The behavioral and institutional determinants of gift exchange: evidence from a real-effort labor market experiment*

Miguel A. Fonseca† and João Madeira‡

January 27, 2014

Abstract

We present results from a real-effort labor market experiment where workers and firms trade multiple units of labor but quality is non-contractible. We systematically vary market structure, workers' knowledge of firm payoffs, and a pure redistributive income tax. While there is gift exchange in our baseline conditions, it disappears when workers do not know firms' payoffs and there is excess supply of labor. This result reconciles the discrepancy between laboratory data, where evidence of gift exchange is robust, and field data, where it is not. Interestingly, gift exchange is also crowded out by an income redistribution tax.

Keywords: labor markets, gift exchange, real effort experiment.

JEL classification numbers: **C72, C92, D43**

*We are very grateful to Tim Miller for his outstanding help in programming the software and running the sessions, as well as Hywel Williams for his algorithm to compute some of the variables in our data analysis. We also thank Ramon Cobo-Reyes for comments and suggestions. Financial support from the British Academy grant SG091122 and the University of Exeter Business School is also gratefully acknowledged. Any errors are ours alone.

†University of Exeter; Tel: +44 (0)1392 262584; Fax: +44 (0)1392 263242; Email: *m.a.fonseca@exeter.ac.uk*

‡University of Exeter; Tel: +44 (0)1392 264486; Fax: +44 (0)1392 263242; Email: *j.madeira@exeter.ac.uk*

1 Introduction

A large literature in experimental economics has emerged based on the seminal work by Akerlof (1982), who proposed a model of labor markets which rationalizes non-competitive equilibrium wages by means of gift exchange: workers reciprocate higher-than-equilibrium wages with higher-than-minimum effort. However, evidence of gift exchange in experimental labor markets is mixed: while laboratory studies with financially costly effort have consistently found positive evidence of gift exchange, field experiments where subjects exert real effort in a naturalistic setting have failed to uncover similarly robust results. Therefore, what drives gift exchange remains an open question.

It has been difficult to reconcile results from the laboratory with results from the field, because the two types of studies typically differ along a number of dimensions. The first is the nature of the experimental task: in most lab studies, subjects choose effort from a payoff table, while they exert real effort in the field. The second is whether the quantity of labor output is contracted upon: it is fixed in the lab, while it is variable in the field. The third is the information workers have about the firm's production function: to the best of our knowledge it is always present in the lab but not in the field. Finally, there is market structure: while field experiments only consider markets with excess labor supply, laboratory studies have considered multiple market structures.

The purpose of this paper is to understand the behavioral and institutional determinants of gift exchange. In particular, we study the role of market structure, income redistribution, and information workers have about the production function of the firm, as well as individual characteristics (such as worker productivity or other-regarding preferences). We do this using a real-effort labor market experiment. Our design combines the high degree of control over the environment typical of laboratory studies, with a real-effort task which is an important characteristic of field experiments.

Our paper implements a real-effort laboratory experiment on gift exchange in which firms and workers contractually agree on a quantity and a piece rate for a commodity. Each unit of that commodity can be of high or low quality, whereby high-quality units are more valuable for the firm, but more difficult to produce and therefore costlier to the worker. The possibility of gift exchange emerges because quality is not contractible: workers can compensate higher piece rates by producing a higher fraction of high-quality output through non-contracted effort. As in the classic gift exchange game, in the equilibrium with self-interested preferences, workers produce only low-quality output, and firms pay competitive piece rates. In addition, unlike the previous lab studies, where it is inexistent or the field studies where it is unobservable, we are able to directly measure individual worker productivity, and estimate its effect on gift exchange.

We consider two types of labor market structures, one in which there is excess demand for labor and another where there is excess supply of labor. Our baseline treatment is designed to verify whether the gift exchange result from an environment without real effort and a single unit of output extends to a real effort experiment with multiple units of contracted output, as per most field experiments (we consider an additional treatment where output is fixed, which we describe further below). Indeed, we find evidence of gift exchange in our baseline condition both when there is excess demand for labor and when there is excess supply of labor.

Our first treatment tests for the robustness of gift exchange by adding the real-life feature that workers do not know the firm's production function, and therefore cannot calculate the financial consequences to the firm of exerting costly non-contracted effort. We hypothesized that taking away information about the firms' profit function from workers would reduce gift exchange because workers can no longer work out the implications of their costly effort on the size and split of the surplus from trade. Our data supports that hypothesis, but only when there is excess supply of labor. When firms compete for workers, the absence of the firms' payoff information does not impact gift exchange. This result broadly reconciles the evidence from gift exchange experiments conducted in the field with the evidence from experiments done in the lab. Most field experiments are operationalized in markets in which there is excess supply of labor and where workers have no direct knowledge of the firm's production function. In these studies, much like the relevant treatment in our own experiment, there is limited evidence that workers exert costly effort that is consistent with gift exchange. In our case, costly effort is likely driven by intrinsic preferences over surplus distribution.

By slightly modifying our design, we are also able to make a contribution to the literature on social preferences, which is tightly related to the literature on gift-exchange experiments. In particular, we wish to understand the different channels through which inequality aversion and reciprocity concerns operate when determining behavior in real-effort labor markets. To test for the relevance of preferences over income distributions, we implemented a purely redistributive income tax. We hypothesized that gift exchange would increase as a result of the tax if workers have other-regarding preferences. Workers not only benefit financially from the extra wealth generated by gift exchange, but they also gain utility from the fact that income inequality at the market level is reduced, not only with respect to firms, but also relatively to the unmatched player in the market, who now receives positive income. Surprisingly, a pure redistribution tax crowds out gift exchange in both market configurations. Together with individual-level data, we show that fairness preferences mostly determine baseline levels of effort made independently of wages, rather than gift exchange itself.

In contrast, gift exchange, whenever present, seems to be driven mostly by reciprocity concerns. By analyzing a treatment in which quantity is exogenously fixed, we take advantage of heterogeneity in worker ability to show that contracted quantity successfully functions as a signal of the firm's kindness towards workers. In the treatment where quantity is endogenously determined, workers' effort is correlated with wage only when contracted quantity is below a particular ability-determined threshold. However, in the treatment where quantity is exogenously determined by the experimenter, workers' effort is correlated with wages irrespective of whether contracted quantity is above or below the same ability-determined threshold.

The rest of the paper is organized as follows. Section 2 places our study in the experimental literature on gift exchange. Section 3 outlines the theoretical model on which the experiment is based and Section 4 describes the experimental design and the procedures. Section 5 presents our results, and Section 6 concludes the paper.

