Cooperative behavior cascades in human social networks

James H. Fowler,1,2 and Nicholas A. Christakis3,4

*Political Science Department, University of California, San Diego, La Jolla, CA 92103; and 3Harvard Faculty of Arts and Sciences and 4Harvard Medical School, Boston, MA 02115

Edited* by Daniel Kahneman, Princeton University, Princeton, NJ, and approved January 25, 2010 (received for review November 12, 2009)

Theoretical models suggest that social networks influence the evolution of cooperation, but to date there have been few experimental studies. Observational data suggest that a wide variety of behaviors may spread in human social networks, but subjects in such studies can choose to befriend people with similar behaviors, posing difficulty for causal inference. Here, we exploit a seminal set of laboratory experiments that originally showed that voluntary costly punishment can help sustain cooperation. In these experiments, subjects were randomly assigned to a sequence of different groups to play a series of single-shot public goods games with strangers; this feature allowed us to draw networks of interactions to explore how cooperative and uncooperative behaviors spread from person to person. We show that, in both an ordinary public goods game and in a public goods game with punishment, focal individuals are influenced by fellow group members’ contribution behavior in future interactions with other individuals who were not a party to the initial interaction. Furthermore, this influence persists for multiple periods and spreads up to three degrees of separation (from person to person to person). The results suggest that each additional contribution a subject makes to the public good in the first period is tripled over the course of the experiment by other subjects who are directly or indirectly influenced to contribute more as a consequence. These results show experimentally that cooperative behavior cascades in human social networks.

behavioral economics | cooperation | public goods | social influence | pay-it-forward

S

cholars studying the evolution of cooperation in humans recently have turned their attention to the role of social networks in structuring human interactions (1–10). Interacting with others in large populations without structure greatly reduces the likelihood of cooperation (11), but in a fixed social network cooperation can evolve as a consequence of repeated interactions because of “social viscosity,” even in the absence of reputation effects or strategic complexity (1, 2). Different network structures can speed or slow selection, and, in some cases, network structures can determine completely the outcome of a frequency-dependent selection process (3). Heterogeneity in the interaction topology can improve prospects for cooperation (4), and adaptive selection of network ties by individuals on evolving graphs also can influence the evolution of behavioral types (5–7).

However, this theoretical literature has not explored whether cooperative behavior actually spreads across ties in human social networks, and recent experimental work has tended to focus on coordination rather than cooperation (12, 13). A growing number of observational studies suggest that diverse phenomena, including obesity (14), happiness (15), ideas (16), and many other behaviors and affective states (17–20), can spread from person to person. However, causal effects are difficult to extract from observational network studies because similarity in observed attributes among connected individuals may result from homophily, the tendency to connect to others who exhibit similar traits or behaviors (21). For example, past work has shown that people who engage in acts of altruism tend to befriend others who do the same (22). Causal effects also are difficult to extract from observational studies because associations between connected individuals might result from shared exposure to contextual factors that simultaneously engender various behaviors (including cooperation) in both parties.

Experimental studies can overcome these problems via random assignment of interactions in a controlled fashion. For example, a recent field experiment (23) showed that a door-to-door canvasser who encourages a resident to vote influences not only the person who answers the door but a second person in the household as well, even though there was no direct contact between the second person and the canvasser. Such studies experimentally showing person-to-person-to-person effects are rare, however. Prior experimental work on spreading processes in networks has focused primarily on direct person-to-person effects—for example, the dyadic spread of studiousness (24), positive moods (25, 26), and weight loss (27).

Not every interaction between connected individuals, of course, and not everything that spreads does so by the same mechanism. For example, although the spread of emotional states, smiling, or yawning may be rooted in fundamental neurophysiological processes (28), the spread of behaviors may arise from the spread of social norms or from other psychosocial processes, such as various types of innate mimicry (29). In the particular case of cooperation or altruistic behavior, it is well known that one person’s altruism toward another can elicit reciprocal altruism in repeated paired interactions (direct reciprocity) (30) and also in groups (31, 32). Indeed, many individuals are “conditional cooperators” who give more if others give more and who are influenced in their interactions with others by what the others are doing during the interaction (33, 34). It also is well known that reputation mechanisms that provide information about a person’s past behavior can help sustain cooperation (indirect reciprocity) (35). One study even showed that in a two-stage gift-exchange game, people who hypothetically receive a larger sum of money in the first stage tend to give more money to a third party in the second stage (36).

