Journal of Economic Psychology 57 (2016) 153–171

Contents lists available at ScienceDirect

Journal of Economic Psychology

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

Carry a big stick, or no stick at all
Punishment and endowment heterogeneity in the trust game

Vicente Calabuig [a], Enrique Fatas [b,*], Gonzalo Olcina [a], Ismael Rodriguez-Lara [c]

[a] ERICES, Universidad de Valencia, Spain
[b] University of East Anglia, United Kingdom
[c] Middlesex University London, United Kingdom

ARTICLE INFO

Article history:
Received 11 November 2014
Received in revised form 14 September 2016
Accepted 21 September 2016
Available online 13 October 2016

JEL Codes:
C91
D02
D03
D69

Keywords:
Trust game
Endowment heterogeneity
Punishment
Deterrence hypothesis
Crowding-out
Intrinsic and extrinsic motivation
Experimental economics

ABSTRACT

We investigate the effect of costly punishment in a trust game with endowment heterogeneity. Our findings indicate that the difference between the investor and the allocator's initial endowments determines the effect of punishment on trust and trustworthiness. Punishment fosters trust only when the investor is wealthier than the allocator. Otherwise, punishment fails to promote trusting behavior. As for trustworthiness, the effect is just the opposite. The higher the difference between the investor and the allocator's initial endowments, the less willing allocators are to pay back. We discuss the consistency of our findings with social preference models (like inequality aversion, reciprocity), the capacity of punishment (i.e., the deterrence hypothesis) and hidden costs of punishment (i.e., models of intrinsic and extrinsic motivation). Our results are hardly coherent with the first two (inequality aversion and deterrence), but roughly consistent with the latter.

1. Introduction

We study the different effects of punishment on trust and trustworthiness in the presence of endowment heterogeneity. Trust and trustworthiness determine economic performance at the macro level and grease the wheels of governments (Arrow, 1974; Bachmann & Zaheer, 2006; Gambetta, 1988; Knack & Keefer, 1997). At the micro level, they both play an essential role in a variety of strategic environments such as incomplete contracts, the hold-up problem or the principal-agent relationship, which may include sequential investment decisions (e.g., the purchase of a good in which the seller chooses the quality of the product after the buyer bought it).

* Corresponding author at: School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom.
E-mail address: E.Fatas@uea.ac.uk (E. Fatas).
URL: http://www.uea.ac.uk/~amr11jwu (E. Fatas).

http://dx.doi.org/10.1016/j.joep.2016.09.006
0167-4870/

Starting with Berg, Dickhaut, and McCabe (1995), trust has been studied as a stylized two party sequential game where a trustor (the investor) unconditionally invests in the first stage of the game. Any amount invested is multiplied before a trustee (the allocator) unilaterally decides how much to give back. Because the sub-game perfect equilibrium under the assumption of purely self-interested subjects is that allocators will return nothing to investors, the investor's decision has been usually identified in the literature as the level of trust, whereas the level of trustworthiness (or reciprocity) has been frequently measured by the allocator's payback.

The determinants of trust and trustworthiness have been extensively studied in the laboratory (see Eckel & Wilson, 2011; Johnson & Mislin, 2011; Cooper & Kagel, 2016 for recent surveys), albeit most studies do not explicitly investigate how to foster them.[1] Recent theoretical work from anthropology and evolutionary game theory highlights the positive effects of individually costly punishment on cooperation (Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Fischbacher, 2003; Hauert, Traulsen, Nowak, Brandt, & Sigmund, 2007; Olcina & Calabuig, 2008). Laboratory experiments do also suggest that sanctions significantly increase contribution to public goods games (Chaudhuri, 2011), even between strangers (Fehr & Gächter, 2002).

Although the potential benefits of sanctions on trust-governed interactions like the ones described above seems straightforward (e.g., by using sticks (punishment) principals might increase the cooperation of their employees, and buyers may get higher quality products from sellers, trying to elude sanctions), the literature on the effects of punishment on trust and trustworthiness is still scarce (see Fehr & List, 2004; Fehr & Rockenbach, 2003; Houser, Xiao, McCabe, & Smith, 2008; Rigdon, 2009). Rigdon (2009) is the first paper in which punishment comes at a cost for investors. Punishment significantly increases trust only when it is very effective or relatively cheap for investors; i.e., when investors spending one token results in large losses for allocators.[2] In settings where punishment is costless for investors, and punishment is weak in that the cost of receiving punishment is smaller than the benefit of violation, the experimental evidence suggest that punishment crowds out trustworthiness, as allocators might be less likely to reciprocate when they get threats from investors (see Fehr & List, 2004; Fehr & Rockenbach, 2003; Houser et al., 2008).[3]