2 Literature Review

Fehr, Kirchsteiger and Riedl (1993) conducted the first test of the gift exchange hypothesis by implementing a competitive labor market in the laboratory. In their experiment, firms bid for a single unit of labor and workers selected financially costly, non-contractible effort from a payoff table; the higher the worker's effort level, the higher the profitability of the firm. Since it is a dominant strategy for workers to select zero effort, in a subgame perfect equilibrium wages will equal workers' marginal product of zero effort. However, the study found that average wages were higher than the competitive equilibrium, and that workers' effort was positively correlated with wages.

This paper spawned a large literature in experimental economics, reviewed by Charness and Kuhn (2011), seeking to understand the behavioral drivers of gift exchange. The abstract nature of the experimental setup in the laboratory has allowed researchers to modify the original experimental design in order to study a number of determinants of gift exchange, which include social preferences (Fehr, Gächter and Kirchsteiger, 1997; Fehr, Kirchsteiger and Riedl, 1998; Charness, 2004), social norms (Fehr et al., 1998), market structure (Brandts and Charness, 2004), strategic delegation of wage decisions (Charness et al., 2012) or the role of reference points in wage contracts (Abeler et al., 2011).[1]

In contrast, evidence of gift exchange in the field is mixed. In field experiments, workers exert real effort, as opposed to picking financially costly effort levels. Importantly, they are set in a naturalistic environment and subjects are unaware of being in an experiment. In some cases, evidence supporting gift exchange is short-lived and gift giving by employers is inefficient, in the sense that it would have been more profitable to hire workers at market clearing wages (Gneezy and List, 2006; Bellemare and Shearer, 2009). In other cases, the effectiveness of gifts depends on the nature of the gift itself (Kube et al., 2012).[2]

[1]Gift exchange experiments are typically framed as an interaction between buyers and sellers, where subjects make choices with the aid of a payoff table. While evidence for gift exchange in the laboratory is reasonably robust, it is not immune to seemingly innocuous changes in the presentation of payoffs. Charness et al., (2004) find that presenting the payoffs to both worker and firm in a comprehensive payoff matrix, drastically reduces the extent to which workers engage in gift exchange. Using a similar design, Hannan et al., (2002) find similar results with undergraduates, but not MBA students.

[2]We note there is stronger evidence of gift exchange in the field in non-labor market settings — see for

In the laboratory, researchers have more control over the environment, but the task is abstract: picking a value from a payoff table.[3] Workers and firms exchange a single unit of labor, and there is little or no individual-level heterogeneity in parameters of interest, like worker productivity. Furthermore, the fact that effort is financially costly means workers can easily compute not only the extra surplus from effort, but also how that surplus is divided. Field experiments are more realistic, in that both sides of the market contractually agree on a quantity (e.g. hours worked in Gneezy and List, 2006; or timber in Bellemare and Shearer, 2011) in addition to wages, but they are very context-specific. It is extremely difficult for researchers to design an experiment in which economic parameters of interest, like market structure, are exogenously varied *ceteris paribus*. Also, although there is plenty of individual-level heterogeneity, it is often unobserved and therefore has to be estimated jointly with behavior (Bellemare and Shearer, 2011). Finally, workers often have no information about the production function of the firm and cannot compute the welfare consequences of non-contracted effort.

3 Theory

3.1 The game

Consider a two-stage market for a commodity, with B buyers and S sellers.[4] In the first stage of the market, buyers make proposals of the form (q, w), where $q > 0$ is the number of units of the commodity and $w > 0$ is the per-unit price. Output can be of low or high quality: a high-quality good is more valuable to the buyer, but is more costly for the seller to make. If a seller accepts a given proposal (q, w), he must first decide the proportion of high-quality goods to produce, $g \in [0, 1]$. Therefore, a market outcome is a set of triples of the form (q, w, g).

instance, Falk (2007) who studies this issue in the context of charitable contributions.

[3]We note that there are significant contributions in real effort lab experiments, but not many in the context of gift exchange, for instance Niederle and Vesterlund (2007) or Charness and Villeval (2009).

[4]We employed this nomenclature in the instruction sets and experimental materials to ensure a neutral framing, and we will retain it throughout the paper.

A player (either a buyer or a seller) earns zero payoff if there is no trade. If a buyer and a seller agree on a contract, payoffs to both parties are conditional on the contract being fulfilled. If a seller fulfills the contract, his payoff equals the piece rate, w, times the produced quantity, q, minus the cost of effort. The buyer's payoff is equal to an increasing concave function of quantity, minus the piece rate times the quantity.

If a seller does not manage to complete his contract, his payoff is equal to half of the piece rate, $w/2$, times quantity minus the cost of effort. In this case the revenue of the buyer would be a function of completed units, but the cost would be a function of the originally agreed contract. This contingent contract incentivizes buyers not to request too high a quantity, and it incentivizes sellers to complete the contract. The following equations express the payoffs to buyers and sellers.

$$\pi_B = \begin{cases} 0 \text{ if no trade} \\ \\ \sqrt{(1-g)q + bgq} - wq \text{ if } q = q^c \\ \\ \sqrt{(1-g)q + bgq} - wq^c \text{ if } q < q^c \end{cases} \tag{1}$$

$$\pi_S = \begin{cases} 0 \text{ if no trade} \\ \\ qw - c_l(1-g)q - c_h gq \text{ if } q = q^c \\ \\ (qw)/2 - c_l(1-g)q - c_h gq \text{ if } q < q^c \end{cases} \tag{2}$$

where $b > 1$ represents the relative value to the firm of high-quality goods over low-quality goods (in our experiment $b = 2$ in all treatments); $c_l > 0$ is the cost to the seller of producing low-quality output; $c_h > c_l$ is the cost to the seller of producing high-quality output; q is the produced number of units and q^c is the contracted number of units.

3.2 Equilibrium

We assume all players have identical preferences. There are always five players in a market, either three buyers and two sellers ($B = 3, S = 2$) or two buyers and three sellers ($B = 2, S = 3$). This implies there will always be an unmatched player. We will consider the general case where all players pay a tax τ on their earnings and receive transfers in the form of a lump sum payment, T, financed from the redistributive tax. Transfers are set at $T = \frac{\tau}{B+S}(B\pi_B + S\pi_S)$ in order to keep a balanced budget and are considered to be exogenous (we assume that agents do not take into account their contribution to the revenue used to finance the transfers they receive). Note that by setting $\tau = 0$, we are in the environment in which the majority of our treatments took place. In the interest of keeping the exposition of the material brief, we relegate proofs to the Appendix. We will assume utility functions which are linear in terms of players' payoffs: $U_i = (1-\tau)\pi_i + T$, where π_i is player i's payoff.

