



**HAL**  
open science

## Group reputations An experimental foray

Steffen Huck, Gabriele K. Lünser

► **To cite this version:**

Steffen Huck, Gabriele K. Lünser. Group reputations An experimental foray. Journal of Economic Behavior and Organization, 2009, 73 (2), pp.153. 10.1016/j.jebo.2009.09.001 . hal-00732667

**HAL Id: hal-00732667**

**<https://hal.science/hal-00732667>**

Submitted on 16 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Title: Group reputations An experimental foray

Authors: Steffen Huck, Gabriele K. Lünser

PII: S0167-2681(09)00216-9  
DOI: doi:10.1016/j.jebo.2009.09.001  
Reference: JEBO 2447



To appear in: *Journal of Economic Behavior & Organization*

Received date: 18-2-2008  
Revised date: 5-9-2009  
Accepted date: 6-9-2009

Please cite this article as: Huck, S., Lünser, G.K., Group reputations An experimental foray, *Journal of Economic Behavior and Organization* (2008), doi:10.1016/j.jebo.2009.09.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Group reputations

## An experimental foray

STEFFEN HUCK AND GABRIELE K. LÜNSER\*

June 2009

### Abstract

Often information structures are such that while individual reputation building is impossible, groups of agents would have the opportunity of building up a reputation. We experimentally examine whether groups of sellers in markets that suffer from moral hazard are able to build up reputations and, thus, avoid market breakdown. We contrast our findings with situations where sellers alternatively can build up an individual reputation or where there are no possibilities for reputation building at all. Our results offer a comparatively optimistic outlook on group reputations as long as groups are small. Even though sellers only receive some of the reputation benefits of withstanding short-run incentives to exploit trust, they are able to overcome the dilemma and successfully exploit the information structure. However, the ability to build successful group reputations depends on group size with trust breaking down in larger groups.

### Keywords

Trust; Group reputations; Moral hazard; Information conditions

### JEL Classification Codes

C72; C92; D40; L14

---

\* Dept. of Economics and ELSE, University College London, Gower St, London WC1E 6BT; email s.huck@ucl.ac.uk / g.luenser@ucl.ac.uk. Both authors acknowledge financial support from the Economic and Social Research Council (UK) via ELSE and an additional grant on “Trust and competition”. Huck is also grateful for additional funding by the Leverhulme Trust.

*It's a capital mistake to theorize before one has data.*  
Sherlock Holmes in Arthur Conan Doyle's  
"A Scandal in Bohemia"

*Ninety percent of the politicians  
give the other ten percent a bad reputation.*  
Henry Alfred Kissinger

## 1 Introduction

In his recent New Palgrave entry on 'reputation' Cripps (2009) identifies the question of how small groups may establish a collective reputation as one of the most fertile regions for future research on reputation building. Often information structures are such that individual reputation building is practically impossible. Take, for example, the string section of an orchestra, or other cases of team production where individual output is unobservable. Similarly, sellers of horizontally undifferentiated products may face the problem that consumers do only remember their average experience with the type of product rather than the individual quality of the good procured through a specific seller. A topical example of this is the market for organic and/or ethical food where sellers have not converged on uniform standards.

While the problem is of obvious practical interest, its theoretical analysis is, as Cripps points out, fraught with difficulties.<sup>1</sup> We take these difficulties as a point of departure for an *experimental* investigation. In an experiment, we can simply implement different information structures (for example, feedback about individual behavior or feedback about group behavior) and observe their empirical consequences—focusing on the one question of first-order importance: Can groups, despite the inherent public good problem, establish reputations such that market breakdown can be avoided?

While there is a substantial body of experimental research on reputation building,<sup>2</sup> the specific question of whether or how groups can establish reputations has not been studied yet. The paper closest to ours is Bohnet et al. (2001), where subjects in a trust game do receive group-

---

<sup>1</sup> Tirole (1996) develops an interesting model of collective reputations as an aggregate of individual reputations. In his model, groups are *large* and group member's history can only be observed with noise. The large group assumption makes individual decision independent of each other which greatly enhances tractability.

<sup>2</sup> See, for example, Camerer and Weigelt (1988) who explicitly introduce "crazy types" in a repeated trust experiment; Andreoni and Miller (1993) who compare strangers and partners in PD games; Bolton et al. (2004) who study trust games and compare partner treatments with random matching environments with full information; or Bohnet et al. (2005) who study how subjects can learn the mechanics of reputation building from each other.

based information. Alas, their focus is on other parameters of the game and there are no comparisons to other information structures.

