Ifo Institute for Economic Research at the University of Munich

Group Reputations
An Experimental Foray

Steffen Huck
Gabriele K. Lünser

Ifo Working Paper No. 51

November 2007

Ifo Working Paper No. 51

Group Reputations
An Experimental Foray

Abstract

Often information structures are such that while individual reputation building is impossible groups of agents would have the possibility 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 rather optimistic outlook on group reputations. Even though each seller only receives some of the reputation benefits of withstanding short-run incentives, sellers are able to overcome the dilemma and successfully exploit the information structure.

JEL Code: C72, C92, D40, L14.
Keywords: Trust, group reputations, moral hazard, information conditions.

Steffen Huck
University College London
Gower Street
London WC1E 6BT, UK
Tel.: +44(0)20/7679-5895
s.huck@ucl.ac.uk

Gabriele K. Lünser
University College London
Gower Street
London WC1E 6BT, UK
Phone: +44(0)20/7679-5852
g.luenser@ucl.ac.uk

1 Introduction

In his recent New Palgrave entry on 'reputation' Cripps (forthcoming) 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.

While the problem is of obvious practical interest its theoretical analysis is, as Cripps points out, fraught with difficulties.[1] 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?

More 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.[2] 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

[1] 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.

[2] Instructions are framed neutrally, see appendix.

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%).

Table 1: Treatments

	Identification of seller	Reputation building
GROUP	no	group reputation
INDIVID	yes	individual reputation
NO	no	no reputation

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.[3] This makes for an interesting quantity-precision comparison. Quadrupling the amount of information as we move from treatment NO (where first movers only observe the outcome from one match—their own) to treatment GROUP (where they

[3] Higher levels of efficiency can, however, be achieved by providing detailed information about the entire history of all matches in a group to every of its members; see Bohnet *et al.* (2005).

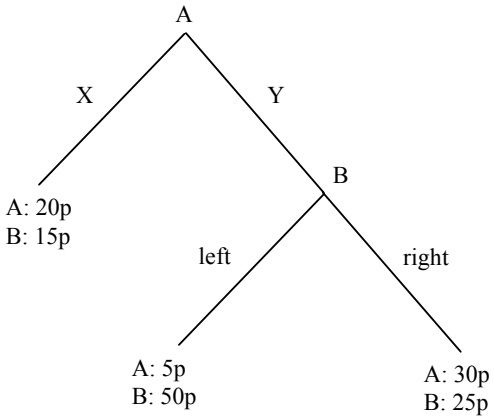
observe all four) is roughly equivalent to making one single observation more precise by attaching the seller's name to it.

In all, we consider our results as providing a rather 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 sellers are able to overcome this dilemma and successfully exploit the provided information structure. Groups are able to build up reputations and, thus, avoid market breakdown.

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 then not buying at all (X). If sellers do not sell the good they are worse off then 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, 144 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.)

3 Experimental Results

Table 2 summarizes the data we observed in our three treatments. 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

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

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.[4] 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.

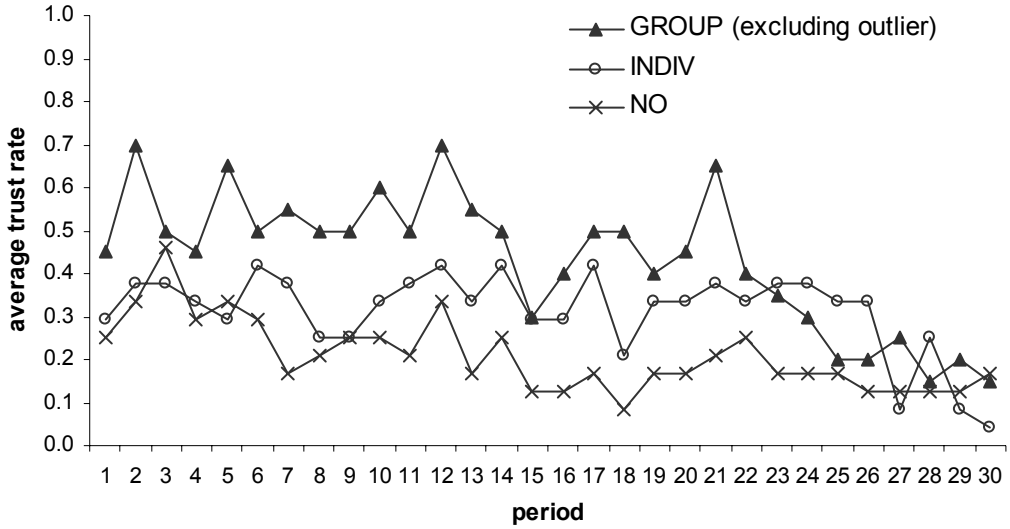
How is average quality affected by different reputation building possibilities? Information structures that allow for group reputation are effective in significantly raising quality. Average quality is more than doubled when comparing GROUP to NO. However, the average share of high-quality goods in GROUP 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 this is not surprising.