Proposition 1 *There will be no gift-exchange in equilibrium, irrespective of market structure. The equilibrium in a market where there is excess supply of workers ($S = 3$, $B = 2$) is given by (q^*, w^*, g^*), where $q^* = (\frac{1}{2c_l})^2, w^* = c_l, g^* = 0$. The equilibrium in a market where there is excess demand for workers ($S = 2$, $B = 3$) is given by (q^*, w^*, g^*), where $q^* = (\frac{1}{2c_l})^2, w^* = 2c_l, g^* = 0$. Furthermore:*

1. *An increase (decrease) in the cost of effort will lead to a fall (rise) in equilibrium quantity and a rise (fall) on equilibrium prices.*

2. *Equilibrium price and quantity will not change if workers do not have access to firms' payoff functions in a market where there is excess supply of workers ($S > B$).*

3. *Imposing a redistributive tax τ on payoffs will have no effect on equilibrium quantity or equilibrium prices.*

In summary, we have that: a) equilibrium quantity is inversely related to the cost of effort and independent of market structure; b) prices are positively related to the cost of effort and are higher when sellers have market power than when buyers have market power; c) there is no gift-exchange in equilibrium. In other words, the market always produces the efficient quantity. Market structure only affects the way surplus is divided between buyers and sellers.

4 The Experiment

In our experiment, each subject was randomly allocated to a group of five, which constituted a market. A market had, depending on the treatment, either three buyers and two sellers, or two buyers and three sellers. Subjects were informed their markets existed for at least 20 periods; from period 21 onwards, there was a 1/6 chance that the experiment would conclude at the end of each period. We did this in order to avoid end-game effects. We used the same sequence of random draws in all sessions to facilitate the comparison of results across treatments.

Each period started with a contract trading stage, where buyers posted contract offers specifying a quantity and a piece rate. Sellers could then accept the contract by clicking on a button on the computer screen — the first seller to accept a contract would get that contract. If a contract was not accepted within 15 seconds, the relevant buyer could post a new contract offer. The trading stage lasted for two minutes. Neither buyers nor sellers were given any identification codes, so reputation formation was not possible.

After the trading stage was over, sellers had to decide what percentage of the contracted output would be of high quality. To do that, sellers used an on-screen slider tool: the further to the right the slider was placed, the higher the percentage of high-quality output the seller committed to delivering. The default position of the slider was on the far left, equal to 0% high quality output.

Once sellers decided the fraction of high-quality and low-quality puzzles they would deliver in that period, they moved to the real-effort stage itself. Sellers had 30 seconds in which to deliver the contracted quantity. Each unit was an arithmetic operation: low-quality output was an operation of the form $x + y =?$, where $0 < x, y \leq 9$, and high-quality output was an operation of the form $x + y - z =?$, where $0 < x, y, z \leq 9$ and $x + y - z \geq 0$. Subjects were not explicitly told of the constraints on x, y and z.

Before the actual experiment started, all subjects were given two opportunities to

	Ctrl	AsymInf	Tax	SliRgt	FixPuz
BuyerPower	6	6	6	6	6
SellerPower	6	6	6		

Table 1: Experimental Design

practise each type of puzzle for 30 seconds, which was the duration of the real effort stage. They did so before finding out if they were a buyer or a seller in the experiment. The practice stage allowed subjects to understand their relative ability at solving each type of operation, and as a consequence, it gave us a direct measure of ability (or cost of effort) for each subject. Once the practice stage was over, the actual experiment began.

We considered a baseline condition, Ctrl, in which sellers had complete information about payoffs and no taxes. We modified our baseline condition in two ways. In one treatment, which we denote as Asym, sellers knew that high quality output was more valuable to firms than low quality, but they did not know firms' actual payoff function. The other treatment, which we denote as Tax, modifies Ctrl by introducing a pure redistributive tax.

We interacted the three aforementioned treatments with two market structures: BuyerPower, where there were two buyers and three sellers in each market; and SellerPower, where there were three buyers and two sellers in each market. Table 1 outlines the experimental design. The numbers in each cell indicate the number of independent markets we collected for each condition.

The purpose of the Asym treatment was to mimic the conditions faced by most workers, as well as most participants in field experiments on gift exchange. In most industries, workers are not aware of, or are unable to calculate the marginal product of labour, either because they lack access to the firm's accounting data, or because the firm's production function is too complex. This is potentially a serious obstacle to gift exchange in most models of fairness and/or reciprocity, because it prevents workers from calculating the 'bliss point' in their utility function which determines a kind offer. As such, we expected gift exchange to be significantly lower in Asym than in Ctrl in either market structure.

The purpose of the Tax treatment was to investigate the role of distributional preferences on gift exchange. A significant portion of the gift exchange literature (see Charness and Kuhn, 2011) has focused on inequality aversion as a source of explanation for the non-equilibrium effort and wage levels in experimental labor markets. To better understand the role of redistribution preferences, we took advantage of the fact that there is always one player who earns zero payoff in both market configurations of the Ctrl condition. We impose a pure redistribution tax, in which all players incur a 10% tax on their end-of-period payoffs, which are then redistributed equally across the five players in the market. This introduces additional incentives for inequality-averse sellers to produce more high-quality output: doing so increases the surplus from trade, which in turn increases the taxable income of buyers and therefore increases the payoff to the unmatched player, as well as their own. As such, we expected to observe higher levels of gift exchange under Tax than Ctrl.

The fact that we interact Tax with both market configurations also allows us to study the role of horizontal comparisons of welfare, that is, the concerns sellers may have for the welfare of other sellers, as opposed to buyers. Again, the existing literature has mixed evidence on this matter: while some studies find little evidence for horizontal welfare comparisons amongst workers (e.g. Charness and Kuhn, 2007; Maximiano et al., 2007), others find strong evidence of worker-worker comparisons of welfare (e.g. Abeler et al., 2010; Gächter et al., 2010).

In addition to the main treatments, we ran two additional controls in the BuyerPower condition. In the first control, named FixPuz, the number of puzzles workers could solve was fixed to 10. This treatment not only checks that the gift-exchange finding is robust to the variation in contracted quantity, but it also aids us to identify the role of reciprocity in determining gift exchange. In our setup, keeping wages fixed, by requesting low quantities, buyers signal their intentions regarding gift-exchange, to which sellers may then reciprocate with high-quality output. In contrast, buyers that request high quantities signal they do not expect any reciprocity from sellers. Comparing the behavioral response of sellers in the case where quantity is determined by buyers to the case where it is fixed, allows us to identify the role of reciprocity. The second control treatment was SliRgt, in which the default position of the slider when sellers chose the fraction of high-quality output to produce was set to 100%. This is to control for inertia in workers' decisions; that is, to control for the fact that workers may be more likely to pick 0% high-quality output in our main treatments, because the default option is to do so.

