Binding promises and cooperation among strangers[*]

Gabriele Camera[1]
University of Basel
and
Chapman University

Marco Casari
University of Bologna

Maria Bigoni
University of Bologna

27 November 2012

Abstract

In an experiment, a group of strangers was randomly divided in pairs to play a prisoners' dilemma; this process was indefinitely repeated. Cooperation did not increase when subjects could send public messages amounting to binding promises of future play.

Keywords: coordination, cheap-talk, deception, repeated game, social norms.

JEL codes: C90, C70, D80

[*] We thank an anonymous referee for helpful suggestions. Camera thanks the NSF for research support through the grant CCF-1101627. Financial support for the experiments was also partially provided by a grant from Purdue's CIBER and Einaudi Institute for Economics and Finance.
[1] Corresponding author: Gabriele Camera, Faculty of Business and Economics, Department of Macroeconomics, University of Basel, Peter Merian-Weg 6, CH - 4002 Basel, Switzerland. Tel: +41 (0)61 267 2458. e-mail: gabriele.camera@unibas.ch

1. Introduction

Non-binding pre-play communication has been shown to promote welfare in some experiments on social dilemmas (Ostrom et al., 1992) but not in others (Wilson and Sell, 1997, Duffy and Feltovich, 2006). Such ineffectiveness may stem from shortcomings in communication, such as deception or lack of credibility of messages (Aumann 1990, Farrell and Rabin, 1996). Indeed, Wilson and Sell (1997) explicitly conjecture that communication would be more effective if promises could be binding. This paper tests such hypothesis by considering two questions: Would subjects in an experiment take steps to address such communication shortcomings? And if so, would such enhanced communication foster trust and be more effective?

We consider interactions among strangers, where there truly is a need of communication and coordination to achieve cooperation. In a *Baseline* treatment, subjects were assigned to a group of four members; in every period they were randomly divided into pairs to play a prisoners' dilemma (PD). Subjects could observe average group cooperation but not individual histories. Group interaction was *indefinitely* repeated; hence, multiple equilibria were possible, including the efficient outcome. In the *Pledge* treatment, subjects could send a message to the entire group, before playing the PD game; they could also *audit* messages, i.e., sanction all those group members whose actions and messages differed. Messages could thus represent pledges, i.e., *binding* promises of future play.

The data reveal that subjects sought to solve deception problems in communication: auditing took place in almost half of the periods, while pledges were rarely breached. However, cooperative behavior in *Pledge* did not increase relative to *Baseline* but rather declined. Pledging cooperation made cooperators easy prey for defectors which led to a collapse of communication and trust. This suggests that the possibility of sending messages amounting to binding promises

is *not* sufficient to make communication effective in sustaining cooperation and, in fact, can backfire.

2. Experimental Design

The stage game was the PD in Table 1, where *Y=cooperate* and *Z=defect*.

	Y	*Z*
Y	25, 25	5, 30
Z	30, 5	10, 10

Table 1: Payoffs

Each session involved twenty subjects and five supergames consisting of an indefinite sequence of periods achieved by a random continuation rule (Roth and Murnighan, 1978). At the end of each period, the computer drew an integer between 1 and 100 from a uniform distribution, and showed it to all participants. The supergame terminated simultaneously for all with a draw of 96 or higher. Hence, in each period the supergame was expected to continue for 19 additional periods.

We built twenty-five economies in each session by creating five groups per supergame. In each period subjects interacted only with members of their group, in randomly formed pairs. Subjects could not identify opponents, hence could not use reputational mechanisms; they could observe their payoff, the group's average cooperation rate, but not individual histories. No two participants ever interacted together for more than one supergame.

This design admits multiple equilibria, ranging from full defection to the efficient outcome. Given that monitoring was public, the efficient outcome can be sustained as a sequential equilibrium for all discount factors above 0.25 when everyone adopts the following strategy:

start cooperating and keep cooperating unless someone defects, in which case defect forever.[1] This follows from the Folk Theorem-type results in Kandori (1992) and Ellison (1994).

Stage game, continuation probability, and matching protocols were identical across treatments. In the *Pledge* treatment, there was a communication stage before period one, and then every four periods. Communication was structured, public, costless, and anonymous. Subjects could send all group members a three-part message, consisting of a suggestion Y, Z, or "not sure"(i) for the subject, (i) for her anonymous match, and (iii) for everyone else. Here, we focus on part (i), which could be used to *signal intentions* of play. We identify Y and Z as *explicit* messages because the message-action mapping is clear, and "not sure" as *neutral*.