However, no work thus far has considered the possibility that, by various mechanisms including innate mimicry, witnessing cooperative or uncooperative behavior might promote changes in cooperative behavior that can be transmitted across social network ties in sequence to others who were not part of the original interaction. That is, quite distinct from prior work, we are concerned here with whether such behavior can create cascades of similar cooperative or uncooperative behavior in others, spreading from person to person to person, even when reputations are unknown and reciprocity is not possible. Such a cascade would suggest that social contagion also may play an important role in the evolution of cooperation.

Author contributions: J.H.F. and N.A.C. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

*This Direct Submission article had a prearranged editor.

†To whom correspondence should be addressed. E-mail: jhfowler@ucsd.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0913149107/DCSupplemental.
To study the spread of cooperative and uncooperative behavior in human social networks, we analyzed a set of previously published public goods game experiments (37). In these experiments, subjects were placed into groups of four, and each subject was given 20 money units (MUs). They then had to decide how many MUs (between 0 and 20) to keep or contribute to a group project. Each MU contributed to the group project would yield 0.4 MUs for each of the four group members. This set-up allows us to study cooperative behavior because each MU contributed is costly to the individual (0.4 MU − 1 MU = −0.6 MU) but beneficial to the group (4 × 0.4 MU − 1 MU = +0.6 MU). If each group member keeps all MUs privately, each will earn 20 MUs, and if each group member makes the maximum contribution of 20 MUs to the group project, each will earn 32 MUs. Despite this opportunity to improve group outcomes, each individual can save more by contributing less.

In the basic public goods game analyzed here, subjects made contribution decisions without initially seeing their fellow group members’ decisions, but all contributions were revealed to each group member at the conclusion of the game, along with payoff outcomes. In an alternate version, subjects played an identical public goods game, but, after viewing their fellow group members’ contributions, they were allowed to spend up to 10 MUs to “punish” each of the other group members. Each MU spent reduced the target’s income by 3 MUs. In both versions of the experiment, subjects participated in a total of six repetitions (“periods”) of the public goods game. This repetition allowed the researchers to show that contributions tend to decline in the basic public goods game and to increase in the public goods game with punishment (37).

Importantly, to distinguish the effect of punishment on cooperation from the effect of direct reciprocity (30), indirect reciprocity (35), reputation (38), and costly signaling (39), the research design enforced strict anonymity between subjects, and ego has no direct network connection to any of the alters (E, I, and M in period 2). The ego has a direct network connection to alters because she sees each of their contributions to the public good before proceeding to the next period. “Alter’s alters” are the individuals in the alters’ groups in the period before the previous period (F, G, H, J, L, N, O, and P in period 1). Note that the ego has no direct network connection to any of the alters’ alters and has not seen any of their contributions. However, the ego is indirectly connected to the alters’ alters by two degrees of separation via the alters (E, I, and M in period 2). The requirement that no two subjects be placed in the same group twice guarantees that we can draw a network like this for all 24 subjects in period 3.

Results
A summary of the raw data (Fig. 2) shows that, indeed, future contributions by focal individuals (“egos”) are significantly related to the amount contributed by each group member with whom the ego interacted in the previous period (“alters”) in both the basic public goods game and the public goods game with punishment. However, the raw relationship does not take into account constraints on the amount subjects can give (specifically, many subjects choose to contribute the minimum or maximum possible amounts). Nor does it take into account the fact that there are multiple observations for each ego and each alter within and between periods. We therefore use an interval regression technique with clustered standard errors (Materials and Methods) to estimate the size of the causal effect of one subject’s contribution behavior on another. We focus on the effect of each alter independently rather than on each group of alters because that focus allows us to take into account ceiling and floor effects that occur at the individual level (Materials and Methods). It also helps us conceptually identify the spillover effect of a single individual on the people with whom he or she is connected as well as the way this effect subsequently can spread to others in the interaction network.

