American Sociological Association

Social Psychology Quarterly
2020, Vol. 83(1) 70–87

DOI: 10.1177/0190272519882389
journals.sagepub.com/home/spq
SAGE

# Aggression, Conflict, and the Formation of Intimidating Group Reputation

**Aron Szekely[1,2], Giulia Andrighetto[2,3,4], Nicolas Payette[5], and Luca Tummolini[2]**

***Abstract***

*From inmates in prison gangs to soldiers in elite units, the intimidating reputation of groups often precedes its members. While individual reputation is known to affect people's aggressiveness, whether one's group reputation can similarly influence behavior in conflict situations is yet to be established. Using an economic game experiment, we isolate the effect of group reputation on aggression and conflict from that of individual reputation. We find that group reputation can increase the willingness to inflict costs on others but only when individuals are able to punish their fellow members. Even if internal discipline can sustain their shared reputation, more intimidating groups provide fewer benefits to their members in the short run. Using an agent-based simulation, we show that this might not be the case in the long run. Our findings yield insights into the effects of group reputation on aggression, conflict, and possible consequences for group survival.*

**Keywords**

aggressive behavior, conflict, group reputation, peer punishment, social dilemma

Establishing a reputation for being aggressive—the belief held by others that someone is willing and able to inflict costs on other people (Barclay 2015)—can be useful in many social situations, from personal protection (Nisbett 1996) to the deterrence of free-riders (Raihani and Bshary 2015) but also from bargaining with the power to hurt (Schelling 1970) up to predation and extortion (Buss and Shackelford 1997). Theoretical models (Johnstone and Bshary 2004; McElreath 2003) and experimental studies (Benard 2013; Przepiorka et al. 2019; Szekely and Gambetta forthcoming) demonstrate that one's willingness to behave aggressively and fight in diverse contexts is influenced by the potential benefits that can be derived from one's personal reputation (*individual reputation*).

In many situations, however, these ensuing benefits depend less on one's

[1]Collegio Carlo Alberto, Turin, Italy
[2]Institute of Cognitive Sciences and Technologies, Italian National Research Council, Roma, Lazio, Italy
[3]Institute for Futures Studies, Stockholm, Sweden
[4]Mälardalen University, Vasteras, Sweden
[5]University of Oxford, Oxford, UK

**Corresponding Author:**
Aron Szekely, Collegio Carlo Alberto, Piazza Arbarello 8, Turin, TO 10122, Italy.
Email: aron.szekely@carloalberto.org

personal reputation and more on the reputation of the group that one is associated with (*group reputation*). In intergroup conflicts, for instance, the perceived aggressiveness of groups as a whole takes precedence over that of its individual members (De Dreu et al. 2016). Given the importance of group reputations for inflicting costs on others, many groups take steps to manage it. The underworld of prison gangs and organized crime provides relevant examples (Skarbek 2014). The Aryan Brotherhood, a violent neo-Nazi prison gang, values its violently created reputation and ensures that it remains intimidating by specifying and enforcing internal norms of "honor" and using would-be members to conduct violent acts as part of their initiation (Grann 2004). Sicilian Mafia families similarly maintain their violent reputation by forcefully shutting down competitors and employing internal monitoring and sanctions (Gambetta 1993; Orlando 2001). Analogously, a concern for sustaining an intimidating group reputation is also found in the "upper world" of police, the military, and among states.

Harboring an intimidating reputation at the group or aggregate level is not only consequential for already established groups but also affects social interactions before a formal, organized group actually exists. Consider that many conflicts in modern societies happen among strangers: anonymous individuals who meet occasionally and do not have access to information about their reciprocal personal past. It is well known that when personal experience is lacking, information about the social category one can be associated with becomes more salient (Fiske and Neuberg 1990). Beliefs that someone else has a particular trait can be grounded on such spontaneous association of strangers with social categories. Indeed, it has been hypothesized that by painting different individuals with the same brush, observers create reputational spillovers (Barnett 2016) that also affect how individuals who are not directly involved in an interaction will be treated in the future. Consequently, strategic interdependences can arise between people who were not previously linked in any structured way (King, Lenox, and Barnett 2002), thereby creating the need to manage an emerging reputation at the group level. For group reputation to matter and affect how its "members" behave, being grouped together from the outside can be enough.

Despite its relevance and ubiquity, the scientific literature has paid only limited attention so far to the formation of aggressive reputation at the group level. Moreover, it has not been established whether a concern for an intimidating group reputation, like that for one's personal reputation, has effects on the willingness of individuals to be aggressive and engage into overt conflict. Here we use an economic game experiment (N = 238) and an agent-based simulation to study the effect that a group reputation for aggression has in conflict situations. Specifically, we ask: How do groups create and maintain their reputation for aggression? Does this process differ from individual reputation formation? Can intragroup enforcement of norms using a punishment mechanism allow the formation of intimidating group reputations, and what are its effects on conflict?

## THEORY AND HYPOTHESES

### *Intimidating Reputations: From Individuals to Groups*

The effects of personal reputation on individual prosociality and group cooperation have been explored by an extensive interdisciplinary literature. Theoretical and empirical studies, including image scoring (Nowak and Sigmund 1998a, 1998b) and standing strategy models (Leimar and Hammerstein 2001; Ohtsuki and Iwasa 2004), show that when past

behavior is observable, individuals are more willing to provide benefits to others and groups are able to maintain cooperation at high levels (for reviews, see Rand and Nowak 2013; Tennie, Frith, and Frith 2010). Although less extensive, theoretical (e.g. Fearon and Laitin 1996; Healy 2007; Tirole 1996) and experimental works (e.g. Kimbrough and Rubin 2015; McIntosh et al. 2013) have also explored whether the possibility to form and maintain group reputation can similarly affect individual prosociality and cooperation. While findings are somewhat mixed (Engelmann, Herrmann, and Tomasello 2018; Huck and Lünser 2010), evidence is emerging that group reputation itself is harder to create and sustain, with less stable effects on prosociality and cooperation (Duca and Nax 2018; Nax et al. 2015).

Compared to cooperation, reputation for aggression—inflicting costs on others—has, in contrast, received little attention. This literature has primarily focused on personal reputation and asked two separate but related questions. First, does the opportunity to create an intimidating reputation increase or decrease aggressive behavior? Second, does this possibility affect the amount of conflict that ultimately arises (e.g., Benard 2013, 2015; Gambetta 2009; Przepiorka et al. 2019; Szekely and Gambetta forthcoming)? While the former concerns individual-level action, the latter is focused on the outcome that arises from interactions. Aggressive behavior in this literature is considered to be an action that inflicts costs, or harm, on others and is often related to the concept of "instrumental aggression" (Anderson and Bushman 2002). While aggression is necessary for conflict to emerge, it is not sufficient because whether conflict actually arises also depends on what the others in the interaction do. Only if the other parties are also either aggressive or willing to defend themselves does conflict, or a "fight," actually occur. This means that there is the potential to coordinate to avoid actual conflict even if (or precisely because) one of them is thought to be highly aggressive (McElreath 2003).

Game theoretic models provide some insights and suggest that reputation systems can increase aggressive behavior. Building on Selten's (1978) chain store game, theoretical models have shown that the possibility to create a tough reputation allows both aggressive behavior and fighting (predatory pricing and market retaliation in their context) to emerge (Kreps and Wilson 1982; Milgrom and Roberts 1982; for an early experimental test, see Jung, Kagel, and Levin 1994).