Messages could be transformed into something that amounted to binding promises: subjects could pay one point to "audit." If at least one subject audited, ten points were deducted from all group members (auditor included) who sent an explicit message in the last communication round, but did not choose the corresponding action in the period; selecting "not sure" protected from sanctions. The number of sanctioned subjects was made public.

We recruited 80 subjects through e-mail and in-class-announcements. Sessions were run at Purdue University. No eye contact was possible among subjects. Instructions were read aloud with copies on all desks. Average earnings were $27.28 (no show-up fee). Sessions comprised on average 102 periods, and lasted about 3 hours, including instruction reading and a quiz.[2]

In comparison to the *Baseline*, communication in *Pledge*: (1) maintains the multiplicity of equilibria, because subjects could simply ignore messages, and (2) maintains anonymity because individual histories, approval or disapproval, or explicit threats could not be conveyed. Hence,

---

[1]For a proof and further details see the anonymous public monitoring treatment in Camera and Casari (2009).

[2] Session dates: 27.4.05, 1.9.05 (*Baseline*, also analyzed in Camera and Casari, 2009), 5.4.07, 11.4.07 (*Pledge*). Exchange rate: $.13 for every 10 points.

even if actions *Y* and *Z* are both part of a sequential equilibrium, messages are not necessarily credible (Farrell and Rabin, 1996).[3]

Auditing could eliminate deception because the sender of an explicit message had incentives to behave accordingly. Yet, auditing cannot sustain the efficient outcome *per se* because it is costly, and in a cooperative equilibrium there is no reason to audit.

Three elements differentiate this design from previous studies. First, there exists a continuum of equilibria, including the Pareto efficient outcome, while in experimental studies of deception in finitely repeated social dilemmas defection is the *unique* equilibrium. Second, subjects could send explicit or neutral messages, which helps in detecting and quantifying deception. Third, auditing allows us to assess whether or not a mechanism to remove deception emerges endogenously.

3. Results

Result 1: *A mechanism to enhance the credibility of communication emerged endogenously.*

Auditing took place in 44.6% of the periods, which enhanced the credibility of messages and transformed explicit messages into credible pledges. Messages largely reflected truthful intentions–pledges were breached only 5.6% of the times–and were perceived as truthful: subjects cooperated more when they saw more Y messages. Those who signaled a cooperative intention were more likely to cooperate even when no one else manifested a similar intention, but all subjects tended to cooperate more when observing more Y messages (Table 4).

[3] E.g., a subject who is playing "always defect" may send a message *Y* to induce cooperation. That is to say, structured messages are not necessarily self-signaling and self-committing and can be outright deceptive.

Message sent by a subject	Cooperative messages (Y)sent by others in the group 0	1	2 or 3
Not sure	20.6% *(680)*	38.6% *(853)*	39.8% *(420)*
Y	71.4% *(301)*	81.9% *(386)*	96.9% *(288)*
Z	2.9% *(68)*	4.0% *(50)*	5.9% *(34)*

Table 4: Frequency of cooperation in Pledge

Notes: *One observation per subject, per period (total number in parentheses).*

A probit regression (supporting materials) confirms that explicit messages amounted to binding promises and were used to facilitate cooperation.

One would expect that the possibility to sanction untruthful messages would foster cooperation, not only because messages amounted to binding promises, but also given previous results on costly personal punishment. We find that structured, repeated communication opportunities did not foster cooperation even if it *is* part of a Nash equilibrium.

