrick for her help in running the experiment, and Gani Aldashev, Maria Bigoni, Frédéric Koessler, Bentley McLeod, Nathan Nunn, Jan Sonntag, Sri Srikandan and Francisco Ruiz Aliseda as well as participants to seminars at ENS-Cachan, ECARES, Montpellier, PSE, U. Paris Sud and to the 2016 ASFEE conference in Cergy-Pontoise, the 2016 SIOE conference in Paris, the 2017 JMA in Le Mans, the 2018 BEPC workshop in Lille and the 3rd Psychological game theory workshop in Soleto for their useful remarks on earlier versions of this paper. The authors gratefully acknowledge funding from CEPREMAP (program 3); Jacquemet acknowledges funding from ANR-10-LABX-93-0.

†Sciences Po and CNRS. Department of Economics, 28 rue des Saints-Pères 75007 Paris, France. roberto.galbiati@sciencepo.fr

‡Sciences Po and CEPR. Department of Economics, 28 rue des Saints-Pères 75007 Paris, France. emeric.henry@sciencepo.fr

§Paris School of Economics and University Paris 1 Panthéon-Sorbonne. Centre d'Economie de la Sorbonne, 106 Bd de l'hopital, 75013 Paris, France. nicolas.jacquemet@univ-paris1.fr

We argue that past fines may affect the future through two different channels: direct spillovers—past fines increase the current willingness to cooperate—but also indirect spillovers: past fines increase cooperation among people met in the past, and this cooperation results in higher willingness to cooperate today. A natural behavioral driver of this second effect is upstream indirect reciprocity, a well-documented phenomenon referring to the fact that cooperation of past partners is reciprocated by a higher propensity to cooperate with unrelated current partners; In the words of Nowak and Roch (2007) "*if someone is nice to you, you feel good and may be inclined to be nice to somebody else*".

These two mechanisms (direct and indirect spillovers) give rise to entirely different dynamic linkages in cooperation. Direct spillovers only change behavior in accordance with their own persistence. If direct spillovers are short-lived (i.e., only fines experienced in the immediate past change current behavior) then older fines have no impact on current cooperation. Indirect spillovers, by contrast, trigger snowball effects. Individuals who experience fines cooperate more and generate indirect spillovers on people they interact with. These partners will cooperate more in their next interactions and in turn generate indirect spillovers for their subsequent partners. This results in a chain of cooperative behaviors.

To empirically identify these mechanisms, our research design relies on a laboratory experiment that allows us to control the exposure to fines. Participants play a series of Prisoner's Dilemma (PD), a common paradigm to study situations where there is a conflict between private benefits and socially optimal outcomes in a controlled environment (Axelrod and Hamilton, 1981; Turner and Chao, 1999; Press and Dyson, 2012; Wang, Jusup, Wang, Shi, Iwasa, Moreno, and Kurths, 2017). From one game to the next, it is randomly determined whether a monetary fine, will be imposed on individuals who do not cooperate. This design ensures that each participant *(i)* has a different history of exposure to fines and of past behavior of partners; and that this history both *(ii)* does not depend on self-selection into particular environments, and *(iii)* is independent from the current environment faced by each individual. This design allows us to separately identify direct and indirect spillovers. It also allows us to demonstrate directly the existence of the snowball effect: we show that individuals respond to past environments with fines that they themselves never experienced.

Results

Direct and indirect spillovers

Figure 1 provides graphical evidence of the effect of current and past enforcement environments. Contemporaneous enforcement of fines that punish deviation do align private and social incentives (left-hand side vs right-hand side). The effect is large: as reported in Table 1, it leads to a 35% increase in cooperation. Moreover, having faced a fine in the previous match increases the current level of cooperation (left panel vs right panel in each subgraph), in the order of 5%—a statistically