We conducted two separate sets of sessions. The first set concerned the main experiment as described above. Three months after finishing the sessions for the main experiment, we invited all participants to take part in another experiment, which was designed to collect auxiliary information. The second experiment consisted of a risk aversion elicitation task (Holt and Laury, 2002) and a set of mini-ultimatum games (Charness and Rabin, 2002). The purpose of the second experiment was to generate individual-level measures of risk and inequality aversion, respectively. We briefly describe the two tasks below. The methodology for the generation of these behavioral variables, is available in the Appendix.

In the risk elicitation task, subjects were presented with a sequence of pairs of lotteries of the form $A = (20ECU, p_i; 16ECU, 1 - p_i)$, $B = (38.5ECU, p_i; 1ECU, 1 - p_i)$, where $p_i \in \{0.1, 0.2, 0.3, \ldots, 0.9, 1\}$, and ECU stands for Experimental Currency Unit. The pairs of lotteries were presented in increasing order of p_i. If $p_i < 0.5$, lottery A has a higher expected value than lottery B, and the opposite is true if $p_i \geq 0.5$. We can infer a subject's aversion to risk by the point at which he switches from choosing lottery A to choosing lottery B. Subjects who switch when $p_i < 0.5$ are risk seeking, and those who switch when $p_i \geq 0.6$ are risk averse. From this task, we generated a variable which we will denote as $Risk_i$.

Each mini-ultimatum game consisted of a simple decision task in which every subject was given an option between two income allocations, each of which determined the payoff to himself and another person in the room. For example, a subject would have to choose between option A which gave 6.34 ECU to himself and 4 ECU to another person in the session, or option B which gave 5.5 ECU to himself and 5.5 ECU to the other person. The different choices systematically vary the overall size of the pie being split and the differences in payoffs between the dictator and recipient. By examining individual options, we were able to construct a measure of an individual's aversion to advantageous and disadvantageous inequality — we will denote respectively these variables as AIA_i and DIA_i.

Subjects did not know they would be invited to a second experiment when taking part in the main experiment, and participation in the second experiment was voluntary. We did this for two reasons: firstly we wished to avoid the potential problem of behaviour in the main task influencing behaviour in the second experiment, like hedging (Blanco et al., 2010);

secondly, pooling the two tasks would have made the experimental sessions exceedingly long, which leads to participant fatigue and is generally considered bad practice (Friedman and Sunder, 1994). However, this approach entails the risk of attrition in the rate of participation: indeed, 37 participants (approximately 15% of our sample) did not take part in the second experiment.

We recruited 240 participants through our lab's ORSEE recruitment database (Greiner, 2004). In both sets of sessions, participants sat at a computer booth upon arriving at the lab. Participants had 10 minutes in which to read the instructions. The instructions for the main experiment included a short quiz to check that participants understood the instructions. Participants made all decisions and received all the feedback through a computer terminal using z-Tree (Fischbacher, 2007). The main experiment's sessions lasted on average 90 minutes and the second experiment's sessions lasted on average 60 minutes. In both sets of experiments, participants were paid in cash at the end of the session. The average payment was £16.63 ($26.87) for the first experiment, and £12.79 ($20.67) for the second experiment.

5 Results

We begin by briefly examining the agreed contracts in each condition, and then we move to our analysis of gift exchange.

5.1 Contracts

Table 2 displays average contracted wages and average contracted number of puzzles in each of the seven treatments. We find a large and significant effect of market configuration on average contracted wages. Wages are significantly higher in markets with an excess demand for sellers (i.e., SellerPower) for all treatments (Ctrl: $p = 0.004$; Asym: $p = 0.004$; Tax: $p = 0.016$, Mann-Whitney (henceforth MW) test). Excess demand for sellers leads to lower average contracted quantity. This effect is less pronounced than for price: we find a marginally significant effect on Asym ($p = 0.077$, MW test), but not on Ctrl or Tax ($p = 0.262$ and $p = 0.337$ respectively, MW test). This is consistent with Proposition 1.

	SellerPower			BuyerPower				
	Ctrl	AsymInf	Tax	Ctrl	AsymInf	Tax	FixPuz	SliRgt
w	20.64	21.67	25.96	7.29	8.54	12.19	12.06	7.74
	(3.22)	(5.90)	(8.24)	(3.43)	(1.23)	(6.81)	(4.51)	(3.44)
q	10.97	9.73	9.75	11.68	11.84	10.39	10	11.91
	(0.67)	(1.98)	(1.07)	(1.82)	(1.23)	(1.44)	-	(1.38)
N	6	6	6	6	6	6	6	6

Standard deviations in parenthesis. N denotes number of independent markets.

Table 2: Average contracted quantity and price.

Keeping market configuration fixed and looking at treatment effects, when there is excess demand for sellers (SellerPower), we find no statistical difference either between Ctrl and Asym or between Ctrl and Tax. When there is excess supply of sellers (BuyerPower), we see significant differences in wages when we compare Ctrl to Asym and Tax ($p = 0.062$ and $p = 0.056$ respectively, MW test). Finally, we find no significant difference between average wage in FixPuz and Ctrl, when we restrict the quantity in the latter to be the same as in the former, i.e., 10 ($p = 0.200$, MW test). We find no difference in average wages between Ctrl and SliRgt ($p = 0.873$, MW test). We do not observe any difference in average contracted quantity between any pair of treatments in either market configuration. Figure 1 shows scatter plots of the average (price, quantity) contracts for Ctrl, Asym and Tax, in both market structures. It confirms the finding from Table 2 that average prices are significantly lower when there is excess supply of sellers. When there is excess supply of sellers, most markets specify on average contracts with quantities in the region of 12 units, while when there is excess demand for sellers, most markets specify on average contracts with quantitates between 8 and 10 units in Asym and around 11 units in Ctrl.

Observation 1: *The principal driver of contract formation is market structure: prices are higher in markets with excess supply of labour. Neither a pure redistribution tax nor absence of information on buyers' payoffs have a consistent effect on contracted quantity and/or*

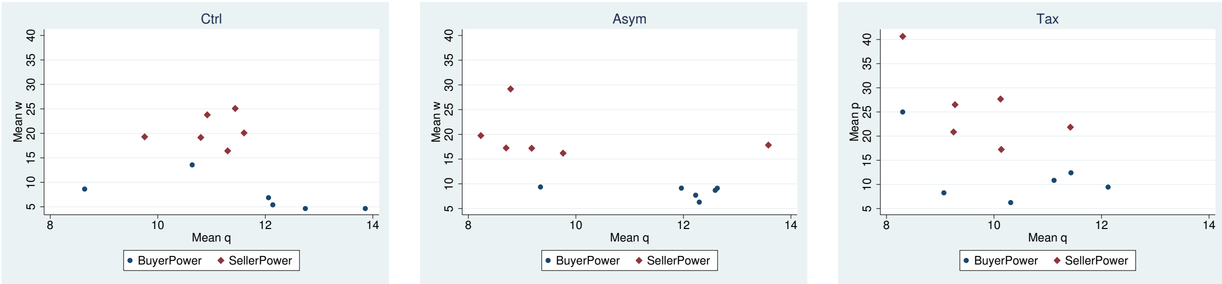


Figure 1: Scatter plots of average contracted quantity and price.

price.