The results show that alters significantly influence egos’ behavior, both directly and indirectly (Fig. 3). For each MU contributed by an alter, an ego contributes an additional 0.19 MUs [95% confidence interval (CI) 0.14–0.24, P < 0.0001] in the next period in the basic public goods game and an additional 0.18 MUs (95% CI 0.14–0.21, P < 0.0001) in the public goods game with punishment. Note that these results summarize the spread of both cooperative and uncooperative behavior: Alters who give less influence egos to give less, and alters who give more influence egos to give more.
Fig. 3. The total effect of alter’s contribution to the public good on ego’s contribution is significant and extends up to three degrees of separation. For each 1 MU contributed by alter, ego contributes an additional 0.19 MUs (0.18 MUs in the version with punishment) in the next period. For each 1 MU contributed by alter’s alter (a contribution ego did not observe), ego contributes an additional 0.07 MUs (0.05 MUs in the version with punishment) two periods later. For each 1 MU contributed by alter’s alter’s alter (three degrees of separation), ego contributes an additional 0.06 MUs in the public good game with punishment three periods later. Mediation analyses show that indirect total effects are mediated by the direct effect of alter on ego (SI Appendix). Alters are randomly assigned to egos, and they are assessed only at the minimum degree of separation at each point in time. Estimates are from interval regressions, controlling for multiple observations of the same ego, multiple observations of the same alter, the ego’s initial contribution in the period in which alter’s contribution was observed, and period fixed effects. Vertical bars show 95% confidence intervals.

Remarkably, even though egos do not observe the contributions of their alters’ alters (two degrees of separation), the alters’ alters also significantly affect egos’ contribution decisions. Each MU contributed by an alter’s alter increases an ego’s contributions by an additional 0.07 MUs (95% CI, 0.03–0.10, P = 0.0005) two periods later in the public goods game and by an additional 0.05 MUs (95% CI, 0.03–0.08, P = 0.0001) in the public goods game with punishment. Furthermore, in the public goods game with punishment, we find evidence that cooperative behavior spreads one degree farther, up to three degrees of separation. Each MU contributed by an alter’s alter increases an ego’s contribution by 0.06 MUs (95% CI, 0.02–0.09, P = 0.001).

It is important to note that the results for two and three degrees of separation represent total effects. They reflect the indirect traces of an individual’s actions on others via a chain of direct pairwise effects. For example, if subjects tend to give 25% more for each MU received, then when an alter’s alter gives 16 additional MUs to an alter, it will cause that alter to give 4 additional MUs to ego, who will give 1 additional MU to the next subject. The total effect of an alter’s alter on ego therefore is the product of the effect of the alter’s alter on the alter and the alter’s effect on the ego. Indeed, as expected given the experimental set-up, a Sobel test (40) shows that the total effect of the alter’s alter on ego is mediated by the indirect effect that spreads from the alter’s alter to the alter to the ego (SI Appendix).

As a separate matter, many of these direct and indirect effects endure, influencing the ego’s behavior long after the initial period of influence (Fig. 4). For example, F may influence E to give more, not only in the following period when E interacts with A, I, and M but also in the period after that when E interacts with K, P, and T (Fig. 1) and in other future periods as well. In the basic public goods game, each MU contributed by the alter causes ego to contribute an additional 0.15 MUs (95% CI 0.09–0.21, P < 0.0001) two periods later, 0.08 MUs (95% CI 0.00–0.16, P = 0.04) three periods later, 0.17 MUs (95% CI 0.07–0.27, P = 0.001) four periods later, and 0.17 MUs (95% CI 0.00–0.33, P = 0.05) five periods later. In the public goods game with punishment, each MU contributed by the alter causes the ego to contribute an additional 0.15 MUs (95% CI 0.10–0.19, P < 0.0001) two periods later, 0.11 MUs (95% CI 0.05–0.17, P = 0.0001) three periods later, 0.14 MUs (95% CI 0.07–0.21, P = 0.0001) four periods later, and 0.14 MUs (95% CI 0.03–0.25, P = 0.02) five periods later. The effect of the decision by the alter’s alter (and not just the effect of alter’s decision) also persists in the basic public goods game after four periods (0.07 MUs; 95% CI, 0.01–0.13, P = 0.02) and in the public goods game with punishment after three periods (0.14 MUs; 95% CI 0.03–0.25, P = 0.007).

