Journal of Economic Behavior & Organization 70 (2009) 93–105

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

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

Trust and reciprocity in incentive contracting☆

Mary Rigdon

Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, United States

ARTICLE INFO

Article history:
Received 15 January 2007
Received in revised form 9 January 2009
Accepted 12 January 2009
Available online 22 January 2009

JEL classification:
C91
C70
D63
D81

Keywords:
Trust
Reciprocity
Incentive contracts
Principal–agent
Experimental economics

ABSTRACT

Principals can attempt to get agents to perform certain actions preferable to the principal by using *ex post* punishments or rewards to align incentives. Field data are mixed on whether, and to what extent, such informal incentive contracting (paradoxically) crowds out efficient solutions to the agency problem. This paper explores, via a novel set of laboratory experiments, the impact of *ex post* incentives on informal contracts between principals and agents in bargaining environments in which there are gains from exchange and when there is an opportunity for the principal to relay a no-cost demand of the division of those gains. Incentive contracting in these environments does not crowd-out off-equilibrium cooperation, and at high incentive levels cooperation is crowded in.

"Homer, the key to worker motivation is trust." Hank Scorpio, CEO Globex Corporation

"Behold! The greatest breakthrough in labor relations since the cat o' nine tails." C.M. Burns, Owner, Springfield Nuclear Power Plant

1. Introduction

A key feature of many bilateral bargaining situations is that they involve one party (the principal) enlisting the other (the agent) to affect some action for a fee. Delegating the action by trusting the agent is costly to the principal, but there are potential joint gains from delegating—and it is the agent's action that determines the distribution of those gains. There may be many actions open to the agent, and the principal and agent may value the outcomes of each such action differently. In particular, the action the agent most values may be the least-valued alternative to the principal. Since there is no guarantee that an agent will take the principal's most valued action, a principal can try to get the agent to perform certain actions and not others by structuring incentives in the right way—i.e., by imposing an *ex post* punishment or reward. The interesting economic issue is to find out what ways of structuring incentives are efficient: should principals try to mitigate the agency problem by appealing to explicit rewards and punishments, and if so, what *size* of carrots and sticks work best? Or, does such explicit incentive contracting (paradoxically) destroy the seemingly robust natural tendency of agents to reciprocate trusting behavior, crowding cooperation out?

☆ The quotations in the epigraph are from *The Simpsons* —a resource, according to some of my students, for real-world examples.

E-mail address: mrigdon@umich.edu.

0167-2681/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2009.01.006

Concrete examples of principal–agent relationships like this are easy to find. One such example is the interaction between firms and workers. Both parties have an interest in a workplace in which each trusts the other—the firm, in particular, has an interest in high-worker productivity with little shirking and low monitoring costs. Another example is in the interaction between shareholders and CEOs. Notoriously, in recent years shareholders of large corporations have enlisted CEOs who have chosen actions to the detriment of the company but have nevertheless reaped lucrative severance packages for themselves. While sometimes these incentives are pre-committed (e.g., stock options and performance-based incentives in sports contracts), often they are *ex post*—agents know that some incentive *may* be enacted, but whether the principal rewards or punishes is a decision made *after* the agent performs the delegated action. This is the standard, for instance, in the distribution of year-end bonuses and (in New York schools) the assignment of lunchroom duty to under-performing teachers.[1] The effect of *ex post* incentives on principal–agent relationships – what we might call "incomplete contracts" – are the focus of this paper.

Although there is field evidence suggesting that carrots and sticks do align incentives (Yamagishi, 1988; Groves et al., 1994; Fernie and David, 1996; Prendergast, 1999; Eisenberger et al., 1999; Lazear, 2000; Fernie and Metcalf, 2000; Ensminger, 2001), there is also a considerable body of work suggesting that, paradoxically, making incentives explicitly tied to performance *erodes* an agent's intrinsic motivations with an overall net effect of decreased efficiency (Deci, 1971; Lepper and Greene, 1978; Deci and Ryan, 1999; Gneezy and Rustichini, 2000a; Benz et al., 2002). In managerial circles, this is known as the "Paradox of Organizational Trust" (Barnes, 1981; Kohn, 1993): by showcasing carrots and sticks, managers signal that they do not trust their employees to perform well on their own, and this undermines the goal of cooperation. This substitution effect is also known within economics as "motivation crowding out" (Frey, 1997).[2] With field data, it is difficult to vary parameters in a systematic way to separate the effects on trust due to incentive schemes, as most settings have both. The laboratory is useful here since it allows us to systematically vary the parameters and environments to help illuminate the paradox.