We explore further the nature of agreed contracts by examining the determinants of the piece rate. Our model predicts a flat relationship between quantity and price: buyers should request the maximum feasible quantity at any given price. To test this prediction, we estimated the following equation.

$$w_{it} = (1 + q_{it})D\beta_1 + X_{it}\beta_2 + t\beta_3 + u_{it} \tag{3}$$

where w_{it} is the agreed price, q_{it} is the agreed quantity, D is a vector of treatment dummy variables, X_{it} is a vector of individual-specific characteristics, t is a time trend, and β_1, β_2 and β_3 are the vectors of coefficients to be estimated, and u_{it} is an error term. We note that the nature of the experimental design imposes some restrictions: on one hand, we have usually two observations per market per period. On the other hand, in treatments with three buyers and two sellers, we always have an unmatched buyer, which means we have multiple missing observations for buyers in those treatments. In either case, we cannot set up a proper panel when performing any econometric analysis of our data. We deal with this by estimating a standard OLS model and clustering observations at the market level.

Regression (1) in Table 3 summarises the estimation results of the restricted model ($\beta_2 = 0$).[5] We observe large, positive and significant intercept coefficients for all treatments, as well as negative and significant coefficients on q_{it} for each treatment. In other words, contracts agreeing to a low quantity are typically contracts that offer a generous piece rate,

[5]We report results on the restricted and unrestricted models because we lose some observations when we include individual characteristics, since not all subjects took part in the second experiment.

DV: w	(1)		(2)	
q × Ctrl × BuyerPower	-0.509***	(0.118)	0.040	(0.442)
q × Asym × BuyerPower	-1.262***	(0.308)	-0.862**	(0.321)
q × Tax × BuyerPower	-1.300*	(0.694)	-0.531	(0.386)
q × SliRgt × BuyerPower	-0.919**	(0.346)	-0.229	(0.190)
q × Ctrl × SellerPower	-0.540***	(0.185)	-0.967*	(0.543)
q × Asym × SellerPower	-0.772*	(0.419)	-0.883	(0.652)
q × Tax × SellerPower	-1.818*	(1.000)	-0.641	(0.399)
Ctrl × BuyerPower	15.686***	(1.694)	10.066	(7.665)
Asym × BuyerPower	25.931***	(3.789)	21.146***	(4.886)
Tax × BuyerPower	28.134***	(9.639)	18.110***	(4.744)
SliRgt × BuyerPower	21.140***	(4.982)	11.619***	(3.015)
Ctrl × SellerPower	29.039***	(2.889)	35.520***	(8.035)
Asym × SellerPower	29.632***	(5.282)	31.851***	(8.151)
Tax × SellerPower	46.140***	(11.995)	33.317***	(4.571)
DIA_i			-4.748	(6.612)
AIA_i			3.297	(2.567)
Risk aversion$_i$			-0.198	(0.297)
Male$_i$			1.468	(0.999)
Period	-0.187**	(0.083)	-0.253***	(0.073)
R^2	0.761		0.817	
N	2,082		1,196	

Market level clustered standard errors in parenthesis.

***, **, *: statistical significance at 1%, 5% and 10% level.

Table 3: OLS estimates of relationship between contracted quantity and price.

and contracts demanding large quantities specify ungenerous piece rates. Furthermore, we observe a negative time trend, indicating a decrease in contracted wages over time. Adding buyer-specific risk aversion, inequality aversion and gender leads to some loss of significance in the quantity interaction variables. However, since none of the individual-level coefficients is significant, it is not possible to attribute the change in significance in the quantity variables to the fact that we are estimating the restricted model. A joint-test of significance of the new coefficients only marginally rejects the null of non-significance ($F(4, 38) = 2.23, p = 0.084$).

In either regression, fixing market structure, there are no significant differences between the intercept coefficients on CTRL and ASYM or between the coefficients on CTRL and TAX, which confirms Observation 1. We can therefore rule out productivity differences across different treatments as an explanation for the differences in average contracts. Keeping treatment conditions fixed and testing for the effect of market structure, we do observe a significant difference in the coefficients on CTRL ($F(1, 38) = 7.51, p = 0.009$), as well as on TAX ($F(1, 38) = 6.70, p = 0.014$), though not in ASYM. In summary, the primary determinant of contract formation is competitive pressures, rather than individual differences in social or risk preferences. However, this effect is diminished when sellers do not know the payoff function of the buyers, and therefore are unable to compute the gains from trade.

Observation 2: *We find little evidence to suggest other-regarding preferences play a role in determining prices and quantity.*

5.2 Output Quality

The main unit of analysis of seller behavior is the fraction of high quality output produced. However, analyzing seller behavior in a setting where quantity is contracted upon and depends on real effort requires some care. The first issue we need to address is whether or not contracts were feasible. To do so, we look at the data from the practice periods: recall that every subject, regardless of being a buyer or a seller, had two consecutive practice periods in which they could solve as many $x + y$ puzzles (i.e., low-quality output) as they could, followed by two consecutive practice periods in which they could solve as many $x + y - z$ puzzles (i.e., high-quality output) as they could. In all four practice periods, the time limit

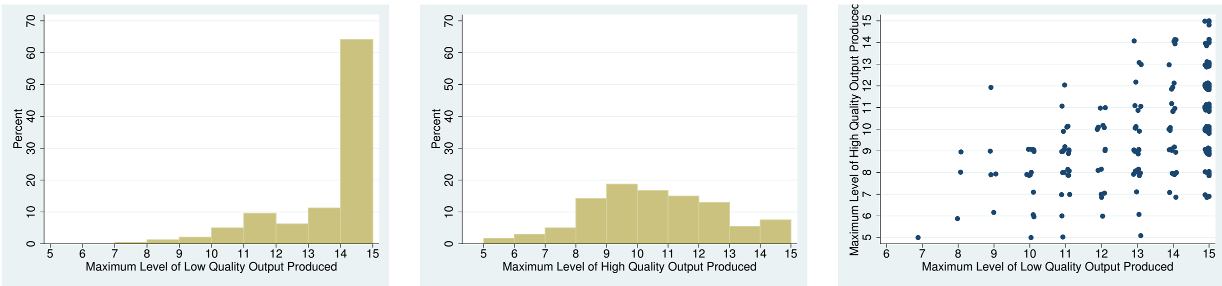


Figure 2: Histogram and scatter plots of maximum low- and high-quality output solved in practice stage.

was the same as in the real experiment, and subjects were paid a piece rate of £0.20. We now show some summary statistics on the highest number of puzzles of each type solved by subjects.