In the context of prison fighting, Gambetta (2009:78–110) has argued that the aim of much violence is precisely to create an intimidating reputation but that the "more notorious an agent is for violence, the less he has to commit to prove his reputation." Interpreted more generally, this argument suggests that one can become more aggressive to establish an intimidating reputation but that once an individual's reputation is clear, he or she has little further motivation to behave aggressively: information is negatively associated with fighting. Since people know who would win or lose in case of an actual conflict, there is no need to undertake it (see also Gould 2003). A few experiments have tested and found support for both ideas. There is evidence that reputation systems can cause aggression (Benard 2013; Griskevicius et al. 2009) but also that they reduce both aggression and resulting conflict (Przepiorka et al. 2019; Szekely and Gambetta forthcoming). Yet whether and how reputation at the aggregate or group level can affect aggression and conflict remains an open question.

### *Hypotheses*

The two main hypotheses of the present study concern aggressive behavior. Consistently with previous theoretical work (King et al. 2002), we conjecture that the creation and maintenance of group reputation is fundamentally different from that of individual reputation. Individuals often have direct incentives to maintain their own fearsome reputation. In contrast, when a reputation is shared with others, it becomes a resource that group members hold in common and can either sustain or exploit. As the cost of maintaining an intimidating group reputation is borne by each individual while the benefits are distributed across the group, the production and maintenance of an intimidating group reputation can be modeled as a social dilemma. Based on this, we hypothesize:

> *Hypothesis 1:* The formation of group reputation poses a social dilemma such that it is overexploited. In our experiment, this implies that aggressive behavior is reduced when reputation is at the group level relative to when reputation is at the individual level.

Among the many mechanisms that are available to solve social dilemmas (see Kollock 1998; Rand and Nowak 2013), a common one is peer punishment—costly or otherwise. If the creation and maintenance of an intimidating group reputation is a social dilemma, then it is possible that peer punishment can allow a group to enforce a norm of contributing, and thereby they can form and maintain an intimidating reputation. Hence, we expect:

> *Hypothesis 2:* When peer punishment is possible, groups will be able to uphold their reputation, and thus aggressive behavior will increase at the individual level.

Whether the opportunity to build a group reputation increases or decreases conflict in our experiment is unclear, and we did not make a prediction about this. Previous research on conflict and information has reported mixed results, and none have empirically studied this question at the group level.

## EXPERIMENTAL STUDY

### *Methods*

The experiment was conducted at the CESARE laboratory (Libera Università Internazionale degli Studi Sociali [LUISS] Guido Carli, Rome) on 238 healthy volunteer subjects (110, 46.22 percent female; mean age = 21.8 years, SD = 2.22, minimum = 18, maximum = 31) recruited using Online Recruitment System for Economic Experiments (ORSEE) (Greiner 2015) from the local subject pool. The experiment lasted for about 1 hour, the maximum possible earning was €16, the minimum was €0 (excluding the €4 show-up fee), and subjects' average earnings, excluding show-up fee, was €9.17. Each subject participated in 1 of 12 sessions. Each session comprised 20 subjects except for one session in which there were 18 subjects. We received approval from the CESARE laboratory ethical committee and the Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie della Cognizione (CNR-ISTC) Institutional Review Board. All research was performed in accordance with the relevant guidelines and regulations, and informed consent was obtained from all participants.

Our sample size and stopping rule for data collection were determined before running any of the sessions and was not changed during the experiment. Our sample size was planned using the public goods game literature (Fehr and Gächter 2000) and the two closest articles on group reputation (Huck and Lünser 2010; Kimbrough and Rubin 2015). To

maximize statistical power, we obtain repeated measures on each subject throughout our experiment. We do not exclude any data, and no observations are treated as outliers. Our predetermined research objective was to compare individual-level reputation and group-level reputation, and both of our hypotheses were prespecified.

The experiment is programmed in z-Tree (Fischbacher 2007). Subjects received printed instructions, and the instructions were read aloud by an experimenter.[1]

*Procedure.* Subjects participate in our game (Figure 1, top panel) as either in the role of Person A (*A*) or of Person B (*B*), and they keep their roles for the duration of the experiment.[2] At the beginning of the session and before the random allocation of roles, subjects made one practice decision as *A* and one practice decision as *B*. Subsequently, subjects answered control questions about these roles on their computer screens, and an experimenter read out the correct answers and provided an explanation. Subjects were then randomly allocated to play as either *A* or *B*. They participated in one of four experimental conditions that they were randomly allocated into at the session level (N = 60 in three treatments and 58 in one treatment; Figure 1, bottom panel).

Subjects in all treatments were told the following. *A* receives €16 and *B* receives €8, *A* decides between keeping all the endowment (sending €0) and sending €8 to *B*, and in case *A* chooses to keep, *B* can reduce *A*'s earnings at a 1:4 ratio (*punishment*) (see Figure 1). This means that if *A* decides to keep and *B* chooses not to punish *A*, then *A* earns €16 and B earns €8. The inequality of initial endowments is designed to represent a situation in which *B* is potentially motivated to take the earnings of *A*—consider mafiosi attempting to extract protection money from shopkeepers or prisoners taking the resources of others. If *A* sends, then the inequality of earnings is reversed such that *B* now received more than *A*. As such, this game induces a direct conflict in incentives between *A* and *B*, one gains what the other loses, and it is not possible to achieve a cooperative outcome (for a more general discussion of this game, see Selten 1978).

We elicit the decisions of subjects in different ways according to the role they

[1]Our instructions, analyses, and data for the experiment and the simulation are available at https://osf.io/6faj8/.

[2]Beside common knowledge of their roles, *A* and *B*, neither in the group history nor in the group history + peer punishment treatments, are explicitly told that they belong to different groups, nor are they given group tasks to prime or artificially create a sense of group membership. We do not adopt the minimal group paradigm (Tajfel et al. 1971) or natural groups in our study for two reasons. First, to compare the effects of intimidating reputation at the individual and at the group or aggregate level, we aimed to keep the individual and the group history treatments as similar as possible. Instead of observing different signals for each *B* as in individual history, *A*s in our group history treatments observe a common signal about the (aggregated) behavior of *B*s, which creates reputational spillovers among *B*s and potentially incentivizes them to manage an emerging reputation at the group or aggregate level by making it impossible for *A*s to discriminate between them (Healy 2007; Huck and Lünser 2010; King et al. 2002). Whether subjects in the role of *B*s without being prompted with group membership cues spontaneously realize their shared fate, expect one another to realize it, and appropriately respond to it is part of our research question. Second, creating or priming one's social identity can be conceived as another mechanism able to overcome a social dilemma (Kollock 1998; for a recent meta-analysis see Balliet, Wu, and De Dreu 2014) that is potentially alternative to the use of internal discipline, which we explore in our group history + peer punishment. Whether social identification can sustain group reputation without any other enforcement mechanism is an interesting question that, however, goes beyond the scope of the present study.

