Contents lists available at ScienceDirect

European Economic Review

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

Intuitive help and punishment in the field

Luis Artavia-Mora, Arjun S. Bedi, Matthias Rieger*

International Institute of Social Studies (ISS), Erasmus University Rotterdam, Netherlands

ARTICLE INFO

ABSTRACT

We test whether humans are intuitively inclined to help or punish strangers using a natural field experiment. The experiment manipulates the time available to decide whether to help or punish a stranger in an everyday situation. Our findings suggest that humans intuitively tend to help. However, time delay significantly reduces helping rates. In line with lab experiments, delay seems to override prosocial intuitions. There is no evidence that time delay affects rates of punishment. However, time delay magnifies self-reported fears of retaliation in response to direct punishment. We discuss our results with respect to previous research on the cognitive origins of human cooperation.

1. Introduction

Prosocial behavior and cooperation are at the heart of modern social interactions (see for instance, Axelrod and Hamilton, 1981; Axelrod, 1984; Fehr and Gächter, 2002; Fehr and Fischbacher, 2003; Nowak and Sigmund, 2005; Nowak, 2006; van Veelen et al., 2012; Gächter, 2012; Capraro, 2013). Conventional theoretical explanations for the origins of cooperation in today's societies pertain to kin selection or direct and indirect reciprocity (Hamilton, 1964; Trivers, 1971; Axelrod and Hamilton, 1981; Nowak et al., 1995; Fehr and Gächter, 2002; Nowak and Sigmund, 2005; Nowak, 2006). However, consider for instance one-shot encounters between strangers in daily life. It is somewhat puzzling that strangers often help or act for others' benefit at a personal cost (Fehr and Gächter, 2002; Balafoutas et al., 2014a). If such interactions are not solely governed by self-interest as standard economic models or theories of evolution would predict (Nowak, 2006; Sigmund, 2010), then the natural question that arises is: why do strangers help each other? And which factors and mechanisms can promote or undermine prosocial behavior? Answering such questions is central to our understanding of human cooperation and the evolution of societies.

A more recent literature has focused on the cognitive mechanisms underlying human cooperation (Rand et al., 2012; Rand and Nowak, 2013; Rand and Epstein, 2014; Rand, 2016; Capraro and Cococcioni, 2016). This literature rests on the dual-process cognitive framework featuring two competing systems of decision making: (1) *intuitive*, automatic, emotional, unconscious and faster decisions based on prior experience and beliefs versus (2) *deliberative*, more controlled, rational, reflective, effortful and slow decisions (Sloman, 1996; Kahneman, 2003; Loewenstein and O'Donoghue, 2004; Frankish and Evans, 2009; Kahneman, 2012; Evans and Stanovich, 2013). The resulting line of reasoning which motivates current experimental work is that *intuition* promotes prosocial behavior, while *deliberation* overrides such behavior (Rand, 2016).

In order to shed light on intuition versus deliberation in cooperative decision making, a series of lab experiments has employed decision time manipulations. Time pressure is used to trigger intuitive decisons and time delay is meant to elicit deliberative

---

* Corresponding author.
  E-mail address: matthias.rieger@graduateinstitute.ch (M. Rieger).

decisions.[1] A host of studies suggests that time delay indeed reduces prosocial behavior (Cone and Rand, 2014; Rand et al., 2012, 2013, 2014; as well as a meta study by Rand, 2016). A second group of studies has found no effect of such time manipulations (Tinghög et al., 2013; Verkoeijen and Bouwmeester, 2014), or, when there is an effect, it is limited to some groups of people and to specific social contexts (Rand et al., 2013; Capraro and Cococcioni, 2016). For instance, effects seem to disappear among experienced experimental subjects (Rand et al., 2013).[2] Effects can also depend on an experimental subject's general trust level (Rand and Kraft-Todd, 2014) and the overall trust environment (Capraro and Cococcioni, 2016).[3]

A related literature has also examined the impact of time manipulation on the decision to punish individuals. There is evidence that time delay moderates impulsive desires to punish offenders by decreasing negative emotions such as disagreement and resentment (Sutter et al., 2003; Grimm and Mengel, 2011; Wang et al., 2011; Neo et al., 2013).

To rationalize these patterns Rand et al. (2014, p.2) proposed the *social heuristics hypothesis*. The link links learning from experience and daily interactions with decision outcomes in unusual contexts such as a lab. Put differently, it theorizes that "daily life typically involves factors such as repetition, reputation and the threat of sanctions, all of which can make cooperation in one's long term self-interest" which in turn shapes "generalized cooperative intuitions." And these internalized inclinations may or may not lead to intuitive cooperation in the lab or other atypical contexts (see also Jordan et al., 2014).