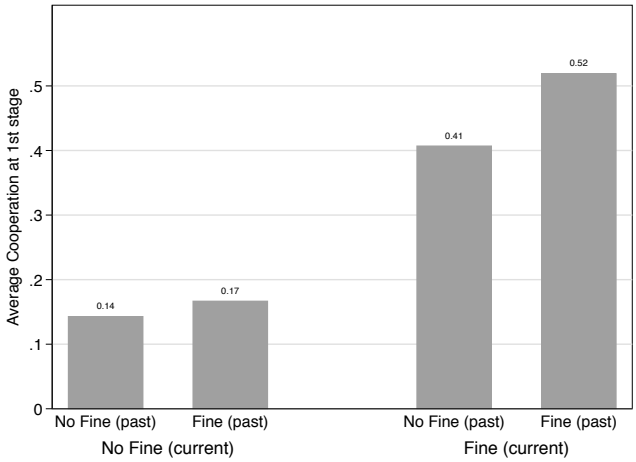

Figure 1: The effect of current and past enforcement

significant and sizeable effect, as reported in Table 1, column (2).

In Figure 2 we illustrate the two alternative mechanisms that can explain this effect of past enforcement. The left panel focuses on individuals who did not face a fine in the match played just before, while the right panel reports behavior of individuals who did experience a fine in the previous match. Within each panel we separate individuals whose partner cooperated in the previous match versus those who did not. Comparing behavior in the right and left panels, the striking pattern is that direct spillovers, once we account for the behavior of the partner, are weak: current cooperation is not affected by the presence of a fine in the previous match. On the contrary, comparing situations where the partner cooperated versus deviated in the last match shows that indirect spillovers are strong, independently of the current or past enforcement environments. This is confirmed by the statistical analysis in Table 1, column (3): whether the partner cooperated in the previous match captures all of the effect of past enforcement. In terms of magnitudes, cooperation by the partner in the previous match increases current cooperation by 13%.

Snowball effect of indirect spillovers

We have shown that the effect of past enforcement on current cooperation does not occur due to direct spillovers but rather because of strong indirect spillovers. We now present empirical evidence that indirect spillovers can lead to dynamic linkages that can create snowball effects of past enforcement. Consider the extreme example of short-lasting shifts in values arising solely from experience in the immediate past. In this case, there will be no persistent effects in time of past enforcement through direct spillovers: enforcement in the immediate past changes current cooperation but, since the shift in values is temporary, has no effect in the future. With indirect spillovers, by contrast, the enforcement of fines affecting immediate cooperation of partners also

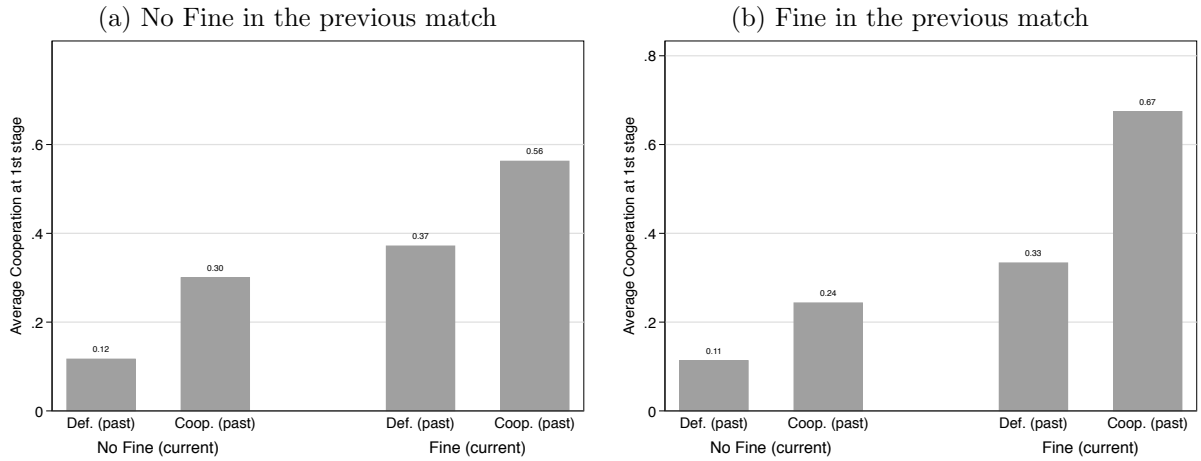

Figure 2: Indirect spillovers: the effect of previous partner's behavior

affects cooperation in the next period. This in turn transmits to cooperation in later periods, through subsequent indirect spillovers. Overall this creates a snowball effect of initial enforcement of fines.