On average, our 240 subjects solved 13.60 (s.d. 1.89) low-quality puzzles, and 10.07 (s.d. 2.15) high-quality puzzles. Figure 2 displays the histograms of maximum low- and high-quality output produced in the practice stage, as well as the scatter plot of low- and high-quality output produced at the individual level. Over 50% of all subjects could solve 15 low-quality puzzles, which was the highest allowed number of puzzles subjects could solve in a given period. Furthermore, 91% of subjects could solve 11 or more low-quality puzzles, which was the average quantity across all treatments. Given that subjects should improve their performance as the experiment progressed, it is very unlikely that subjects would not be able to complete any given contract during the experiment. In contrast, only 41% of subjects could solve 11 or more high-quality puzzles. The right panel of Figure 2 also reveals the degree of heterogeneity in our sample: for a given level of ability in solving low-quality puzzles, there is great heterogeneity in high-quality puzzle ability – however, the correlation between the two measures is positive and significant (Spearman's $\rho = 0.509, p < 0.001$).

Table 4 displays the average fraction of high-quality output, for each treatment in our experiment. Keeping treatments constant and varying market structure, we do not observe any significant difference in the fraction of high-quality output produced (Ctrl: $p = 0.262$, Asym: $p = 1.000$, Tax: $p = 0.423$, MW test). Fixing market structure and varying treatments, in BuyerPower, we see a marginally significant difference between Ctrl and

	Ctrl	Asym	Tax	SliRgt	FixPuz
BuyerPower	0.250	0.152	0.381	0.297	0.213
	(0.108)	(0.097)	(0.120)	(0.075)	(0.128)
SellerPower	0.311	0.224	0.370		
	(0.118)	(0.195)	(0.223)		

Standard deviations based on market-level averages in parenthesis.

Table 4: Average fraction of high-quality output

Tax ($p = 0.078$), though not between Ctrl and Asym ($p = 0.199$). In SellerPower we see no difference either between Ctrl and Tax ($p = 0.423$), or between Ctrl and Asym ($p = 1.000$). Figure 3 provides a time series of average fraction of high quality output produced in each of the three treatments in both market configurations. In addition to broadly confirming the information in Table 4, it also shows that there is no consistent time trend in any treatment. Importantly, the average fraction of high-quality output remains positive and never declines to zero in any treatment.

Observation 3: *We find high-quality output is produced in all treatments.*

5.3 Gift Exchange

We now turn to the main questions of the experiment: do we observe gift exchange in markets where output is contracted upon, and how does that gift exchange vary with our treatments? For the purposes of the present experiment, we define gift exchange as a positive correlation between the fraction of high-quality output produced by sellers and the contracted wage. Since our primary measure of seller behavior is a variable between zero and one, it is possible that the optimal choice made by subjects is a corner solution (i.e., no gift exchange or full gift exchange). Hence, we may observe mass points at either end of the distribution of the fraction of high-quality output produced by sellers. Therefore, the OLS estimator is not appropriate, and we should instead employ a corner-solution model (Wooldridge, 2010). In particular, we opted for a two-limit Tobit model, which models the dependent variable, g^*_{it} as a censored case of a latent variable, g_{it}, where $g_{it} = 0$ if $g^*_{it} \leq 0$, $g_{it} = g^*_{it}$ if $0 < g^*_{it} < 1$,

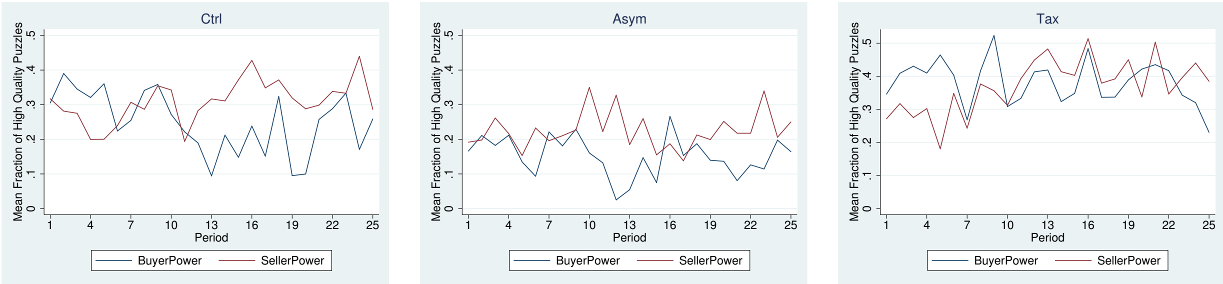


Figure 3: Time series of mean fraction of high-quality output produced.

and $g_{it} = 1$ if $g^*_{it} \geq 1$.[6]

Our first specification considers the effect of agreed price and quantity on gift exchange, where both price and quantity variables interact with a set of treatment dummies:

$$g^*_{it} = (1 + q_{it} + w_{it})D\beta_1 + X_{it}\beta_2 + t\beta_3 + \varepsilon_{it} \tag{4}$$

where q_{it} is output, w_{it} is price, D is a vector of treatment dummy variables, X_{it} is a vector of individual-specific variables, which we describe below, t is a time trend and β_1, β_2 and β_3 are vectors of coefficients to be estimated; ε_{it} is an error term.

Column (3) in Table 5 summarizes the outputs of a restricted version of the model described above, by setting $\beta_2 = 0$. For ease of exposition, we express the coefficients on treatment effects directly, rather than differences from an omitted category. We start by examining the coefficients on the intercept dummy variables. Their economic interpretation is the expected fraction of high-quality output a seller would be willing to provide for free. We find positive and significant coefficients only on Ctrl × BuyerPower, and both Tax conditions. While the former case is somewhat surprising, especially since the coefficient on Ctrl × SellerPower is not significant, the latter two cases are somewhat expected. In the Tax conditions there is a benefit, for sellers with inequality-averse or maximin preferences, of increasing surplus by producing high-quality output, as they are able

[6]While the Tobit model is a popular application for data which follow distributions that are bounded, some argue against its use, as it is really meant for cases in which the distribution is censored: that is, unobservable above and/or below a particular set of values. We considered as an alternative the QML regression model proposed by Papke and Wooldridge (1996) to tackle fractional response variables. The results are robust to this alternative specification, and we include them in the appendix to this paper.