	Avg. cooperation rate (all periods)	Avg. cooperation rate (period 1)	Avg. coordination on cooperation
Baseline	58.6%	70.5%	28.6%
Pledge	57.7%	65.0%	17.4%

Table 2: Cooperation by treatment

Result 2: *In the Pledge treatment, average cooperation did not increase, while average coordination on cooperation fell.*

The *cooperation rate* is the fraction of Y actions in the group, in a supergame. *Coordination on cooperation* is the fraction of periods in a supergame in which everyone in a group cooperated.[4] Considering all periods, average cooperation rates in *Baseline* are not significantly lower than *Pledge*. A statistical test does not reject the null hypothesis that observations from the two treatments are drawn from the same population (Mann-Whitney tests, n1=n2=50, p-value>0.10).[5] The difference in period 1 cooperation across treatments is not statistically significant (Mann-Whitney test, n1=n2=50, p-value>0.10). Only in the first supergame, cooperation rates are lower in *Baseline* than in *Pledge* (Tobit regression, all periods) but this result is not robust to including all supergames. However, the availability of structured communication significantly reduces *coordination on cooperation* with respect to the *Baseline* (Mann-Whitney pairwise tests, n1=n2=50, p-value=0.076). A Tobit regression confirms this result (supporting materials).

Communication was ineffective in facilitating efficient Nash play. The reason was not deception (Result 1) and cannot be entirely attributed to limitations in the message space. Instead, the transformation of cheap talk into something akin to binding messages created a holdup problem, which ultimately reduced the usefulness of communication as a way to coordinate on efficient Nash play. The widespread use of auditing did not aim at building trust, but at extracting rents from cooperators, which reduced trust.

Result 3: *Defectors audited more often than cooperators.*

---

[4]Unless otherwise specified, the unit of observation is the group in a supergame (50 observations per treatment).

[5]The results of statistical tests rely on assuming that all observations are independent. All tests are two-sided.

Auditing exposed those who pledged cooperation to the risk of becoming prey of opportunistic subjects, who sent neutral messages, defected and *also* audited to ensure compliance. Cooperators who sent a Y message made 33.9% of all auditing requests; defectors who sent a neutral message made 46.2% of all auditing requests (Table 7). Through auditing, defectors could "hold up" cooperators for the next four periods.

	Message sent			**Total**	*N*
Action	Not sure	Y	Z		
Y	11.1%	33.9%	0.2%	**45.3**%	*191*
Z	46.2%	4.0%	4.5%	**54.7**%	*231*
Total	57.3%	37.9%	4.7%	**100.0**%	*422*
N	*242*	*160*	*20*	*422*	

Table 7: Frequency of auditing

Notes: *One observation per subject per period. The 422 auditing requests encompass 100% of the observations.*

4. Conclusions

The experiment shows that subjects did take steps to make communication truthful. Yet, contrary to Wilson and Sell (1997)'s conjecture, the ability to send messages that amounted to binding promises did not improve the effectiveness of communication due to an exploitative use by defectors. Defectors "held up" those who pledged cooperation, which ultimately lowered the value of communicating cooperative intentions as a mean to achieve a cooperative outcome. Such result might not survive if we removed the option of sending a neutral message. On the one hand, this would eliminate the possibility to hide the intention to hold-up cooperators by sending a non-committal message. And on the other, it would essentially force *every* individual to

commit to either defection or cooperation. Therefore, it is hard to tell whether cooperation would increase or decrease in this case. Future research can shed some light on this.

References

Aumann, R. (1990), "Nash Equilibria are not Self-enforcing." in Economic Decision Making: Games, Econometrics and Optimization (J.J. Gabszewicz; J.F. Richard and L.A. Wolsey Eds.) Elsevier, Amsterdam, New York.

Camera, G., and M. Casari (2009), "Cooperation among strangers under the shadow of the future." *American Economic Review*, 99(3), 979–1005.

Duffy, J. and N. Feltovich (2006), "Words, Deeds and Lies: Strategic Behavior in Games with Multiple Signals." *Review of Economic Studies*, 73, 669-688.

Ellison, G. (1994), "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching." *Review of Economic Studies,* 61, 567-588.

Farrell, J. and M. Rabin (1996), "Cheap talk." *Journal of Econ. Perspectives* 10, 103–118.

Kandori, M. (1992), "Social Norms and Community Enforcement." *Review of Economic Studies,* 59, 63-80.

Ostrom, E., J. Walker, and R. Gardner (1992), "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review*, 86, 404-417

Roth, A. E., and K. Murnighan (1978), "Equilibrium Behavior and Repeated Play of The Prisoner's Dilemma." *Journal of Mathematical Psychology,* 17: 189-198.

Wilson, R. K., and J. Sell (1997), "Liar, Liar... Cheap Talk and Reputation in Repeated Public Goods Settings." *Journal of Conflict Resolution*, 41 (5), 695-717.