

Higher-order Beliefs and Social Preferences in Trust Games

Anthony S. Gillies Mary L. Rigdon
Department of Philosophy Institute for Social Research
Rutgers University University of Michigan

Abstract

When principals and agents interact, implicit contracts are the rule not the exception. But if the future choice of the agent is not contractable then efficiency can be hard to come by: if people are rational and self-interested then conventional equilibrium analysis favors no exchange at all between parties. This is both inefficient and at odds with the truism that principals and agents (both in the laboratory and in the wild) do manage to reach cooperative outcomes that Pareto dominate the equilibrium, even in one-shot encounters. That makes understanding the motivations that drive this choice behavior both theoretically and empirically significant.

We look to probe this question by turning it on its head: Are there ways of distributing information about the possible payoffs facing a principal-agent pair that open the door for agents not to cooperate — offering them degrees of plausible deniability — and do they walk through it? The intuition is that some portion of agents exhibit higher-order belief dependent motivations: they treat implicit contracts as having binding force precisely because they view the interactions as contracts. But a contract is only as good as the distribution of information about it between the parties is. So we manipulate that information, seeing whether agents who would otherwise cooperate defect instead.

Our focus here is on an experimental analysis of the connection between higher-order information about the space of monetary outcomes and cooperation in a simple two-person one-shot trust game (Figure 1). The game has enough of the structural properties to make it a reasonable model of a principal-agent interaction under an implicit contract: A can

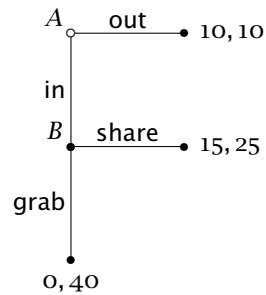


Figure 1: Trust Game

enlist B , but B 's choice at her decision node is beyond A 's control and there is no enforcement agency or shadow of the future or punishment mechanism that can be relied on. But the game is also simple enough to allow us to easily manipulate subjects' higher-order beliefs about the space of distributions in a play of the game. We examine the effects of asymmetric distributions of higher-order information — creating a degree of plausible deniability for B — on cooperation.

Our main hypothesis is that whether people in B 's position cooperate depends on whether they know that A knows that they know the space of possible monetary payoff bundles. There are, of course, a variety of models of motivation that attempt to explain the standard trust game data. Roughly, they fall into two classes: models of distributional fairness that posit, for instance, a taste for fairness or inequity aversion (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 1997) and higher-order belief-dependent motivation models such as intentions-based reciprocity or guilt aversion (e.g., Falk and Fischbacher, 2005; Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2009). Although testing between specific models is not our main focus here, our results do make contact with this literature. In particular, we find strong evidence in favor of higher-order belief-dependent motivations; thus if that phenomenon is difficult to accommodate in a given model, then our results will be likewise difficult for such a model to explain.

KEY WORDS: strong reciprocity, social preferences, trust game, interactive epistemology