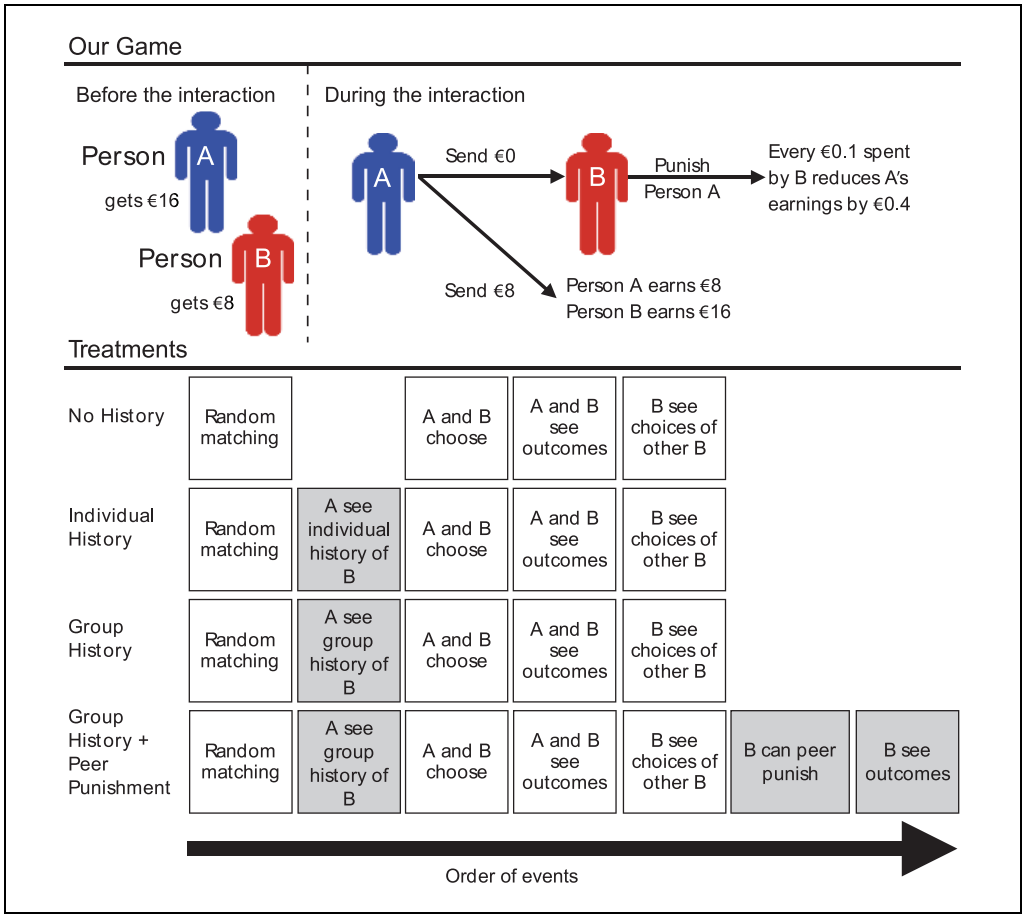


**Figure 1.** The Conflict Situation that Participants Engage in (top panel) and the Treatments in the Experiment (bottom panel)

play. *A* decides between sending €8 or sending €0 (keeping), while *B* decides how much they would spend on punishing in case *A* decided to keep (the strategy method; see Selten 1967). In other words, *B* preselects what would happen if *A* chooses to keep, and in case *A* actually chooses to keep, this action plan is implemented. By recording *B*'s decisions in this way, we can observe the choices of all *B*s instead of those made only by the subset of *B*s who faced *A*s that decided to keep.

We operationalize aggression as the amount spent by *B* on punishing *A*. The level of aggression is solely controlled by *B*, and since it is actually carried out if *A* keeps, it represents a commitment to inflict harm on *A* by *B*. We operationalize conflict as the actual amount of punishment that is inflicted by *B*s on *A*s. To have conflict, both *B* needs to be willing to punish *A* and *A* has to defy the implicit request of *B* and keep. Specifically, our measure of conflict is the average punishment actually inflicted by *B* on *A* when *A* keeps. This captures both the proportion of *A*s who keep and are subsequently punished and the intensity of such punishment.

Subjects are also told that they will observe the outcome of their interaction,

that *B*s observe the punishment strategy of other *Bs* (see the following); that the session lasts for 11 periods; and that they are randomly rematched with other subjects after each period. Subjects know how many *A*s and *B*s there are in each session. What else subjects are told depends on the treatment (Figure 1, bottom panel).

- In the *no history* treatment, there are no additional features.
- In the *individual history* treatment, *A*s are shown the history of punishment strategies of the *B*s that they are matched with.
- In the *group history* treatment, *A*s are shown the history of the average punishment strategies of the *B*s in the session.
- In the *group history + peer punishment* treatment, subjects have the same information as in group history and *B*s have the opportunity to reduce the earnings of other *B*s at a 1:4 ratio (*peer punishment*).

What changes between treatments is whether *As* observe the history of punishment strategies chosen by *B*s, whether this is the specific history of the individual they are matched with or is aggregated at the group level, and whether *B*s have the possibility to use within-group peer punishment. In contrast, in all treatments, each *B* has the opportunity to observe the punishment strategy of the other *B*s in his or her session. Such mutual observability among *B*s is necessary to implement the within-group peer punishment mechanism but could also enable peer effects among *B*s. To avoid this potential experimental confound, we keep this feature constant in all treatments.

In individual history, group history, and group history + peer punishment, *B*s transmit their strategies, and not actions, to *A*s. This means that *A* observes either the actual amount of punishment that *Bs* have inflicted on previous *As* or the amount they would have inflicted if the *A* they were previously matched with had decided to keep. By transmitting strategies (instead of actual actions), we are able to provide a cleaner identification of the effect of reputation on the behavior of *A*. Implemented this way, history contains only one type of information—average intensity of punishment—and it prevents the decision of *A* from affecting the punishment history created by *B*. Transmitted strategies can be considered as an operationalization of the circulation of subjects' commitment and willingness to punish that can be obtained through threats, gossip, and body language.

In addition to actions and strategies, we elicit the beliefs of *A* regarding the amount of punishment that *B* has chosen and the beliefs of *B* about how likely *A* is to send €8. At the end of the experiment we asked subjects to fill the Buss-Perry Aggression Questionnaire (BPAQ; Buss and Perry 1992), a short questionnaire on their demographic characteristics, and a nonincentivized pen-and-paper version of the lottery risk elicitation (Holt and Laury 2002). We conducted exploratory analyses of *B* punishing behavior using the BPAQ but found no relationships between BPAQ and punishing. We also explored whether risk preferences, as measured by the lottery risk elicitation, moderated punishing by *B* in the different treatments; however, we ultimately decided against using this analysis because the elicitation is not incentivized and a nontrivial proportion of subjects did not fully complete it.

To legitimize peer punishment decisions and avoid antisocial punishment (Faillo, Grieco, and Zarri 2013), *B* in the group history + peer punishment treatment can only reduce the earnings of other *Bs* who spent less than them on punishing.

*Predictions.* The behavior of subjects in no history provides an empirical baseline for *A* sending and *B* punishing. This is because in this treatment it is not possible for *B* to create a reputation, and thus, punishing is likely to be driven by concerns with inequality of initial endowments. Individual history serves to check that people in our experiment are concerned with individual reputation and as a comparison to group reputation creation. If Hypothesis 1 is correct, the social dilemma will arise in group history, and aggression will decrease. More specifically, *B* should punish less in group history than in individual history and at a similar level to no history. If the creation of an intimidating group reputation suffers from a social dilemma, we predict in Hypothesis 2 that peer punishment allows groups to overcome the issue (Fehr and Gächter 2000; Ostrom, Walker, and Gardner 1992), so peer punishment in the group history + peer punishment treatment should allow the social dilemma to be solved. This means that punishment of *As* in group history + peer punishment should be higher than in no history and group history.