[4] In period 25 the average trust rates in GROUP and NO are equal.

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 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 sellers fail to cooperate and where also demand is extremely low, 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 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.

Figure 2: Average trust rate over time



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.

References

Bohnet, Iris; Harmgart, Heike; Huck, Steffen and Tyran, Jean-Robert (2005): Learning Trust, *Journal of the European Economic Association* 3, 322-329.

Cripps, Martin W.: Reputation, *New Palgrave Dictionary of Economics*, Palgrave Macmillan, forthcoming.

Fischbacher, Urs (2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10, 171-178.

Tirole, Jean (1996): A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality), *Review of Economic Studies* 63, 1-22.

Instructions (treatment GROUP)

Welcome to our experiment!

Please read these instructions carefully! Do not speak to your neighbors and keep quiet during the entire experiment! In case you have a question raise your hand! We will then come to you.

In this experiment you will repeatedly make decisions. Doing this you can earn money. How much you earn depends on your decisions and on the decisions of other participants. All participants receive the same instructions.

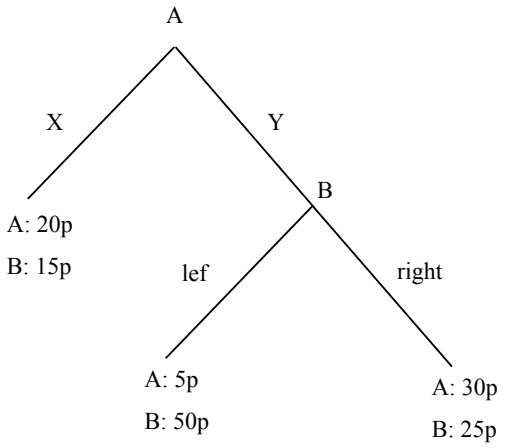
All participants stay anonymous to the experimenter and also to other participants.

In the experimental situation there are two agents called A and B, respectively. In each round every A will be ***randomly*** *matched with a B: Each A interacts with exactly one B and vice versa. Altogether there are eight participants. All participants are randomly assigned a role (A or B) at the beginning of the experiment and roles are kept throughout the experiment. There will be four A-participants and four B-participants.*

At the beginning of each round every A-participant is randomly matched with a B-participant by the computer. That is, in each round the computer randomly chooses for each A-participant one B-participant to interact with. This process is repeated in the next round. Therefore, you may meet with the same participant in two consecutive rounds (in one out of four cases) or meet a different participant in the next round (in three out of four cases). However, you never know with whom you have been matched in a particular round nor will you learn this at a later date.

Once participants have been matched, it is A's turn to make a decision. More specifically, A has to choose between option X and option Y. If he picks option X, A will earn 20 p and B will earn 15 p. If he picks option Y, the payoffs depend on B's choice who has to decide whether he wants to go "left" or "right". If he decides to pick "left", A will earn 5 p and B will earn 50 p. If he decides to pick "right", A will earn 30 p and B will earn 25 p.

These rules are illustrated in the following „tree":



The experiment consists of 30 rounds. All participants keep the role assigned to them throughout the experiment. After each round you will be informed about what has happened and you will be reminded of your earnings and your total earnings so far.

Moreover, all participants (A's and B's) can keep track of the entire aggregated history. There will be a screen depicting the aggregated history of all eight participants. For each round the screen will show

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

Note that there is always a total of four A-B-matchings so that in each round the three numbers will add up to four.