More recently, Bear and Rand (2016) have formalized these empirical patterns within a cognitive game theory model of cooperation. Individuals in the model can use intuition or deliberation when deciding how to interact with others. However, deliberate responses require individuals to make an additional effort as they need to reason through their decision to cooperate or defect. In other words, there is a "trade-off" between making costly but informed decisions and cheap intuitive ones. That said, individuals play two sorts of games: (i) one-shot, or (ii) reciprocal, repeated prisoner's dilemma games (occurring at probabilities *1-p* and *p*, respectively). Only in repeated games do choices of individuals impact future encounters. The "social environment" then shapes the evolution and ultimately the nature of individuals. If the environment is sufficiently reciprocal (in other words, the probability *p* of repeated interactions is relatively high), individuals may follow a dual-process cognitive framework. That is, they will be intuitive cooperators who are, however, able to use costly deliberation. In the case of repeated games, deliberation may also lead to cooperation, while in one-shot games it favors defection. In sum, the model provides a rationale for why time delay (i.e. inducing deliberation) reduced cooperation in some of the previous experiments.

We contribute to the literature by examining whether time delay impacts the likelihood of helping and punishing strangers using a *natural field experiment*. While the concepts "cooperate" and "defect" do not translate directly to our experimental setting, the model by Bear and Rand (2016) does yield information on the expected impact of time delay. In our setting, we examine individually costly pro-social behavior in terms of helping and direct punishment and we experimentally induce deliberation through time delay (as outlined below). If the overall environment in the field is sufficiently reciprocal, and if helping behavior is correlated with the cooperative nature of individuals and driven by a similar dual-process cognitive framework, then time delay should decrease helping rates. Furthermore, if helping and punishment are driven by similar cognitive and psychological processes, we should see a similar pattern in both cases.

Our experiment has two notable features compared to previous studies in the lab: first, it yields a demographically diverse pool of subjects. Second, we can randomly and very naturally manipulate the time available to make a decision. Our design minimizes human mistakes and examines everyday human behavior (List, 2007, 2011; Balafoutas et al., 2014a). Rather than pushing buttons, the experiment is designed such that subjects may actually help a stranger at a personal cost. However, there is also one important design difference compared to papers which conduct lab experiments. While our paper is located in the emerging literature on intuitive cooperation, we examine pure helping behavior, and in an extension, direct and indirect punishment in one-shot interactions among strangers. Our design does not feature strategic elements nor create room for further interactions.

In order to test for the effect of time delay on helping and punishment rates in the field, we closely follow the methodology by Balafoutas and Nikiforakis (2012) and Balafoutas et al. (2014a). Specifically, we use actors that interact with strangers in the field. In the helping a stranger experiment, we asked an actor to drop a glove in a public park. We then record if strangers (subjects) who are passing by help, that is, pick up or alert the actor about the glove. We also collect basic socio-demographic characteristics of subjects through a post-experimental survey.

The novelty of our paper compared to the original methodology used by Balafoutas and co-authors is that we are adding an exogenous time manipulation treatment to test for intuitive versus deliberative responses. The actor drops the glove when an approaching subject is at a distance of either 4.5 m or 13 m from him/her. In our experiment, the average subject has about 3.5 s or 10 s to decide whether to help or defect. The short distance is our time pressure or baseline case meant to trigger fast and intuitive help, while the longer distance is our time delay treatment that promotes slow and deliberative responses.

In an extension we also estimate the impact of time delay on direct and indirect punishment in the field by building on Balafoutas et al. (2014a) and (2012). Punishing a stranger who violates a social norm may benefit society, but also comes at a personal cost and involves the fear of retaliation or counter-punishment (Fehr and Fischbacher, 2004; Gächter et al., 2008; Nikiforakis, 2008;

---

[1] Another strand of the literature has correlated response times with cooperation (Nielsen et al., 2014; Piovesan and Wengström, 2009; Recalde et al., 2014). However, recent work has shown that such an approach does not support a test of intuition versus deliberation but rather reflects decision conflicts (Krajbich, Bartling, Hare and Fehr, 2015; Evans, Dillon, and Rand, 2015). Therefore, we do not discuss these studies in more detail.

[2] Rand et al. (2013) report "vanishing" effect sizes associated with time manipulations in studies among increasingly experienced subjects recruited on Amazon Mechanical Turk.

[3] Capraro and Cococcioni (2016) recruited inexperienced subjects from India, a social environment with low quality institutions and levels of trust. They find that time manipulation has no effect on cooperation rates.