*Statistical approach.* Analyses of the experimental data are conducted using Stata IC v13.1. We use logistic regressions to test the predictors of *A* sending. These included controls for male, period, and punishment received, and we cluster at the individual level to account for the multiple observations per subject. Our results are the same without controls and for an alternative specification of reputation.[3]

To analyze *B* punishing *A*, peer punishment against other *Bs*, conflict among *A* and *B*, and earnings, we use linear regressions with standard errors calculated using the wild cluster bootstrapped-t procedure with clustering at the session level. Session-level clustering accounts for the intrasession correlation of subjects' actions that arises because *B*s can observe how much other *B*s punish in each period. We take specific steps to account for the few session clusters (we have 12 sessions) using the wild cluster bootstrapped-t procedure (Cameron, Gelbach, and Miller 2008; Cameron and Miller 2015; Esarey and Menger 2019; Harden 2011; Kézdi 2004). We implemented this using *boottest* in Stata (Roodman et al. 2019). Following recommendations (Cameron et al. 2008; Cameron and Miller 2015), we impose the null and bootstrap for 10,000 repetitions and use Webb weights (Webb 2014). We also use bias reduced linearization (Bell and McCaffrey 2002), another method designed for cluster robust inference with few clusters, and show the same results.[4] All *p* values are derived from two-sided tests.

### Results

*Reputations for aggression provide benefits to Bs.* We start by analyzing whether the observable history of *B*s affects the behavior of *A*s. This is a precondition because *B*s only have incentives to invest in individual and group reputation if *A*s are intimidated. To test this, we predict the decision of *A*s to send or keep based on the observable history of punishment strategies (more precisely the average punishment that has been committed by *B*s in previous periods of the experiment). The higher this history is, the more, on average, *B*s decided to punish.

[3]See online Appendix A. Experiment Results for models with and without controls and detailed results.

[4]See online Appendix A. Experiment Results.

As expected, *A*s are likelier to send when they observe higher individual histories (odds ratio [OR] = 1.26, $Z = 4.51$, $p < .001$) and higher group histories (OR = 1.53, $Z = 5.15$, $p < .001$); these associations can be seen in Figure 2, Panels B and D. Analyzing the relationship between sending and history of punishment this way assumes that all punishment decisions of *B*s are equally weighted across the rounds by *A*s. Yet, this may not be the case. It is possible that *A*s pay more attention to the more recent punishing behavior of *B*s. In additional analyses, we show that if we consider only the previous period punishing of *B*s, then *A*s are still likelier to send when they observe higher histories.[5]

Next, we test whether the response of *A*s to *B*s' history is driven by their beliefs about the willingness of *B*s to punish. The beliefs of *A*s about *B*s' willingness to punish shows the same relationship with sending (individual history: $b = .15$, $t[118] = 9.96$, $p < .001$; group history and group history + peer punishment: $b = .25$, $t[118] = 7.31$, $p < .001$) when history is included as the average punishment in all of the previous periods. These results hold if we consider the alternative specification of punishment history and only use previous period punishment.[6]

Together, these results imply that since *A*s sends money because of their beliefs about *B*s (because of *B*s intimidating reputation), establishing such reputation can provide substantial benefits to *B*s.

*Group reputation without peer punishment is exploited and reduces aggression.* Consistent with Hypothesis 1, group reputation is naturally exploited and aggression is reduced. Punishment in group history is €1.38 (SD = 1.53; Figure 2A). This is below the €2 level needed to make *A*s indifferent between sending and keeping, and it is comparable to punishing in the no history treatment (M = 1.30, SD = 1.45; $b = .08$, $t[11] = .37$, $p = .744$). Despite the incentives to establish an intimidating group reputation (see previous result), punishment in group history is indistinguishable from punishment in no history. The dynamics of reputation formation confirms that it is exploited: punishment in group history never makes it above €2 in any period.

Conversely, individual reputations for aggression are consistently created and maintained in individual history (M = 2.08; SD = 1.47; Figure 2A). This is significantly more than punishment in no history ($b = .78$, $t[11] = 5.92$, $p = .009$) and group history ($b = .70$, $t[11] = 2.95$, $p = .019$). For much of the experiment (7/11 periods), average punishment in individual history is at or above the €2 indifference point. Subjects spontaneously create intimidating reputations when the history of their own past punishment strategies is observable. Consistent with reputation building concerns, we find an endgame effect in individual history that decreases to €1.40, a level indistinguishable from that found in no history ($b = .31$, $t[11] = .60$, $p = .581$).

*Group reputation can be maintained with peer punishment and increases aggression.* Consistent with Hypothesis 2, when *B* can punish other *Bs* (peer punishment), groups can form and maintain an intimidating group reputation. Average punishment rises to €2.29 (SD = 1.54) in group history + peer punishment (Figure 2A). This is significantly higher than punishment in no history ($b = .99$, $t[11] = 4.64$, $p = .014$) and group history ($b = .91$, $t[11] = 3.12$, $p = .013$) and similar to punishment in individual history ($b = .21$, $t[11] = .87$, $p = .529$). For most of the 11 periods (8/11), punishment in group

[5]See online Appendix A. Experiment Results.
[6]See online Appendix A. Experiment Results.

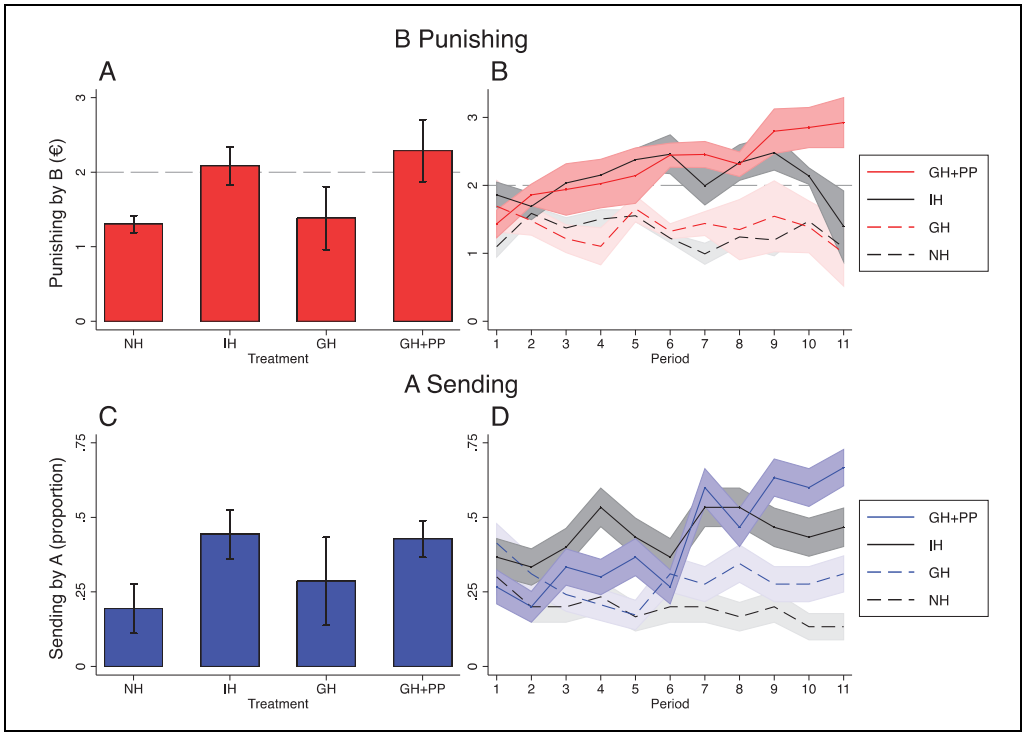


**Figure 2.** Punishing by *B* and Sending by *A:* (A) Average Punishment by *B*, (B) the Dynamics of Punishment, (C) Average Proportion of *A* Sending, (D) the Dynamics of Sending