To measure the overall size of these cooperative behavior cascades (a hypothetical example is given in Fig. 5), we focus only on effects with p values less than 0.05 in Figs. 3 and 4. In the basic public goods game, if a subject increased his/her contribution by an additional MU in period 1, that increase would directly cause the three other subjects in his/her group to increase their total contributions by 1.8 MUs (95% CI 1.3–2.3) over the next four periods. It also would indirectly cause nine other subjects to increase their total contributions by 1.2 MUs (95% CI 0.9–1.5) in periods 3 and 5, for an overall increase of 3 MUs (95% CI 2.4–3.6). In the public goods game with punishment, the direct increase in contributions would be 2.1 MUs (95% CI 1.6–2.7) over the next five periods, and the indirect increase would be 0.9 MUs (95% CI 0.7–1.0) in periods 3 and 4, also totaling 3 MUs (95% CI 2.4–3.6). The reverse is also true, and each reduction in one’s contribution in the first period can generate a cascade of uncooperative behavior through parts of the network subsequently. This exercise suggests that, overall, over the course of each version of the public goods game, the spread of cooperative behavior through the network approximately triples each additional contribution made in the first period, at least in a network of this particular, experimentally controlled structure.

Finally, we note that in the public goods game with punishment, alters’ alters also may influence the ego via their punishment behavior. F might punish E, causing E to contribute more after
one period, an increase that causes A to contribute more after two periods. To test this hypothesis, we regressed ego’s contribution on the punishment alters had received from others two periods previously (SI Appendix). Indeed, we find that punishment can spurt cooperative cascades as well. Each punishment point that alter’s alter gives to alter increases ego’s contribution two rounds later by 0.13 MUs (95% CI 0.02–0.23, \( P = 0.02 \)). However, the effect does not appear to spread any further: The relationship between alter’s alter’s alter’s punishment behavior and ego’s contribution three periods later is insignificant (\( P = 0.25 \)). We also failed to find any evidence of spreading punishment behavior per se; the association between punishment received in the previous round and punishment given in the current round was not significant (\( P = 0.83 \); SI Appendix).

Discussion

It is often supposed that individuals in experiments like the one described here selfishly seek to maximize their own payoffs. However, it is worth reiterating that most of our subjects violated this supposition: Because all interactions are single-shot, the equilibrium prediction is to contribute nothing and to pay nothing to punish noncontributors, but the subjects did not follow this pattern (37). One mechanism that may underlie such deviations from “rational” action appears to be mimicry: When subjects copy the cooperative behavior of others with whom they interact, their doing so causes them to deviate even more from rational self-interest and may help reinforce this behavior.

Observational studies suggest that behaviors, knowledge, and emotions spread between people with personal social ties (14–20). Of course, people can be influenced by strangers too—for example, in catching a germ, following a worn path on a field, imitating a smile, adopting a fashion, or responding to road rage. In this experiment, relationships were anonymous, and contact was not sustained. Nevertheless, there was real interaction, and people observed each others’ behavior in the setting of a game that they cared about. A consistent explanation for both the experimental investigations and the observational studies is that people mimic the behavior they observe, and this mimicking can cause behaviors to spread from person to person. If anything, it seems likely that people who are willing to copy strangers’ behavior may be even more likely to copy similar behavior observed in friends in real-world settings.

Although an act of punishment, like a contribution, can initiate a cooperative cascade, we found minimal differences in the spread of influence between the basic public goods game and the public goods game with punishment. This observation suggests that the existence of punishment does not fundamentally change network dynamics: punishment may not enhance or facilitate the spread of cooperation per se. The reason may be that ego observes only whether alter is cooperating, not the motivations that alter has for behaving in a particular way nor alter’s prior history of interactions that may prompt a particular behavior. However, punishment clearly has a direct effect on contributions, and the network process we describe may help magnify the indirect effect of punishment. Thus, behavioral cascades may be a crucial part of an explanation for how small changes in human institutions (such as informal norms or formal rules about punishment) can yield large changes in a group’s behavior. This effect is all of the more remarkable because we found no evidence that punishment itself spreads.

This multiplier effect also suggests that behavioral imitation and interpersonal spread may be an important factor in the evolution of cooperation in humans. For example, cascades of cooperative (or noncooperative) behavior can promote coordination on a particular strategy, possibly decreasing within-group variance. At the same time, the path-dependent nature of this process within each group may tend to increase between-group variance. In a population with structured groups, both of these effects work in favor of the emergence of altruism (41). Cascades also may help mitigate the negative effect of group size on cooperation (11, 42) because they reduce the number of independent entities in a population, effectively increasing the size of groups in which public goods can be maintained via self-interest. Evolutionary game theorists therefore should consider the possibility that behavioral imitation itself may have coevolved with both cooperation and the emergence of social networks.