We empirically test these ideas by examining how the history of play two games prior to the current one affects current cooperation. The results are presented in column (4) of Table 1. First, enforcement experienced two matches ago by an individual has no impact on her cooperation—suggesting direct spillovers are indeed short-lived. In contrast, if one's own partner in the previous match experienced a fine two matches ago, this has a strong and significant effect: it increases current cooperation by 5%. This provides strong direct evidence for the persistence of indirect spillovers, since this effect is observed on a player who did not experience the fine herself. In column (5) of Table 1, we directly control for whether the partner in the previous match cooperated. The results confirm that this persistence is due to indirect spillovers, since we no longer find any effect of enforcement experienced two matches ago. The enforcement environment two matches ago only matters to the extent it influenced the behavior of the partner in the previous match.

We have shown that even if indirect spillovers are short lived (i.e. only the behavior of the partner in the previous match matters), they create a snowball effect on later matches. We now study directly the question of what length of history matters. To examine the survival of indirect spillovers over time, we introduce the past history of cooperation in Table 2. Column (1), consistent with Table 1, shows that the number of successive matches played with a fine has no effect on current cooperation, i.e direct spillovers are short lived. In contrast, column (2) shows that the total number of matches in a row where partners cooperated has a significant and large effect. Column (3) breaks up this effect and shows that not only does the behavior of the partner in the previous match affect current cooperation, but also the behavior of the partners two and three matches ago. Overall, indirect spillovers affect cooperation several matches forward through

Table 1: Spillovers and the persistent effect of enforcement

	(1)	(2)	(3)	(4)	(5)
Fine in current (t) match	0.358***	0.364***	0.347***	0.359***	0.342***
	(0.072)	(0.070)	(0.058)	(0.052)	(0.039)
Fine in previous ($t-1$) match		0.054**	0.013	0.051*	0.010
		(0.021)	(0.032)	(0.029)	(0.045)
Partner in $t-1$ cooperated			0.129***		0.126***
			(0.032)		(0.040)
Fine in match $t-2$				-0.041	-0.038
				(0.076)	(0.070)
Partner in $t-1$ faced a fine in $t-2$				0.056**	0.044
				(0.028)	(0.028)
Observations	325	325	325	325	325
σ_u	1.416	1.410	1.341	1.423	1.357
ρ	0.667	0.665	0.643	0.669	0.648
LL	-126.962	-126.385	-124.029	-126.022	-123.785

Note. For each model, the table displays the marginal effects of a Probit model on the individual decision to cooperate—i.e., the estimated change in the likelihood to cooperate induced by a marginal change in the corresponding variable in row. See the SI Appendix for a detailed description of the specification, the full results and robustness checks. *Significance levels*: *10%, **5%, ***1%.

two channels: the snowball effect shown in Table 1 and the effect identified in Table 2 that sees behavior of partners faced several matches ago still influence current behavior.

Discussion

This paper provides evidence that past enforcement of cooperation with fines can have future effects even when these fines are no longer in force. This is due mostly to an indirect spillover effect that can create dynamic linkages in cooperation. These results suggest that if fines are costly to implement, a one time investment in an enforcement policy can have future effects. The social returns of enforcement can well be higher than the current benefits of cooperation, as future behavior adjusts to this experience.

Our results also sheds new lights on the empirical literature showing that individuals' propensity to cooperate with others is affected both by institutions experienced in the recent past (Fisman and Miguel, 2007) and by institutions that were in force in the distant past (Lowes, Nunn, Robinson, and Weigel, 2017; Guiso, Sapienza, and Zingales, 2016). Although the external validity of laboratory experiments needs to be treated with some caution, and the institution we consider is of a very specific form, the mechanisms highlighted in this paper can help explain some part of this persistence in the form of the snowball effect in cooperation due to indirect spillovers.