*Note:* Error bars represent 95 percent confidence intervals with clustering at the session level for (A) and at the individual level for (C). Shaded areas indicate ±1 standard error of the mean with clustering at the session level for (B). Reference line indicates the punishment level at which *As* are indifferent between sending and not sending. NH = no history; IH = individual history; GH = group history; GH+PP = group history + peer punishment.

history + peer punishment is at or above the €2 level. Unlike individual reputation, group reputation shows no signs of an endgame effect (group history + peer punishment vs. no history: $b = 1.84$, $t[11] = 4.98$, $p = .012$) so that by the final period, there is a €1.5 difference in punishment spending between individual history and group history + peer punishment ($b = 1.53$, $t[11] = 2.46$, $p = .078$).

Although peer punishment among *Bs* has a dramatic effect on the level at which *As* are punished, it is infrequently used. Peer punishment is used only in 18 percent (213/1186) of opportunities, and even when it is used, little is spent (mean = €.73). Yet peer punishment is targeted: the more *B*s deviate below the group's mean punishment, the more their earnings are reduced ($b = -.02$, $t[2] = -3.56$, $p = .091$). Moreover, it is effective in making *B*s increase their punishing. Following a reduction in their earnings, *B*s increase their punishing in the subsequent period ($b = .81$, $t[2] = 6.26$, $p = .097$). We next examine which *B*s peer punish and test if *B*s who punish *A*s more are also those who engage in peer punishment more. We find that this is not the case: there is little association between *B*s' punishment toward *A*s and *B*s' spending on peer punishment.

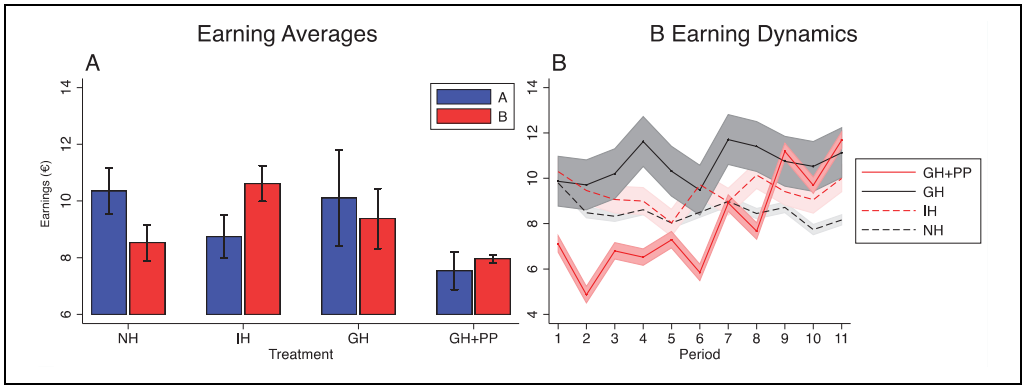


**Figure 3.** Earnings of Subjects by Treatment

*Note:* (A) Peer punishment reduces the earnings of *B* and *A*. Error bars represent 95 percent confidence intervals with clustering at the session level. (B) The dynamics of *B* earnings. Shaded areas indicate ± 1 standard error of the mean with clustering at the session level. NH = no history; IH = individual history; GH = group history; GH+PP = group history + peer punishment.

*Group reputation with peer punishment increases conflict.* Average punishment inflicted in group history + peer punishment is €5.05. This is significantly higher than conflict in no history (€4.10; $b$ = .95, $t$[11] = 5.08, $p$ = .013), individual history (€3.72; $b$ = 1.33, $t$[11] = 8.38, $p$ = .008), and group history (€3.62; $b$ = 1.42, $t$[11] = 4.68, $p$ = .012). This arises from two factors: a greater proportion of interactions between *A*s and *B*s leads to conflict in group history + peer punishment (48.18 percent) and no history (46.97 percent) than in the other two treatments (individual history: 39.70 percent; group history: 37.93 percent), and conditional on not paying, the intensity of punishment inflicted is highest in group history + peer punishment (€8.81 vs. no history: €5.09; individual history: €6.67; group history: €5.07).

*Peer punishment reduces short-run earnings.* Counterintuitively, within-group cooperation to maintain an intimidating group reputation harms both group members and those outside the group. This is because of the higher conflict between *A*s and *B*s and the intragroup cost of peer punishment. Both *A*s and *B*s earn little in the group history + peer punishment treatment (M = 7.53, SD = 4.72 and M = 7.96, SD = 5.60, respectively), leading to an inefficient system overall (Figure 3A). *B*s earn less in group history + peer punishment than in all treatments except in no history (vs. group history: $b$ = 1.41, $t$[11] = 2.60, $p$ = .092; $p$ = .045 for the first 10 periods; vs. individual history: $b$ = 2.65, $t$[11] = 8.11, $p$ = .012; vs. no history: $b$ = .56, $t$[11] = 1.70, $p$ = .160). *A*s earn less in group history + peer punishment than in all treatments except in individual history in which the difference is close to significance (vs. group history: $b$ = 2.56, $t$[11] = 2.76, $p$ = .023; vs. individual history: $b$ = 1.21, $t$[11] = 2.34, $p$ = .093; vs. no history: $b$ = 2.81, $t$[11] = 5.28, $p$ = .010).

## SIMULATION STUDY

A potential explanation for why earnings are low in group history + peer punishment is that there are short-term net costs to building an intimidating reputation but that once it is created and maintained for some time, an intimidating group reputation can indeed provide net

benefits. Although our experiment captures only short-term dynamics, this possibility is consistent with our data: *B*s' earnings in group history + peer punishment increase over time (Figure 3B). Previous research on short-run experiments in the public goods game literature also finds a similar result (Fehr and Gächter 2000; Gürerk, Irlenbusch, and Rockenbach 2006). Yet longer experiments show that after an initial costly period, peer punishment increases earnings (Gächter, Renner, and Sefton 2008). To test this possibility in our setup, we create an agent-based model of our experiment and simulate what happens in a longer version of our experiment.

### Methods

Our simulation uses the same rules as the experiment, and we infer agents' decision-rules from the data.[7] The only entities in the model are the players. Like in the experiment, we have two types of simulated roles: Person A (*A*) and Person B (*B*).

Agent *A* chooses whether to pay *B*. This decision is based on the history of *B* (when history is available in a treatment) and the punishment received in the previous period—the two factors we find in the experiment that influence *A* sending. The precise strength of the association is based on logistic regressions for sending or keeping.[8]

Agent *B* chooses whether to take two actions: whether and how much to punish *A* and whether and how much to punish other *Bs* (peer punishment).