Such models also might help explain the influence of genetic variation on social network structure (43). Egocentric network characteristics such as network centrality can make some individuals more susceptible than others to contagions with negative outcomes for fitness (e.g., germs, misinformation, and violence). However, the results here suggest that one fundamental justification for the existence of elaborate social ties in the form of social networks may be that these ties may allow humans to benefit from the actions of widely distributed others and also may allow humans to spread beneficial strategies widely enough to benefit others on whom they depend. Genetic variation in social network position suggests that networks may influence reproduction or survival via a frequency-dependent selection process or rapid (relative to evolution) environmental variation; in either case, given that cooperation itself also appears to have a genetic basis (44), it makes sense to think about how networks may play a role in evolution.

Finally, these results provide experimental support for a conjecture about human social networks. To explain a variety of observational data showing that behavior in social networks is correlated up to three degrees of separation but rarely extends farther, a “three degrees of influence rule” (45) has been described which suggests that (i) behavior can spread from person to person to person via a diverse set of mechanisms, subject to certain constraints, and (ii) as a result, each person in a network can influence dozens or even hundreds of people, some of whom he or she does not know and has not met. The present results show experimentally that such cascades can occur in a controlled environment where people are making decisions about giving to others. Other researchers have shown that giving behavior can spread from person to person in
natural settings, whether in workplace donations to charity (46) or the decision to donate organs (47). However, whether such “pay it forward” behavior spreads more widely from person to person in natural human networks remains an open question.

### Materials and Methods

The procedures for implementing the public goods game experiments for the 240 subjects analyzed in this report have been described elsewhere (37). Fig. 1 illustrates that the requirement in these experiments that no two subjects meet each other twice ensures that any ego who is directly connected to an alter (one degree of separation) cannot also be connected indirectly to the same subject by two degrees of separation (an alter’s alter). It also ensures there are no redundant paths at one and two degrees of separation, and no subject can be connected to herself by two degrees. However, there are four and five degrees of separation, such combinations are possible, so we remove from the analysis all self-connections and all redundant paths, and we keep just one observation from among those with the shortest path length (smallest degree of separation). For example, if at period t subject B is subject A’s alter’s alter (three degrees) via two paths and also is his alter’s alter’s alter (four degrees) via five paths, for the purpose of analysis, we assign a single, randomly chosen observation for this pair to the data in which subject B’s contribution behavior depends on subject A’s behavior at t = 3.

To analyze ego contribution behavior, we use interval regression (also known as “Tobit” regression), a method typically used in the literature on public goods games (31). This type of regression model treats responses at the minimum (0 MUs) and maximum (20 MUs) as censored. Past work has shown that applying ordinary least-squares regression to data like these yields inconsistent results (slope coefficients are biased toward zero, and intercepts are biased away from zero), whereas interval regression yields consistent results (48). However, the coefficients in interval regression apply to the latent outcome variable (what subjects would do if they were not constrained) rather than the observed outcome variable (what subjects actually do).

To estimate the influence of one subject’s contribution on another subject’s contribution, we include in these regression models the alter’s contribution in the period t − s, where s is the degree of separation (alter: s = 1, alter’s alter: s = 2, and so on). To control for serial correlation, we also include ego’s contribution in the period t − s; alternative specifications that add additional lags (SI Appendix) generate identical results. To control for period effects, we include an indicator variable for all but one of the periods in which ego contributions were observed. To control for multiple observations of the ego and the alter, we use Huber–White sandwich errors that account for errors clustered on each ego and each alter. As a robustness check, we examined whether the effect of alter on ego varies depending on whether alter’s contribution is high or low (it does not). We also included the other two group members’ contributions as a control variable; this inclusion did not change the results.

We further replicated all results by treating the group contribution, rather than the alter’s contribution, as the unit of analysis. When we analyzed the effect of others’ contributions on alter’s influence over ego, we found that alter’s influence remained significant under all conditions, suggesting that analysis at the individual level rather than the group level is appropriate (SI Appendix).

We emphasize that all activity in the experiments was completely anonymous. Group composition changed randomly every period, so that no one played with the same person more than once. The subjects were ignorant of other players’ experimental history; neither past payoffs nor past decisions were known. Different group composition in each period and the absence of any history of play ensured that subjects could neither develop reputations nor target other subjects for revenge.

### ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grants P01 AG031093 and P30 AG044420 from the National Institute on Aging, by a Pioneer Grant from the Robert Wood Johnson Foundation, and by the University of Notre Dame and the John Templeton Foundation as part of the Science of Generosity Initiative.