These are the rules. You can trust us that everything will happen exactly according to these rules. Take your time going over these instructions again. And feel free to ask questions. But don't shout! Simply raise your hand.

Your total earnings that equal the sum of your individual earnings in the 30 rounds (plus your show-up fee of £5) will be paid to you right after the experiment in cash.

Your role is displayed to you on the first screen at the beginning of the experiment.

Screenshots

INDIV

GROUP

Ifo Working Papers

No. 50 Meier, V. and G. Schütz, The Economics of Tracking and Non-Tracking, October 2007.

No. 49 Buettner, T. and A. Ebertz, Quality of Life in the Regions – Results for German Counties, September 2007.

No. 48 Mayr, J. and D. Ulbricht, VAR Model Averaging for Multi-Step Forecasting, August 2007.

No. 47 Becker, S.O. and K. Wohlrabe, Micro Data at the Ifo Institute for Economic Research – The "Ifo Business Survey", Usage and Access, August 2007.

No. 46 Hülsewig, O., J. Mayr and S. Sorbe, Assessing the Forecast Properties of the CESifo World Economic Climate Indicator: Evidence for the Euro Area, May 2007.

No. 45 Buettner, T., Reform der Gemeindefinanzen, April 2007.

No. 44 Abberger, K., S.O. Becker, B. Hofmann und K. Wohlrabe, Mikrodaten im ifo Institut – Bestand, Verwendung und Zugang, März 2007.

No. 43 Jäckle, R., Health and Wages. Panel data estimates considering selection and endogeneity, March 2007.

No. 42 Mayr, J. and D. Ulbricht, Log versus Level in VAR Forecasting: 16 Million Empirical Answers – Expect the Unexpected, February 2007.

No. 41 Oberndorfer, U., D. Ulbricht and J. Ketterer, Lost in Transmission? Stock Market Impacts of the 2006 European Gas Crisis, February 2007.

No. 40 Abberger, K., Forecasting Quarter-on-Quarter Changes of German GDP with Monthly Business Tendency Survey Results, January 2007.

No. 39 Batchelor, R., Forecaster Behaviour and Bias in Macroeconomic Forecasts, January 2007.

No. 38 Sülzle, K., Innovation and Adoption of Electronic Business Technologies, December 2006.

No. 37 Overesch, M. and G. Wamser, German Inbound Investment, Corporate Tax Planning, and Thin-Capitalization Rules – A Difference-in-Differences Approach, December 2006.

No. 36 Kempkes, G. and C. Pohl, The Efficiency of German Universities – Some Evidence from Non-Parametric and Parametric Methods, October 2006.

No. 35 Kuhlmann, A., German Productivity – A Reassessment via the New Ifo Productivity Database, October 2006.

No. 34 Kuhlmann, A., What is the X-Factor in the German Electricity Industry?, September 2006.

No. 33 Temple, J. and L. Wößmann, Dualism and cross-country growth regressions, August 2006.

No. 32 Baumann, F., V. Meier and M. Werding, Transferable Provisions in Individual Health Insurance Contracts, July 2006.

No. 31 Abberger, K., Qualitative Business Surveys in Manufacturing and Industrial Production – What can be Learned from Industry Branch Results?, May 2006.

No. 30 Ruschinski, M., Investigating the Cyclical Properties of World Trade, May 2006.

No. 29 Holzner, Chr., V. Meier and M. Werding, Time Limits in a Two-tier Unemployment Benefit Scheme under Involuntary Unemployment, April 2006.

No. 28 Eggert, W. and A. Haufler, Company Tax Coordination cum Tax Rate Competition in the European Union, April 2006.

No. 27 Lachenmaier, S. and H. Rottmann, Employment Effects of Innovation at the Firm Level, April 2006.

No. 26 Radulescu, D.M. and M. Stimmelmayr, Does Incorporation Matter? Quantifying the Welfare Loss of Non-Uniform Taxation across Sectors, March 2006.

No. 25 Lessmann, Chr., Fiscal Decentralization and Regional Disparity: A Panel Data Approach for OECD Countries, March 2006.

No. 24 Fuchs, Th., Industry Structure and Productivity Growth: Panel Data Evidence for Germany from 1971–2000, December 2005.