Specifically, we study reputation building in a simple moral-hazard environment. Groups of four first and four second movers repeatedly play a simple binary trust game with random matching between rounds. First movers ('buyers') decide whether or not to buy an experience good, that is, a good whose quality they can only infer when using it, but not when inspecting it in the shop.<sup>3</sup> Second movers ('sellers') decide about whether to supply high or low quality. Sellers earn most if they sell low quality and least if the buyer does not buy. Buyers earn most if they buy high quality and least if they buy low quality. Accordingly, the unique Nash equilibrium predicts that buyers will abstain and markets break down completely.

In this setup, we vary the information provided to first movers between rounds. In particular, in our main treatment, treatment GROUP, that we designed to study group reputations, we inform first movers after each period about *the distribution of outcomes* in their group of eight. In this treatment we find that 18% of all matches result in mutually beneficial trade. In 38% of all cases, first movers trust the second mover and second movers reward this, on average, 41% of the time with high quality. In order to benchmark these levels of trust, trustworthiness and efficiency, we contrast this treatment with two controls in which first movers are only informed about the outcome of their own game. The control treatments differ in whether or not sellers are identifiable, and, hence can or cannot build up an individual reputation. In the first control, treatment NO, second movers do not have labels such that interaction is completely anonymous and individual reputation building is ruled out. In the second control, treatment INDIVID, sellers do carry labels such that individual reputation building becomes possible. Table 1 summarizes our design.

In treatment NO, we basically observe the Nash equilibrium prediction of total market breakdown: Only 5% of all matches result in mutually beneficial trade (with an average trust rate of 21% and average trustworthiness of 19%). In treatment INDIVID, we find, on the other hand, that 19% of all matches result in mutually beneficial trade, almost exactly the same number as in treatment GROUP. However, there is a little less trust in that treatment (31%) despite higher average quality (59%).

---

<sup>3</sup> Instructions are framed neutrally, see appendix.

**Table 1:** Treatments

	Identification of seller	Reputation building	Group size / subjects
GROUP	no	group reputation	4+4 / 48
INDIVID	yes	individual reputation	4+4 / 48
NO	no	no reputation	4+4 / 48
BIG GROUP	no	group reputation	8+8 / 96

Thus, we can benchmark our findings from the main treatment GROUP in two ways. First of all, we find that information structures that allow for group reputation building do enhance efficiency substantially. The share of matches that result in high-quality trade is *quadrupled* compared to the anonymous benchmark of treatment NO. Second, the magnitude of this effect is of precisely the same order as that of individual reputation building with private feedback about own interactions.<sup>4</sup>

Of course, the stability of our finding may crucially depend on group size. As groups grow, the marginal return to an individual investment in the group's reputation is shrinking and, hence, one might suspect that with larger group sizes trust breaks down.<sup>5</sup> In order to investigate this, we run an additional control treatment, BIG GROUP, where we double the number of sellers and buyers. As was to be feared, we observe a complete breakdown of trust in this treatment.

Nevertheless, we consider our results as providing a comparatively optimistic outlook on group reputations. Despite the fact that each seller in our experiment receives only one quarter of the reputation benefits of his investment in the good standing of his group, we find that, in small groups, sellers are able to overcome this dilemma and successfully exploit the provided information structure. Small groups are able to build up reputations and, thus, avoid market breakdown. However, larger groups are not.

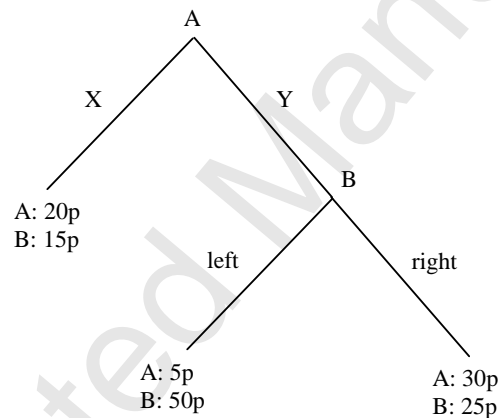
<sup>4</sup> Higher levels of efficiency can, however, be achieved by providing detailed information about the entire history of all matches in a group to all of its members; see Bohnet et al. (2005).

<sup>5</sup> For group size effects in public good games, see the seminal work by Isaac and Walker (1988).