Experimental evidence on the effect of endowment heterogeneity on punishment effectiveness is much more scarce. Endowment heterogeneity is a crucial and common feature in most real life settings, like principal-agent relationships or bargaining situations with a monopolistic seller and competitive buyers. One simple example of how endowment heterogeneity may affect the efficacy of costly punishment is to consider the strategic dilemma faced by someone building a new house, and finding the outcome unsatisfactory. In the absence of a budget constraint, hiring a good lawyer to sue the builder is both an available option, and a credible threat. If resources are scarce, the balance between your resources and the builder's company size will most likely determine the extent to which you can effectively punish the builder. Endowment heterogeneity (rather the mere existence of sanctions, and its technology) thus becomes a relevant determinant of the damage you can credibly inflict to the builder.

In this paper, we investigate the effect of endowment heterogeneity on trust and trustworthiness in a controlled laboratory experiment. We give the investor the possibility of punishing the allocator at a cost, and systematically manipulate the distribution of initial endowments. When sanctions are available, trust could increase because the investor may anticipate the allocator will fear sanctions. Arguably, the allocator will be more or less afraid depending on the magnitude of the investor's endowment, relative to the allocator's one. If the investor is much wealthier than the allocator, we say that the investor's capacity of punishment is high because the investor can destroy most of the allocator's payoff investing a relatively small share of her own payoff. Alternatively, the allocator will fear sanctions relatively less if she is wealthier than the investor, and the investor's capacity of punishment of the investor is low.[4] As a result, the effect of sanctions on trust and trustworthiness could very reasonably be mediated by differences in initial endowments. Interestingly enough, endowment heterogeneity also changes other features of the trust game by exogenously introducing an unequal distribution of resources. Inequality aversion may generate strong behavioral reactions to non-reciprocal behavior in some endowment distributions, critically affecting the likelihood of punishment.

We test four different predictions in our experiment. The first one is related to the existence of *social preferences*. Both inequality aversion and reciprocity predict an effect of endowment heterogeneity in the trust game, as we will explain below (Coleman, 1990; Rodriguez-Lara, 2014; Smith, 2011; Xiao & Bicchieri, 2010). Secondly, the *deterrence hypothesis* assumes that punishment increases both trust and trustworthiness, with the effect of punishment being non-decreasing with the

[1] McCabe, Rigdon, and Smith (2003), Cox (2004), Cox, Kerschbamer, and Neururer (2014) or Ashraf, Bohnet, and Piankov (2006) specifically attempt to analyze the different motivations behind trust and reciprocity, including the role of intentions and the importance of altruism and expectations on behavior.

[2] The low-punishment treatment in Rigdon (2009) is such that investors need to spend 1 unit of their endowment to reduce the allocator's payoff in 1 unit, whereas the allocator's payoff is reduced in 3 units in the high-punishment treatment.

[3] In these setups, investors that are allowed to punish choose the amount they send to the allocator and a desired payback. If the desired payback is very large, this might be perceived as unfair by allocators, and therefore less money is returned. When investors refrain to fine, allocators are more likely to return a higher amount. Thus, the allocators' willingness to fulfill the investors' desired back-transfer depends negatively on the investors' requested amount, as well as on the investors' decision on the fine. Allocators also look at the investor's intention and whether or not the cost of receiving punishment is smaller than the benefit of violation (Houser et al., 2008). We decided not to include a desired-back transfer in our setting with punishment because contracts, even if they are not binding, can foster trust (Abbink, Irlenbusch, & Renner, 2000).

[4] At the same time, the impact of sanctioning institutions on trust is capped by the necessity of keeping enough resources to make the threat of sanctions credible. When the investor is wealthier than the allocator, the investor can invest relatively more and still keep resources to make the punishment credible. Note that in a linear public good game a cooperator always has the chance of punishing free riders with the returns of her investment in the public good. For the study of punishment in other recent bilateral settings, where punishment is less unequivocal than in the public good game see Abbink et al. (2000), Bolle, Breitmoser, and Schlächter (2011) or Nikiforakis, Oechssler, and Shah (2013).