Journal of Economic Psychology 33 (2012) 642–658

Contents lists available at SciVerse ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

The power of words: A model of honesty and fairness

Raúl López-Pérez *

Department of Economic Analysis, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 9 January 2011
Received in revised form 15 November 2011
Accepted 6 December 2011
Available online 13 December 2011

JEL classification:
C72
D01
D62
D64
Z13

PsycINFO classification:
3020 Group & Interpersonal Processes

Keywords:
Communication
Cooperation
Fairness
Heterogeneity
Honesty
Reciprocity
Social norms

ABSTRACT

We develop a game-theoretical model of honesty and fairness to study cooperation in social dilemma games with communication. It is based on two key intuitions. First, players suffer a utility cost if they break norms of honesty and fairness, and this cost is highest if most others comply with the norm. Second, people are heterogeneous with regard to their concern for norms. We show that a model based on honesty norms alone cannot explain why pre-play communication fosters cooperation in simultaneous social dilemmas. In contrast, the model based on norms of honesty and fairness can. We also illustrate other predictions of the model, offering experimental evidence in line with them – e.g., the effect of communication on cooperation depends on how many players communicate, and whether the social dilemma is played simultaneously or sequentially. In addition, ideas for new experiments are suggested.

1. Introduction

Communication often has a very substantial effect on cooperation, something true even if the communicators are anonymous subjects playing one-shot games (for early studies pointing this, see Dawes, McTavish, & Shaklee, 1977; Dawes, 1980). In a meta-analysis of social dilemma experiments conducted from 1958 to 1992, for instance, Sally (1995) concluded that non-binding communication raises cooperation by approximately 30%; a more recent meta-analysis by Balliet (2010) has found as well a large positive effect.[1]

In this paper we offer one possible explanation for this effect. Our model is based on the idea that people care about social norms (Becker, 1996; Bicchieri, 2002; Elster, 1989) in a *conditional/reciprocal* manner – intuitively, they feel painful emotions like shame when they transgress internalized norms that others respect. Crucially, it also assumes that players are

* Tel.: +34 91 497 6801.
E-mail address: raul.lopez@uam.es

[1] See also Bicchieri (2002) for a survey of some evidence on communication in public good games and social dilemmas, and Ellingsen and Johannesson (2004) for a review of some related psychological literature.

0167-4870/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.joep.2011.12.004

heterogeneous concerning the norms that they care about.[2] More precisely, we distinguish between three types of players: (i) selfish types who do not care at all about norms, (ii) honest (H) types who are motivated by a norm of honesty, and (iii) EH types who are motivated by a norm of both distributive justice and honesty – this norm requires players to seek an efficient and egalitarian (E) outcome and also to be honest, which explains the acronym EH.

A key message from our model is that the interaction between the three types is crucial to understand why communication fosters cooperation in social dilemmas. In effect, our first result shows that a model based on honesty norms alone cannot account for this phenomenon, at least in simultaneous social dilemmas. In short, honesty/lie-aversion is not enough to deliver a communication effect. In contrast, our '3-types' hypothesis can account for it. To get some of the intuition in this respect, consider first the following three behaviors, well-observed in numerous studies (Sally, 1995): (a) some people cooperate in social dilemmas even if they cannot communicate, (b) communication increases cooperation (which suggests that some people cooperate *only* if they can communicate), and (c) some people never cooperate in social dilemmas, even if pre-play communication is available. These behaviors correspond in our model to the EH, honest, and selfish types, respectively. In effect, EH types cooperate if they expect their co-players to cooperate as well – that is, they are conditional cooperators – and their presence explains why some people cooperate even if they cannot communicate. In turn, communication increases cooperation because it allows honest types to make promises and hence commit themselves to cooperate. Is that ever an optimal strategy? Yes, if they believe that the promisee is an EH player, that is, the type of person who cooperates conditionally and hence would defect if the sender announced defection. Finally, selfish agents are necessary to explain why cooperation sometimes fails to happen, even if communication is available.

We also hint that the effect of communication subtly depends on a number of variables including: (i) the message content, (ii) the number of message senders, (iii) the order in which players communicate or make choices later (e.g., sequentially or simultaneously), (iv) the possibility for the players to keep silent, (v) the expected price of telling the truth, or (vi) prior history. These predictions are conveyed by means of simple examples, which might hopefully inspire further theoretical research. Further, we report some existing experimental evidence consistent with those predictions, although it seems to us that more data is required to better understand motivations like honesty and the effect of communication. We hope in this respect that a formal, detailed model will help suggest ideas for future experiments.

This paper contributes to two expanding research fields. On one hand, several models of fairness and reciprocity have emerged in the last decades (we review them in Section 5), providing insights on why people cooperate. For instance, these models predict costly conditional cooperation in public good games, trust games, and other social dilemmas, a phenomenon observed in many experiments – see Chaudhuri (2011) for a survey. On the other hand, a number of papers have recently formalized the idea that people are lie-averse or honest, motivated by evidence that people often tell the truth even when it is contrary to their material interest (Belot, Bhaskar, & van de Ven, 2010; Erat & Gneezy, 2009; Fischbacher & Heusi, 2008; Gneezy, 2005). Our model incorporates ideas from both fields, not only with the aim of better understanding how communication affects behavior, but also to improve our knowledge on the above mentioned motivations, which could help to account for behavior related to accounting, auditing, insurance, job interviews, labor negotiations, regulatory hearings, and tax compliance, to provide a few examples.

The paper is organized as follows. To better appreciate the interaction between the three types of players, the next section introduces a simple model in which there are only two types: honest and selfish. We prove that this model is not consistent with the robust phenomenon cited before, by which pre-play communication fosters cooperation in social dilemmas (Sally, 1995). For this reason, we extend the model in Section 3, introducing the EH types. Using this extended model, Section 4 presents our main results. Section 5 compares our account with alternative explanations, and Section 6 concludes by mentioning some possible extensions of the model together with some potential experiments where it could be put to test.

2. A model of honesty

Consider any n-player, extensive form game of perfect recall. Let $N = \{1,\ldots,n\}$ denote the set of players, z a terminal node, h an information set, and $A(h)$ the set of available actions or moves at h. Further, let $u_i(z)$ denote player i's utility payoff at z, and $x_i(z)$ player i's monetary payoff at z.

At some h, a player may be given the opportunity to communicate – i.e., send a message to other players. Given our focus on pre-play communication, we assume that players only exchange messages about future moves. More precisely, the messages available at h are announcements of actions $[\mathbf{a}(h_1),\ldots,\mathbf{a}(h_k)]$, where $h_1,\ldots,h_k$ are information sets *following* h and $\mathbf{a}(h_j) \in A(h_j)$ for any $j \in \{1,\ldots,k\}$. A message announcing action $\mathbf{a} \in A(h_j)$ means 'the mover at h_j would choose $\mathbf{a}$ if she had to move at h'_j. Observe that players are allowed to announce their own future moves and/or other players' potential moves;

[2] The idea that agents are heterogeneous in their pro-sociality is consistent with a large body of experimental data. To start, the evidence from social dilemmas without pre-play communication (Brandts & Schram, 2001; Croson, 2000, 2007; Fischbacher & Gächter, 2010; Fischbacher, Gächter, & Fehr, 2001; Neugebauer, Perote, Schmidt, & Loos,2009) points out that some subjects are conditional cooperators who cooperate if they expect others to cooperate as well, while remaining subjects tend not to cooperate. In addition, recent lab evidence shows as well that subjects differ in their propensity to tell the truth (Gneezy, 2005: Hurkens & Kartik, 2009).