Punishing *As* is based on two inputs. The first input is given by the "type" of *B* that a specific agent instantiates. Using the algorithm specified by Kurzban and Houser (2005), we identify four types of *B* in our data set: *nonpunishers* who consistently do not punish, *punishers* who consistently punish, *positive conditional punishers* who increase their punishment the more others punish, and *negative conditional punishers* who reduce their punishment the more others punish.[9] The first three types are standard in the public goods game literature; the fourth type has also been found in other studies (Kurzban and Houser 2005). The second input is the amount of punishment received from other *Bs*.

The second action of *B* is whether to punish other *Bs* (peer punishment). Specifically, *B* has to decide which other agent *B* to punish and how much. Both decisions are influenced by the amount that the potential target deviates from the group average punishment of *As* and the difference in punishment between the potential peer punisher and the potential target. The more a target deviates below the group descriptive norm (Cialdini, Reno, and Kallgren 1990) the likelier and with greater intensity is *B* peer punished, and the greater the difference in punishment between the actor and the target, the likelier and with greater intensity is the target peer punished.

We run the simulation for 100 periods and repeat it 10,000 times for each treatment to ensure robustness. The agent-based model uses the ODD protocol (Grimm et al. 2010) that is built using NetLogo v6.0.1 (Wilensky 1999), and analyses of the simulation data are conducted using Stata IC v13.1.

### Results

Comparing the earnings of *B* in the agent-based simulation and the

[7]See online Appendix B. Simulation Details.

[8]See online Appendix B. Simulations Details: Table B5.

[9]See online Appendix B. Simulation Details: Input Data.

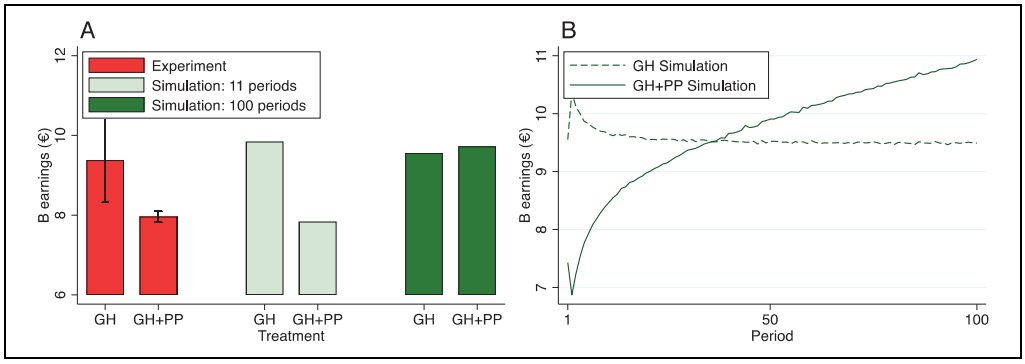


**Figure 4.** Earnings in Our Simulation

*Note:* (A) Peer punishment reduces the earnings of *B* in the experiment and the short run, but it increases *B* earnings in the long run. Error bars represent 95 percent confidence intervals with clustering at the session level. Confidence intervals are not plotted on simulation data as they are negligible. (B) From period 36, peer punishment increases earnings. Simulated data shown for the 100 periods averaged over 10,000 runs for each treatment. Confidence intervals are not plotted on simulation data as they are negligible. GH = group history; GH+PP = group history + peer punishment.

experiment shows that simulated *B* earnings after 11 periods are consistent with that of human participants in all four treatments (Figure 4).[10] Simulated *B*s in group history + peer punishment, like their human counterparts, earn less in the first 11 periods than they do in group history (Figure 4A). Yet, in the long-term, peer punishment does not reduce earnings: the earnings of *B*s in group history + peer punishment increase to higher levels than in group history. By period 36, *B*s in group history + peer punishment earn more (Figure 4B). This is because *B*s in group history + peer punishment punish more, but once their group reputation is well established, *A*s pay more frequently and less conflict occurs.

## DISCUSSION

Our study demonstrates that being able to form and maintain an intimidating group reputation can provide substantial benefits for its members: as the average group history that *A*s observe increases, the probability of paying *B*s also increases. Despite the potential benefits that individuals can derive from an intimidating group reputation, group members have a hard time sustaining it. In situations in which the reputation of individuals cannot be disentangled from that of similar others, the inability to manage their aggregate or group reputation diminishes the probability that they receive favorable treatment. These results are consistent with the idea that the maintenance of a valuable group reputation is a social dilemma.

Our study also shows that a group reputation transmission mechanism by itself is not sufficient to motivate aggressive behavior. Relative to individual-specific reputations for aggression, the opportunity to build a group reputation for aggression reduces aggressive actions. This supports the view that reputation systems at the group level reduce aggression. Yet we also find evidence for the opposing perspective in the individual history treatment: an opportunity to build personal reputation increases aggression. Thus, we find support for both perspectives—that reputation systems can increase or

[10]See online Appendix B. Simulation Details: Simulation Experiments and Results.

decrease aggression—depending on the aggregation level of the reputation.

When we allow *B*s to punish their peers, *B*s increase their punishment strategies, and they are able to maintain an intimidating group reputation. So, a classic solution to cooperation problems, peer punishment (Fehr and Gächter 2000; Ostrom et al. 1992) also works for group reputations concerning aggression. It is possible that also other mechanisms like group identification (Balliet, Wu, and DeDreu 2014), rewards (Rand et al. 2009), or communication (Tavoni et al. 2011) would allow the social dilemma to be solved. Thus, peer punishment combined with group reputation makes group members behave more aggressively than without some form of internal discipline. Peer punishment also affects the dynamics of group reputation creation and maintenance. In contrast to individual reputation, the formation of group reputation is less strategically and instrumentally pursued: group reputation building does not display endgame effects since *B*s in the group history + peer punishment treatment use strong punishing strategies even in the final period of play. Unlike individual reputation, group reputation displays the typical inertia of social norms.

Although groups can create and maintain an intimidating reputation with peer punishment, this ultimately increases the amount of conflict that occurs. *A*s who refuse to pay are punished frequently and at a high intensity. Combined with the costs of peer punishment, this leads to the worst possible outcome, at least in the short term, for *A*s and *B*s. Although subjects take advantage of an internal discipline mechanism to create an intimidating group reputation, their earnings suffer. In our setup, the cost of creating and maintaining the group reputation offsets the benefits reaped from the shared resource. In the longer-term, however, this is not the case. Once an intimidating reputation is sufficiently established, conflict reduces, and it becomes a profitable strategy. Whether, more generally, reputation increases or decreases conflict depends on the aggregation of the reputation.

Taken together, our results showing that a valuable group reputation can be easily depleted but that protecting it is costly in the short term could have an important real-world implication. They suggest that groups have to navigate a narrow path to establish themselves. On the one side, they face the danger that if they fail to use an internal discipline mechanism, their reputation will suffer, negatively affecting how people outside the group treat the group members, reducing membership, and potentially increasing the probability of the group's dissolution. On the other side, if they commit to a costly mechanism like peer punishment, by establishing, for instance, social norms that prescribe to protect the good name of one's group, they could face the danger of poor outcomes for their group members and potentially the disintegration of the group itself. Such difficulties may vary depending on the stage of group emergence and stability. As suggested by some experimental evidence (Huck and Lünser 2010)—although findings are mixed (Capraro and Barcelo 2015)—when groups are small, group reputations may be naturally maintained. Once the group becomes larger and people start exploiting the reputation of their group, the group either avoids peer punishment mechanisms or is prepared to face its short-term costs in anticipation of future benefits. Following its establishment, the costs associated with maintaining an intimidating reputation may decrease again. Groups of many different types, gangs, criminal organizations, and proto-states, among others, may have to go through this high cost "hump" that could be a particularly sensitive stage at

which many groups might simply disappear.