DV: g	(3)		(4)	
w × Ctrl × BuyerPower	0.014**	(0.006)	0.009	(0.006)
w × Asym × BuyerPower	-0.002	(0.004)	-0.004	(0.005)
w × Tax × BuyerPower	0.0001	(0.003)	0.002	(0.007)
w × FixPuz × BuyerPower	0.010**	(0.005)	0.022**	(0.011)
w × SliRgt × BuyerPower	0.011*	(0.006)	-0.009	(0.012)
w × Ctrl × SellerPower	0.007	(0.005)	0.011**	(0.006)
w × Asym × SellerPower	0.020***	(0.008)	0.017**	(0.006)
w × Tax × SellerPower	0.016***	(0.003)	0.003	(0.006)
q × Ctrl × BuyerPower	-0.070***	(0.024)	-0.068***	(0.022)
q × Asym × BuyerPower	-0.012	(0.031)	0.004	(0.058)
q × Tax × BuyerPower	-0.100***	(0.013)	-0.087***	(0.019)
q × SliRgt × BuyerPower	-0.011	(0.031)	-0.072*	(0.043)
q × Ctrl × SellerPower	-0.063	(0.041)	-0.072***	(0.025)
q × Asym × SellerPower	-0.098***	(0.035)	-0.135***	(0.037)
q × Tax × SellerPower	-0.106***	(0.038)	-0.087**	(0.040)
Ctrl × BuyerPower	0.774***	(0.270)	0.569	(0.467)
Asym × BuyerPower	0.009	(0.298)	-0.520	(0.729)
Tax × BuyerPower	1.383***	(0.139)	0.945**	(0.449)
FixPuz × BuyerPower	-0.150	(0.135)	-0.720	(0.448)
SliRgt × BuyerPower	0.175	(0.365)	0.437	(0.605)
Ctrl × SellerPower	0.791	(0.609)	0.485	(0.457)
Asym × SellerPower	0.529	(0.415)	0.816	(0.537)
Tax × SellerPower	0.964***	(0.370)	0.700	(0.588)
Ability$_i$			0.005	(0.026)
DIA$_i$			0.660	(0.495)
AIA$_i$			0.582**	(0.227)
Risk aversion$_i$			0.022	(0.036)
Male$_i$			-0.003	(0.054)
Period	-0.004	(0.003)	-0.002	(0.003)
σ	0.638	(0.056)	0.532	(0.053)
N	2,380		1,512	

Market level clustered standard errors in parenthesis.

***, **, *: statistical significance at 1%, 5% and 10% level.

Table 5: Tobit regressions of determinants of high-quality output – equation (4)

to increase the payoff to the unmatched player this way (in addition to the direct benefit from an increased tax transfer). Although the coefficient on Tax $\times$ BuyerPower is nominally larger than the coefficient on Tax $\times$ SellerPower, that difference is not significant ($F(1, 2061) = 0.99, p = 0.320$). We find negative coefficients on all interactions between quantity, q, and treatment dummies, although they are not all statistically significant. At face value, these coefficients indicate a non-linear trade-off between efficiency and fairness: in contracts specifying large quantities, sellers are unwilling to further increase surplus by producing higher-quality output, since that extra surplus goes wholly to the buyer (and effort is costly).

We now now turn to the relationship between the piece rate and the fraction of high-quality output keeping quantity fixed, which is our primary measure of gift exchange. In our regression model, this is captured by the coefficients on the interaction of w with each treatment dummy variable. We find strong evidence of gift exchange in the SellerPower - Tax and SellerPower - Asym conditions, and little evidence of gift exchange in the BuyerPower conditions — only present in the BuyerPower - Ctrl, BuyerPower - FixPuz and BuyerPower - SliRgt treatments.

Column (4) summarizes the estimates of the unrestricted model, which includes a set of individual-level characteristics: a proxy for cost of effort, *Ability*, equal to the largest number of high difficulty puzzles that subject solved in the practice stage; a measure of each subject's aversion to disadvantageous and advantageous inequality (DIA_i and AIA_i respectively); a measure of risk aversion, $Risk_i$, and a gender dummy, $Male_i$. The addition of the individual-specific variables marginally significantly increases the goodness-of-fit of the model (Wald test: $F(6, 1483) = 1.81, p = 0.093$). Once we account for individual differences in distributional preferences, all (but one) intercept dummies cease to be significant, and all but three coefficients on price become statistically insignificant as well. This suggests that inequality aversion accounts for some gift exchange behavior, but it primarily explains baseline levels of gift giving: in other words, inequality aversion drives not only high-quality output that is generated as a result of higher piece rates, but high-quality output that is generated unconditionally. Importantly, only one of the control treatments in the BuyerPower treatments has a significant price interaction coefficient. This confirms the finding from the restricted model that market structure has an important impact on gift exchange. Interestingly, the coefficient on the variable measuring individuals' aversion to advantageous inequality is positive and significant ($0.582, p = 0.010$). The effect of contracted quantity remains significant in most treatments.

Observation 4: *We find a negative relationship between contracted quantity and high-quality output produced.*

This suggests we should condition the analysis of gift exchange on the level of contracted quantity in a non-linear way, since there may be a range of contracted quantities for which a seller is unable to execute 100% high-quality output in the allotted time for the task. Not taking this into account could lead us to under-estimate the extent to which sellers reciprocate high wages with a high fraction of high- quality output. Looking at the middle histogram of high-quality puzzles solved in Figure 2, we see there is some heterogeneity in ability. In particular, only 41% of subjects would have been able to solve 11 or more high-quality output, and as we can see from the scatter plot on the right of Figure 2, there was some heterogeneity in relative ability to solve high-quality output relative to low-quality output. We will exploit this heterogeneity in ability in our analysis.

In particular, we distinguish contracts that specify quantities that allow for sellers to complete 100% high-quality output from those which do not. This in turn requires us to define a threshold quantity, which we denote as q_i^{FGE}. We assume that a seller should be able to complete 100% high-quality output if the agreed contract specifies a quantity below q_i^{FGE}. We define q_i^{FGE} as the maximum number of high-quality puzzles a subject completed in the two practice rounds prior to the actual experiment commencing.[7]

Large quantity demands could signal buyers do not expect a great deal of high-quality output to be produced, either because they believe the seller cannot provide high-quality output due to low ability (or conversely, high cost of effort), or because the buyer believes sellers will not want to provide high-quality output. In particular, we distinguish between levels of contracted quantity which allow for sellers to produce 100% high-quality output, and

[7] As we mentioned earlier, it is possible that subjects become better at solving puzzles over the course of the experiment, which would shift q_i^{FGE} up. As such, this is a conservative measure of this threshold.

those which do not. In other words, we wish to understand whether we observe a stronger evidence for gift exchange when full gift exchange is possible. As such, we consider a different specification, in which we replace the contracted quantity, q_{it}, by a dummy variable, FGE, which equals one if $q_{it} \leq q_i^{FGE}$ and zero otherwise; and in addition, an interaction between FGE and w_{it}, which is summarized in the equation below.