## 2 Design and Procedures

Figure 1 shows the specific experimental payoffs in pence for buyers (A) and sellers (B) in our moral-hazard environment which reflects the incentive structure already discussed in the introduction—buyers long for consuming a high quality good (‘right’) while sellers long for selling the low quality good (‘left’). If buyers buy (Y) a low quality good they are worse off than not buying at all (X). If sellers do not sell the good they are worse off than selling the high quality good. The framing in the experiment was neutral and strategies and player roles were as labeled in the game tree. Assuming that subjects maximize their own monetary income, it is straightforward to see that the game has a unique Nash equilibrium (X, left) in which the buyer does not buy and the seller provides low quality if he has custom.

**Figure 1:** The trust game



The experiment was conducted in the ELSE/UCL experimental laboratory. Subjects received written instructions and answered some control questions before the start of the experiment to ensure they had a complete understanding of the game rules. All sessions were computerized and the experimental software was developed with Fischbacher’s (2007) z-Tree.

In total, 240 students of various fields participated in the experiment and, on average, earned £11.15 (including a £5 show-up fee). For each of the three treatments we conducted six separate sessions, each with one matching group. Every subject only participated in one session.

In all three treatments, the stage-game is repeated 30 times which is known to subjects from the beginning of the experiment; subjects are either a buyer or a seller for the entire experiment and are randomly rematched at the beginning of each period; matching groups consist of four buyers and four sellers.

Subjects receive different feedback information between rounds in our three treatments (GROUP, INDIVID, and NO). In GROUP, where sellers have no labels, and in INDIVID, where sellers do have labels buyers can track the history they know via a visual interface on the left part of the screen. In INDIVID this history window shows four columns of hash (#) signs, each column representing one seller and each row representing one period. Initially, each column consists of thirty *white* hash signs. Then, period by period, hash signs change their color according to what happened in the game: a hash turns *grey* if a subject did not receive any information about a particular seller in the previous period, it turns *black* if a seller had a non-trusting buyer, it turns *red* if a seller had custom and chose low quality, and, finally, it turns *green* if a seller had custom and chose high quality. In GROUP the history window shows information on an aggregate level. It has three columns, each column representing one of the three outcomes (X; Y/left; Y/right) and each row representing one period. Then, after each period, subjects see how many (of the four total buyer-seller-matchings) reached a particular outcome. (Screenshots are shown in the Appendix.)

It is worthwhile to note that with this design treatment INDIVID also provides the information that is available in treatment GROUP. Buyers can simply aggregate the individual feedback and thus track group performance as well. From that point of view, sellers can build up individual and group reputations in INDIVID.

Our additional control treatment BIG GROUP is just like treatment GROUP but with eight sellers and eight buyers interaction in each group.

### 3 Experimental Results

Table 2 summarizes the data we observed in our three main treatments, GROUP, INDIVID, and NO. The upper part of the table reports average trust rates (the average trade volume), honor rates (the average share of high quality among traded goods) and the performance rate (the average share of buyer-seller pairings that resulted in high-quality trade). Additionally, the lower part of the table reports statistical tests for treatment effects. These MWU tests use group-level averages over 30 periods as a unit of observation.



**Table 2:** Overview of aggregated results from small groups

	trust rate	honor rate	performance rate
<i>GROUP</i>	0.38 (0.16)	0.41 (0.16)	0.18 (0.09)
<i>INDIVID</i>	0.31 (0.14)	0.59 (0.15)	0.19 (0.12)
<i>NO</i>	0.21 (0.16)	0.19 (0.11)	0.05 (0.05)
	<i>treatment effects</i>		
<i>GROUP-INDIVID</i>	$p = 0.131$	$p = 0.075$	$p = 0.316$
<i>GROUP-NO</i>	$p = 0.027$	$p = 0.019$	$p = 0.015$
<i>INDIVID-NO</i>	$p = 0.168$	$p = 0.002$	$p = 0.005$

Standard deviations are given in parentheses. Treatment effects are tested by one-tailed Mann-Whitney U-tests (\*).

As Table 2 shows the market is such that in the absence of any reputation building almost full breakdown occurs. Only 5% of all interactions result in mutually beneficial trade. Things change drastically when information about sellers is present. As already discussed in the introduction the amount of high-quality trade nearly quadruples in both treatments with reputation building.

Compared to treatment NO, demand in treatment GROUP is up by 81%. This difference is significant and holds almost through the entire 30 periods. In all but period 25 and the last period the average trust rate in GROUP is higher than in NO.<sup>6</sup> In fact, the trust rate in treatment GROUP is even slightly higher than the trust rate in INDIV where sellers are labeled and, thus, can build up individual reputations. This difference is not quite significant in our data set but interesting enough to warrant some further attention below.