Our results give insights into the fragility and stability of reputation at the group level. They help us to understand how reputational concerns can shape aggression and conflict, why it is difficult to maintain an intimidating group reputation, and why they often deteriorate. Understanding this may aid us in undermining groups that rely on intimidating reputations to succeed by intervening at the stage when they are most vulnerable. Conversely, for socially and economically beneficial groups, improving and strengthening their reputation at this crucial stage should help them create widespread benefits. Ultimately, the reputation that groups possess is likely to be a key determinant of whether their members are aggressive or peaceful.

## ACKNOWLEDGMENTS

This work is dedicated to the memory of Rosaria Conte. We thank Ozan Aksoy, Pat Barclay, Dominic Burbidge, Cristiano Castelfranchi, Diego Gambetta, Pawel Gola, Nan Zhang, and participants at the CeDeX workshop (University of Nottingham), LUISS Behavioural Economics workshop, and the conference Social Interaction and Society 2016 (ETH Zurich) and Social Dilemma 2017 (Taormina) for their comments and critiques.

## FUNDING

This work was supported by the project Global Dynamics of Extortion Racket Systems (GLODERS), grant agreement 315874, the Knut and Alice Wallenberg Grant "How Do Human Norms Form and Change?" (2016.0167), and the Horizon 2020 Framework Programme Project PROTON "Modelling the PRocesses leading to Organised crime and TerrOrist Networks" under Grant Agreement No.: 699824.

## ORCID iD

Aron Szekely https://orcid.org/0000-0001-5651-4711

## SUPPLEMENTAL MATERIAL

Additional supporting information may be found at https://journals.sagepub.com/doi/suppl/10.1177/0190272519882389.

## REFERENCES

Anderson, Craig A., and Brad J. Bushman. 2002. "Human Aggression." *Annual Review of Psychology* 53(1):27–51.

Balliet, Daniel, Junhui Wu, and Carsten K. W. De Dreu. 2014. "Ingroup Favoritism in Cooperation: A Meta-analysis." *Psychological Bulletin* 140(6):1556–81.

Barclay, Pat. 2015. "Reputation." Pp. 810–28 in *The Handbook of Evolutionary Psychology*. Vol. 2, edited by D. Buss. Hoboken, NJ: Wiley & Sons.

Barnett, Michael L. 2016. "Reputational Spillovers." Pp. 682–84 in *The Sage Encyclopedia of Corporate Reputations*, edited by C. Carroll. Thousand Oaks, CA: SAGE Publications.

Bell, Robert M., and Daniel F. McCaffrey. 2002. "Bias Reduction in Standard Errors for Linear Regression with Multi-stage Samples." *Survey Methodology* 28(2):169–81.

Benard, Stephen. 2013. "Reputation Systems, Aggression, and Deterrence in Social Interaction." *Social Science Research* 42(1):230–45.

Benard, Stephen. 2015. "The Value of Vengefulness: Reputational Incentives for Initiating versus Reciprocating Aggression." *Rationality and Society* 27(2):129–60.

Buss, A. H., and M. Perry. 1992. "The Aggression Questionnaire." *Journal of Personality and Social Psychology* 63(3):452–59.

Buss, D. M., and T. K. Shackelford. 1997. "Human Aggression in Evolutionary Psychological Perspective." *Clinical Psychology Review* 17(6):605–19.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90(3):414–27.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2):317–72.

Capraro, Valerio, and Hélène Barcelo. 2015. "Group Size Effect on Cooperation in One-Shot Social Dilemmas II: Curvilinear Effect." *PLOS ONE* 10(7):e0131419. doi:10.1371/journal.pone.0131419

Cialdini, R., Raymond R. Reno, and Carl A. Kallgren. 1990. "A Focus Theory of

Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58(6):1015–26.

De Dreu, Carsten K. W., Jörg Gross, Zsombor Méder, Michael Giffin, Eliska Prochazkova, Jonathan Krikeb, and Simon Columbus. 2016. "In-Group Defense, Out-Group Aggression, and Coordination Failures in Intergroup Conflict." *Proceedings of the National Academy of Sciences of the United States of America* 113(38):10524–29.

Duca, Stefano, and Heinrich H. Nax. 2018. "Groups and Scores: The Decline of Cooperation." *Journal of The Royal Society Interface* 15(144):20180158.

Engelmann, Jan M., Esther Herrmann, and Michael Tomasello. 2018. "Concern for Group Reputation Increases Prosociality in Young Children." *Psychological Science* 29(2):181–90.

Esarey, Justin, and Andrew Menger. 2019. "Practical and Effective Approaches to Dealing with Clustered Data." *Political Science Research and Methods* 7(3):541–59.

Faillo, Marco, Daniela Grieco, and Luca Zarri. 2013. "Legitimate Punishment, Feedback, and the Enforcement of Cooperation." *Games and Economic Behavior* 77(1):271–83.

Fearon, James D., and David D. Laitin. 1996. "Explaining Interethnic Cooperation." *The American Political Science Review* 90(4):715–35.

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *The American Economic Review* 90(4):980–94.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10(2): 171–78.

Fiske, Susan T., and Steven L. Neuberg. 1990. "A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation." Pp. 1–74 in *Advances in Experimental Social Psychology*. Vol. 23, edited by M. P. Zanna. San Diego, CA: Academic Press.

Gächter, Simon, Elke Renner, and Martin Sefton. 2008. "The Long-Run Benefits of Punishment." *Science* 322(5907):1510. doi:10.1126/science.1164744

Gambetta, Diego. 1993. *The Sicilian Mafia: The Business of Private Protection*. New ed. Cambridge, MA: Harvard University Press.

Gambetta, Diego. 2009. *Codes of the Underworld: How Criminals Communicate*. Princeton, NJ: Princeton University Press.

Gould, Roger V. 2003. *Collision of Wills: How Ambiguity about Social Rank Breeds Conflict*. Chicago, IL: University of Chicago Press.

Grann, David. 2004. "The Brand: How the Aryan Brotherhood Became the Most Murderous Prison Gang in America.' *The New Yorker*, February 16, 156–71.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association* 1(1):114–25.

Grimm, Volker, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback. 2010. "The ODD Protocol: A Review and First Update." *Ecological Modelling* 221(23):2760–68.

Griskevicius, Vladas, Joshua M. Tybur, Steven W. Gangestad, Elaine F. Perea, Jenessa R. Shapiro, and Douglas T. Kenrick. 2009. "Aggress to Impress: Hostility as an Evolved Context-Dependent Strategy." *Journal of Personality and Social Psychology* 96(5):980–94.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312(5770):108–11.

Harden, Jeffrey J. 2011. "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics & Policy Quarterly* 11(2):223–46.

Healy, Paul J. 2007. "Group Reputations, Stereotypes, and Cooperation in a Repeated Labor Market." *American Economic Review* 97(5):1751–73.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *The American Economic Review* 92(5):1644–55.