This paper systematically investigates incentive contracting under a novel set of laboratory conditions. First, I extend the investment game (Berg et al., 1995) by allowing principals to state their (cheap-talk) expectations of the agents, forming a contract with them. Second, I add an *incentive stage* to the game during which agents can be either punished or rewarded, conditional on whether they shirked or fulfilled the terms of the (cheap-talk) contract. In particular, I consider two kinds of incentive environments: environments in which principals may punish under-performing agents (P only), and environments in which principals may reward adequately and over-performing agents (R only). Third, in the incentive stage, I investigate both *low* incentive environments (with a 1:1 exchange rate for rewards/punishments) and *high* incentive environments (with a 1:3 exchange rate for rewards/punishments).[3]

This set of conditions is importantly different from previous work for at least three reasons. First, there is an existing body of work that systematically investigates the impact of the availability of incentives (some mixture of punishments or rewards) in, for example, ultimatum and proposer–responder games (Andreoni et al., 2003; Büchner et al., 2004; Gneezy, 2003). Yet there are no gains from exchange in these contexts, and that is an important feature of the kinds of real-world principal–agent problems we find interesting. Second, much is known about the impact of cheap-talk in coordination settings and in ultimatum games (Cooper et al., 1989, 1992; Clark et al., 2000; Duffy and Feltovich, 2002; Boles et al., 2003; Blume and Ortmann, 2007).[4] Similarly, Ben-Ner and Putterman (2007) and Ben-Ner et al. (2007) explore the effect that pre-play communication has on behavior in the investment game. However, the issue at stake in the research reported here is how the terms of a cheap-talk contract can have binding force for the purposes of *ex post* incentives. We simply do not know how such contracts interact with the ability to punish and the ability to reward in trust settings—like those involving workers and firms where such talk is highly likely to occur. Third, this is the first paper to investigate how such cheap-talk contracts, in a game with gains from exchange, and the *levels* of both punishments and rewards that are costly interact with efficient, cooperative behavior.[5] This is significant since, in general, it may be cheap or costly for a principal (e.g., a firm) to impart a significant (relatively speaking) reward or punishment upon on an agent (e.g., a wage employee).

The most closely related experiments are those testing the role of incentives in an efficiency wage game (Fehr et al., 1997, 1998; Fehr and Gächter, 2002; Fehr and Fischbacher, 2002).[6] However, there are several key differences between these

[1] The first example is nicely illustrated in Ensminger (2001), where Orma herders who perform well receive substantial rewards—in the form of one cow per year. The example in New York schools is due to Andreoni et al. (2003). The relevant point is that these are examples of *ex post* imposing of incentivizing devices.

[2] The crowding-out literature is surveyed by Bowles (2007); see also Gintis et al. (2005).

[3] The messages sent by principals are, strictly speaking, cheap-talk (in the sense of Crawford and Sobel (1982)) only in the baseline environment defined below—in that environment the messages have no direct payoff implications. However, once we consider environments with the incentive stage, the contents of the messages may have payoff implications since it will be on the basis of the terms in the message that a principal has the opportunity to impose *ex post* rewards/punishments. But the promises and threats carried by the message will still be incredible. Rather than change what we call the message between the environments, I will stretch standard terminology a little and continue to call such contracts "cheap-talk" contracts throughout.

[4] See Crawford (1998) for a review in coordination games.

[5] Fehr and Rockenbach (2003) and Fehr and List (2004) experimentally examine how only one level of *pre-committed* punishment interacts with transfers and returns in an investment game so agents know at the time of the decision what level of punishment they will face (if any). Additionally, punishment was *costless* to the principal.

[6] See Fehr and Falk (2002) for an overview of the effects of incentives. Bohnet et al. (2001) explore the relationship between trust and reciprocity and the degree of contract enforcement in a trust game. Lazzarini et al. (2003) examines experimentally how a principal's choice to enter an incentive contract when playing a trust game interacts with the probability of continuation with the agent as well as the cost of enforcement.