How is average quality affected by different reputation building possibilities? Information structures that allow for group reputation are effective in significantly raising quality. Average quality is more than doubled when comparing GROUP to NO. However, the average share of high-quality goods in GROUP is nevertheless lower than the share of high-quality goods in INDIV (41% vs 59%). In light of the public good problem that sellers face in GROUP (where high quality supply cannot be attributed to a specific seller) this is not surprising.

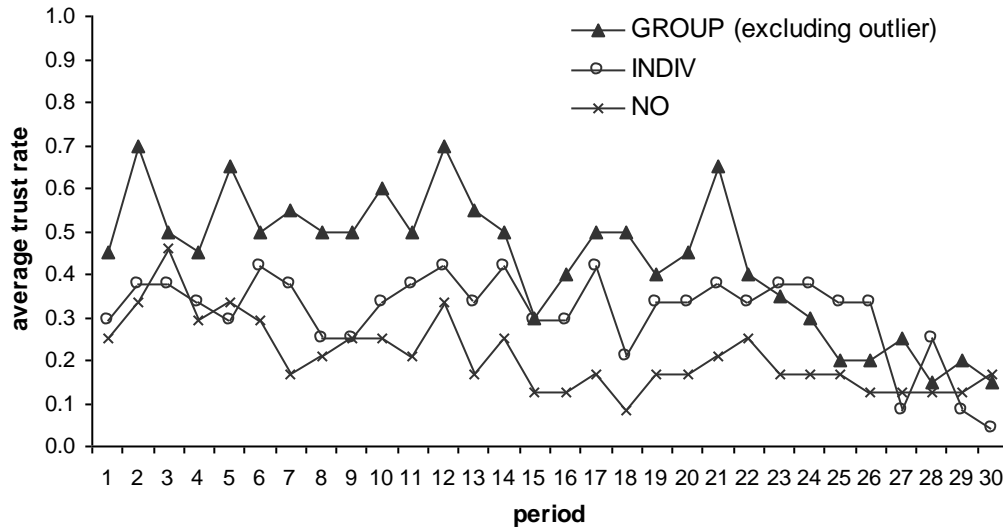
There are also some interesting differences in heterogeneity between GROUP and INDIV. In INDIV average honor rates vary between 0.41 and 0.79. In GROUP there are five observations where the rate varies only slightly between 0.46 and 0.52, while there is one

<sup>6</sup> In period 25 the average trust rates in GROUP and NO are equal.

outlier where the average honor rate is just 0.09. Obviously, sellers in this group failed to cooperate in their ‘reputations-as-a-public-good dilemma’. It is interesting to have a brief look at the dynamics in this group. In the first two periods buyers were extremely skeptical and sellers did not have any custom at all. In the third period two buyers were giving it a try but both received low quality. In the next period followed an instance of high-quality trade but, apparently, this was too late. Demand stayed at a level of virtually zero afterwards. These dynamics suggest it is extremely important to build up a good group reputation *early on*. As mentioned before, sellers face a public good dilemma when it comes to the provision of high quality. But as the story from this group illustrates quite clearly it is a public good game *with endogenous stakes*. If there is little cooperation initially, demand will fall and stakes become very low which makes future cooperation even harder.

We can now come back to the perhaps slightly surprising observation that trust rates are higher in GROUP than INDIV despite average quality being lower. This difference narrowly failed to be significant in the complete data set. However, if we exclude the session where things go badly wrong from the first period, we do find a significant difference (average demand in GROUP without the outlier is .44 while it is .31 in INDIV—a difference that is significant at .034; MWU one-tailed). It appears that most buyers understand that sellers will find it harder to cooperate if demand is low. If buyers do not trust enough, sellers’ incentives to cooperate by choosing high quality diminish. Thus, buyers need to trust more, in particular initially, in treatment GROUP if market breakdown is to be avoided. This is borne out in the data that show that the difference in trust rates between GROUP and INDIV is particularly pronounced in the first half of the experiment. Excluding the outlier in GROUP Figure 2 plots trust rates over time and visualizes the ‘leap of faith’ with aggregate information.

Of course, as the group that is excluded from Figure 2 shows, the alternative to this picture is essentially a completely flat zero line. Overall reputation building in groups appears to be more fragile than individual reputation building. If things don’t work well, they hardly work at all in treatment GROUP. On the other hand, the vast majority of groups does exploit the information structure successfully.