Huck, Steffen, and Gabriele K. Lünser. 2010. "Group Reputations: An Experimental Foray." *Journal of Economic Behavior & Organization* 73(2):153–57.

Johnstone, Rufus A., and Redouan Bshary. 2004. "Evolution of Spite through Indirect Reciprocity." *Proceedings of the Royal Society of London B: Biological Sciences* 271(1551):1917–22.

Jung, Yun Joo, John Kagel, and Dan Levin. 1994. "On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-Store Game." *RAND Journal of Economics* 25(1):72–93.

Kézdi, Gábor. 2004. "Robust Standard Error Estimation in Fixed-Effects Panel Models." *Hungarian Statistical Review* Special English Volume 9:95–116.

Kimbrough, Erik O., and Jared Rubin. 2015. "Sustaining Group Reputation." *Journal of Law, Economics, and Organization* 31(3):599–628.

King, Andrew A., Mew Lenox, and Michael L. Barnett. 2002. "Strategic Responses to the Reputation Commons Problem." Pp. 393–406 in *Organizations, Policy, and the Natural Environment: Institutional and Strategic Perspectives*, edited by A. Hoffman and M. Ventresca. Palo Alto, CA: Stanford University Press.

Kollock, Peter. 1998. "Social Dilemmas: The Anatomy of Cooperation." *Annual Review of Sociology* 24(1):183–214.

Kreps, David M., and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27(2):253–79.

Kurzban, Robert, and Daniel Houser. 2005. "Experiments Investigating Cooperative Types in Humans: A Complement to Evolutionary Theory and Simulations." *Proceedings of the National Academy of Sciences of the United States of America* 102(5):1803–807.

Leimar, Olof, and Peter Hammerstein. 2001. "Evolution of Cooperation through Indirect Reciprocity." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268(1468):745–53.

McElreath, Richard. 2003. "Reputation and the Evolution of Conflict." *Journal of Theoretical Biology* 220(3):345–57.

McIntosh, Craig, Elisabeth Sadoulet, Steven Buck, and Tomas Rosada. 2013. "Reputation in a Public Goods Game: Taking the Design of Credit Bureaus to the Lab." *Journal of Economic Behavior & Organization* 95:270–85.

Milgrom, Paul, and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27(2):280–312.

Nax, Heinrich H., Matjaž Perc, Attila Szolnoki, and Dirk Helbing. 2015. "Stability of Cooperation under Image Scoring in Group Interactions." *Scientific Reports* 5:12145. doi:10.1038/srep12145

Nisbett, Richard E. 1996. *Culture of Honor: The Psychology of Violence in the South*. Boulder, CO: Westview Press.

Nowak, Martin A., and Karl Sigmund. 1998a. "The Dynamics of Indirect Reciprocity." *Journal of Theoretical Biology* 194(4):561–74.

Nowak, Martin A., and Karl Sigmund. 1998b. "Evolution of Indirect Reciprocity by Image Scoring." *Nature* 393(6685):573–77.

Ohtsuki, Hisashi, and Yoh Iwasa. 2004. "How Should We Define Goodness?—Reputation Dynamics in Indirect Reciprocity." *Journal of Theoretical Biology* 231(1):107–20.

Orlando, Leoluca. 2001. *Fighting the Mafia and Renewing Sicilian Culture*. San Francisco: Encounter Books.

Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants with and without a Sword: Self-Governance Is Possible." *The American Political Science Review* 86(2):404–17.

Przepiorka, Wojtek, Charlotte Rutten, Vincent Buskens, and Aron Szekely. 2019. "How Dominance Hierarchies Emerge from Conflict: A Game Theoretic Model and Experimental Evidence." *Social Science Research*, first published on November 26, 2019. https://doi.org/10.1016/j.ssresearch.2019.102393.

Raihani, Nichola J., and Redouan Bshary. 2015. "The Reputation of Punishers." *Trends in Ecology & Evolution* 30(2):98–103.

Rand, David G., Anna Dreber, Tore Ellingsen, Drew Fudenberg, and Martin A. Nowak. 2009. "Positive Interactions Promote Public Cooperation." *Science* 325(5945):1272–75.

Rand, David G., and Martin A. Nowak. 2013. "Human Cooperation." *Trends in Cognitive Sciences* 17(8):413–25.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata Using Boottest." *The Stata Journal* 19(1):4–60.

Schelling, Thomas C. 1970. "The Diplomacy Of Violence." Pp. 64–84 in *Theories of Peace and Security: A Reader in Contemporary Strategic Thought*, edited by J. Garnett. London: Palgrave Macmillan UK.

Selten, R. 1967. "Die Strategiemethode Zur Erforschung Des Eingeschränkt Rationalen Verhaltens Im Rahmen Eines Oligopolexperiments." Pp. 136–68 in *Beiträge zur experimentellen Wirtschaftsforschung*. Tübingen: Mohr.

Selten, R. 1978. "The Chain Store Paradox." *Theory and Decision* 9(2):127–59.

Skarbek, David. 2014. *The Social Order of the Underworld: How Prison Gangs Govern the*

*American Penal System*. Oxford, UK: Oxford University Press.
Szekely, Aron, and Diego Gambetta. (forthcoming). "Does Information about Toughness Decrease Fighting? Experimental Evidence." *PLOS One*.
Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1(2):149–78.
Tavoni, Alessandro, Astrid Dannenberg, Giorgos Kallis, and Andreas Löschel. 2011. "Inequality, Communication, and the Avoidance of Disastrous Climate Change in a Public Goods Game." *Proceedings of the National Academy of Sciences* 108(29):11825–29.
Tennie, Claudio, Uta Frith, and Chris D. Frith. 2010. "Reputation Management in the Age of the World-Wide Web." *Trends in Cognitive Sciences* 14(11):482–88.
Tirole, Jean. 1996. "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)." *The Review of Economic Studies* 63(1):1–22.
Webb, Matthew D. 2014. "Reworking Wild Bootstrap Based Inference for Clustered Errors." Working Paper 1315, Queen's University, Department of Economics.
Wilensky, Uri. 1999. "NetLogo." Center for Connected Learning and Computer-Based Modeling, Northwestern University.

## BIOS

**Aron Szekely** is an assistant professor of sociology at the Collegio Carlo Alberto in Turin. He studies social mechanisms, including reputation, signaling, and social norms, and their effects in situations of cooperation and conflict. Much of his research uses experiments to study microlevel individual decision-making or agent-based models to explore emergent macrolevel social phenomena.

**Giulia Andrighetto** is a researcher at the Institute of Cognitive Sciences and Technologies of National Research Council of Italy in Rome and a Wallenberg fellow at Mälardalen University and at the Institute of Future Studies in Stockholm. Her research examines the nature and dynamics of social norms combining computational models of decision making with (large-scale) laboratory experiments.

**Nicolas Payette** is a senior research associate in agent-based modelling at the School of Geography and Environment of the University of Oxford. He is currently working on the POSEIDON project, modeling fisheries as agent-based socio-ecological systems. He is interested in how programming languages and tools shape the way we model the world and how we can build better tools to help us build better models.

**Luca Tummolini** is a researcher at the Institute of Cognitive Sciences and Technologies of National Research Council of Italy in Rome (Italy). His research interest is in social interaction and the cognitive mechanisms that enable humans to flexibly coordinate and collaborate with one another: from shared deliberation in small groups to conformity with population-wide regularities like conventions and social norms.