Table 2: Speed of decay of indirect spillovers

	(1)	(2)	(3)
Fine in current (t) match	0.365^{***}	0.361^{***}	0.358^{***}
	(0.063)	(0.056)	(0.056)
Number of matches in a row played with a fine	0.013	0.004	0.004
	(0.017)	(0.023)	(0.025)
Number of matches in a row played with a partner who cooperated		0.045^{***}	-0.008
		(0.006)	(0.014)
Partner in $t-1$ cooperated			0.126^{***}
			(0.032)
Partner in $t-2$ cooperated			0.109^{**}
			(0.051)
Partner in $t-3$ cooperated			0.039^{**}
			(0.017)
N	325	325	325
σ_u	1.402	1.346	1.319
ρ	0.663	0.644	0.635
LL	-126.631	-124.323	-121.671

Note. For each model, the table displays the marginal effects of a Probit model on the individual decision to cooperate—i.e., the estimated change in the likelihood to cooperate induced by a marginal change in the corresponding variable in row. See the SI Appendix for a detailed description of the specification, the full results and robustness checks. *Significance levels*: *10%, **5%, ***1%.

Materials and methods

At their arrival to the laboratory, participants receive a consent form. They are told signing the consent is mandatory to participate to the experiment. The design of the experiment received approval from the Paris School of Economics Institutional Review Board, under number 2018-016).

Participants in the experiment play a series of two player games we call matches. In each match, two subjects play several repetitions (called rounds) of a prisoner's dilemma. At the end of each round it is randomly determined whether there will be another round in the current match, so that the length of each match (i.e. the number of rounds in each match) is random. The probability of continuation at the end of each round is fixed at $\delta = 0.75$ and is independent of any choices players make during the match. At the end of each match, players are randomly and anonymously reassigned to a new partner to play the next match.

At the beginning of each match, a random draw determines whether the match is to be played with a fine (a fixed monetary cost imposed on the decision not to cooperate) or without (see the stage-game payoffs in the SI Appendix, Table S1). The two events occur with equal probability. The result from this draw applies to both players of the current match, and to all its rounds.

When matched with a new partner, subjects are not provided with any information about the partner's history.

Thanks to rematching and the random choice of the implementation of fines in every match, this experiment delivers individual observations of cooperation *(i)* under varying enforcement regimes (fine / no fine applying to the current match) and (*ii*) after individual specific histories of fines faced in the past. We apply two restrictions to the analysis of this data so as to focus on spillover effects of the exposure to past enforcement regimes. First, we summarize individual behavior by the decision to cooperate or defect at the first stage of each game.[1] Second, we focus on the games played late in the experiment to abstract from learning effects.[2]

References

Axelrod, R., and W. Hamilton (1981): "The evolution of cooperation," *Science*, 211(4489), 1390–1396.

Dal Bó, P., and G. R. Fréchette (2011): "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence," *American Economic Review*, 101(1), 411–29.

Fisman, R., and E. Miguel (2007): "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets," *Journal of Political Economy*, 115(6), 1020–1048.

Guiso, L., P. Sapienza, and L. Zingales (2016): "Long-Term Persistence," *Journal of the European Economic Association*, 14(6), 1401–1436.

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel (2017): "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom," *Econometrica*, 85(4), 1065–1091.

Nowak, M. A., and S. Roch (2007): "Upstream reciprocity and the evolution of gratitude," *Proceedings of the Royal Society B—Biological Sciences*, 274(1610), 605–609.

Press, W. H., and F. J. Dyson (2012): "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent," *Proceedings of the National Academy of Sciences*, 109(26), 10409–10413.

Turner, P. E., and L. Chao (1999): "Prisoner dilemma in an RNA virus," *Nature*, 398, 441–443.

Wang, Z., M. Jusup, R.-W. Wang, L. Shi, Y. Iwasa, Y. Moreno, and J. Kurths (2017): "Onymity promotes cooperation in social dilemma experiments," *Science Advances*, 3(3), e1601444.

[1] As described in the SI Appendix, Section S2, this is achieved by restricting the sample to observations for which the first stage decision fully summarizes the subsequent course of actions. For those observations, there is no loss of information since future actions can be predicted for sure for any sequence of play of the opponent.

[2] As documented in, e.g., Dal Bó and Fréchette (2011), learning strongly affects cooperation at early stages of the experiment. We thus restrict the analysis to later games. The thresholds defining late games are chosen in such a way that one third of the observed decisions are classified as "early", and one third as "late". All results are robust to alternative definitions of these thresholds as shown in the SI Appendix, Section S3.