**Figure 2:** Average trust rate over time

The question remains how robust our findings on group reputations are once the size of the group varies. In order to investigate this issue in more detail we have run a control treatment, BIG GROUP where the same information is provided as in treatment GROUP but that has eight sellers and eight buyers. The data from BIG GROUP is shown in Table 3. It turns out that the performance measures in this treatment are statistically indistinguishable from treatment NO. The trust rate is .2, the honor rate .27, and the performance rate is .05. This is statistically indistinguishable from treatment NO. Interestingly, there is also very little variance between these bigger groups.<sup>7</sup> Thus, a very clear picture emerges. The ability to build up group reputations successfully depends crucially group size.

**Table 3:** BIG GROUP results

	trust rate	honor rate	performance rate
<i>BIG GROUP</i>	0.20 (0.01)	0.27 (0.03)	0.05 (0.01)

*Standard deviations are given in parentheses.*

<sup>7</sup> The standard deviations for the three measures are .01, .03, and .01 respectively.

## 4 Conclusions

In this paper, we have made a first experimental foray into understanding reputation building by groups. We find that group reputations are built up despite the inherent dilemma problem that agents face. However, cooperative investments into group reputations by trustees require an initial leap of faith by trustors. In the context of markets for experience goods, buyers need to create sufficient initial demand to make high-quality provision for sellers worthwhile.

In terms of efficiency, we find that group feedback about the aggregate behavior of all sellers is just as effective as individual reputation building with private feedback about own interactions, which suggests an interesting quantity-quality tradeoff for feedback information. However, we also find that larger groups fail to solve the dilemma they are faced with. This is perhaps not surprising as the marginal return of an individual investment in the group's good standing is decreasing in group size.

## References

- Andreoni, J, Miller, JH, 1993. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. *Economic Journal* 103, 570-585.
- Bohnet, I, Harmgart, H, Huck, S, Tyran, JR, 2005. Learning Trust. *Journal of the European Economic Association* 3, 322-329.
- Bohnet, I, Frey, BS, Huck, S, 2001. More order with less law: On contract enforcement, trust, and crowding. *American Political Science Review* 95, 131-144.
- Bolton, G, Katok, E, Ockenfels, A, 2004. How Effective are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science* 50, 1587-1602.
- Camerer, C, Weigelt, K, 1988. Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica* 56, 1-36.
- Cripps, MW, 2009. Reputation, *New Palgrave Dictionary of Economics Online*, Palgrave Macmillan, doi:10.1057/9780230226203.1424.
- Fischbacher, U, 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10, 171-178.
- Isaac, RM, Walker, JM, 1988. Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *Quarterly Journal of Economics* 103, 179-199.
- Tirole, J, 1996. A Theory of Collective Reputations (With Applications to the Persistence of Corruption and to Firm Quality). *Review of Economic Studies* 63, 1-22.

Screenshots

INDIV

Period 14 of 30 Remaining time [sec]: 32

Round	B1	B2	B3	B4
29	#	#	#	#
28	#	#	#	#
27	#	#	#	#
26	#	#	#	#
25	#	#	#	#
24	#	#	#	#
23	#	#	#	#
22	#	#	#	#
21	#	#	#	#
20	#	#	#	#
19	#	#	#	#
18	#	#	#	#
17	#	#	#	#
16	#	#	#	#
15	#	#	#	#
14	#	#	#	#
13	#	#	#	#
12	#	#	#	#
11	#	#	#	#
10	#	#	#	#
9	#	#	#	#
8	#	#	#	#
7	#	#	#	#
6	#	#	#	#
5	#	#	#	#
4	#	#	#	#
3	#	#	#	#
2	#	#	#	#
1	#	#	#	#

You are paired with B4  
Choose which way to go:  X  Y

OK

Key:  
 # Not done yet  
 # Didn't participate  
 # Went left  
 # Went right  
 # No information

GROUP

Round 30 of 30 Remaining time [sec]: 0

Round	# X	# Y / left	# Y / right
29	4	0	0
28	4	0	0
27	3	1	0
26	3	1	0
25	2	2	0
24	2	2	0
23	2	0	2
22	2	1	1
21	2	2	0
20	1	1	2
19	3	0	1
18	1	2	1
17	1	3	0
16	2	1	1
15	2	1	1
14	1	1	2
13	1	1	2
12	1	2	1
11	2	0	2
10	0	2	2
9	2	2	0
8	1	1	2
7	2	2	0
6	1	0	3
5	0	1	3
4	2	2	0
3	2	1	1
2	1	1	2
1	2	1	1

Choose which way to go:  X  Y

OK

Key:  
 # X how many A's chose X  
 # Y/left how many A's chose Y and B's went "left"  
 # Y/right how many A's chose Y and B's went "right"