$$g_{it}^* = (1 + w_{it} + w_{it}FGE)\, D\beta_1 + X_{it}\beta_2 + \varepsilon_{it} \tag{5}$$

Column (5) in Table 6 summarizes the estimation results of the restricted version (i.e., where $\beta_2 = 0$) of the new model. We start by looking at the evidence from contracts that allow sellers to produce 100% high-quality output. In this case, there is some evidence of gift exchange. That is, we observe positive and significant coefficients on the interaction between FGE, w and the relevant treatment. In particular, we observe evidence of gift exchange in both Ctrl treatments, irrespective of market conditions. The evidence is nominally stronger in BuyerPower (0.040) than in SellerPower (0.022), but the difference is not significant ($F(1, 2061) = 2.60, p = 0.107$). In the BuyerPower condition, we do not observe evidence of gift exchange either in the Asym (0.0004, p=0.970) or the Tax (0.007, p=0.136) treatment. In contrast, there is strong evidence of gift exchange in the SellerPower - Asym (0.020, p=0.006) and SellerPower - Tax (0.016, p=0.047) treatments. This evidence is consistent with results from the regression analysis of Table 5. When sellers cannot produce 100% high-quality output, we find no evidence of gift exchange in any treatment. In other words, the coefficients on the interactions between w and the relevant treatment dummy are not statistically significant for any treatment.

Observation 5: *We find stronger evidence of gift exchange in markets with excess demand for labor than in markets with excess supply of labor.*

As in the previous set of estimations, we also consider an extended model, which accounts for individual-level characteristics, as well as a time trend. Column (6) summarizes the results from the extended model estimation. We first note that, as before, the unrestricted model has an improved fit (Wald test, $F(5, 1483) = 2.54, p = 0.027$). Most coefficients retain their sign and significance. Some exceptions are worth noting: the first is that the coefficient

DV: g	(5)		(6)	
$FGE \times w \times$ CTRL $\times$ BUYERPOWER	0.040***	(0.006)	0.041***	(0.010)
$FGE \times w \times$ ASYM $\times$ BUYERPOWER	0.0004	(0.011)	-0.022*	(0.013)
$FGE \times w \times$ TAX $\times$ BUYERPOWER	0.007	(0.004)	-0.001	(0.006)
$FGE \times w \times$ FIXPUZ $\times$ BUYERPOWER	0.013	(0.012)	0.019***	(0.005)
$FGE \times w \times$ SLIRGT $\times$ BUYERPOWER	0.004	(0.003)	-0.008	(0.009)
$FGE \times w \times$ CTRL $\times$ SELLERPOWER	0.022**	(0.010)	0.022***	(0.008)
$FGE \times w \times$ ASYM $\times$ SELLERPOWER	0.020***	(0.007)	0.026***	(0.010)
$FGE \times w \times$ TAX $\times$ SELLERPOWER	0.016**	(0.008)	0.017	(0.011)
$w \times$ CTRL $\times$ BUYERPOWER	0.001	(0.004)	-0.003	(0.007)
$w \times$ ASYM $\times$ BUYERPOWER	9.78×10^{-6}	(0.005)	0.009	(0.005)
$w \times$ TAX $\times$ BUYERPOWER	0.003	(0.006)	0.008*	(0.004)
$w \times$ FIXPUZ $\times$ BUYERPOWER	0.004	(0.007)	0.014**	(0.006)
$w \times$ SLIRGT $\times$ BUYERPOWER	0.012	(0.008)	-0.002	(0.009)
$w \times$ CTRL $\times$ SELLERPOWER	-0.003	(0.007)	0.004	(0.006)
$w \times$ ASYM $\times$ SELLERPOWER	0.010	(0.011)	0.002	(0.010)
$w \times$ TAX $\times$ SELLERPOWER	0.009	(0.007)	0.001	(0.009)
CTRL $\times$ BUYERPOWER	-0.118	(0.074)	-0.295	(0.307)
ASYM $\times$ BUYERPOWER	-0.203	(0.150)	-0.492	(0.347)
TAX $\times$ BUYERPOWER	0.225**	(0.103)	0.002	(0.271)
FIXPUZ $\times$ BUYERPOWER	-0.250***	(0.092)	-0.818**	(0.402)
SLIRGT $\times$ BUYERPOWER	-0.029	(0.113)	-0.445	(0.280)
CTRL $\times$ SELLERPOWER	-0.030	(0.116)	-0.409	(0.311)
ASYM $\times$ SELLERPOWER	-0.532***	(0.124)	-0.544**	(0.263)
TAX $\times$ SELLERPOWER	-0.144	(0.113)	-0.309	(0.313)
DIA_i			0.696	(0.509)
AIA_i			0.567***	(0.217)
Risk aversion_i			0.034	(0.037)
Male_i			-0.016	(0.051)
Period			-0.004	(0.003)
σ	0.651	(0.055)	0.547	(0.053)
N	2,380		1,512	

Market-level clustered standard errors in parenthesis.

***, **, *: statistical significance at 1%, 5% and 10% level.

Table 6: Tobit regressions of determinants of high-quality output – equation (5)

on $FGE \times w_{it} \times$ ASYM $\times$ BUYERPOWER is now surprisingly negative and marginally significant $(-0.022, p = 0.078)$. This confirms the results from regression (4), and is our next observation.

Observation 6: *Gift exchange disappears when sellers do not know the production function of buyers. and there is excess supply of sellers.*

The second case worth noting is that the coefficient on $FGE \times w_{it} \times$ TAX $\times$ SELLER-POWER (0.017, p=0.117) is also no longer significant. In other words, once individual-level characteristics are accounted for, gift exchange is crowded out by a purely redistributive tax. This confirms our earlier finding from Table 5, that fairness preferences primarily determine the default fraction of high-quality output sellers will produce irrespective of the price offered by buyers.

Observation 7: *Gift exchange is crowded out by a pure redistribution tax, irrespective of market structure.*

We can also investigate the role fairness and reciprocity have in determining gift exchange. We do so by examining the role of output in signaling buyers' kindness towards sellers. When contracted output is endogenously determined, its level is an indication of whether or not buyers expect gift exchange: by proposing very high levels of output, buyers signal they do not expect much reciprocity from sellers. However, that signaling ability is absent when output is determined exogenously. In that case, we should expect gift exchange to occur regardless of whether or not the contracted output is in excess of sellers' threshold ability, q^{FGE}. This reasoning suggests we should find a positive and significant coefficient on $FGE \times w_{it} \times Ctrl \times$ BUYERPOWER but not on $w_{it} \times Ctrl \times$ BUYERPOWER, while we should see positive and significant coefficients on both on $FGE \times w_{it} \times Ctrl \times$ FIXPUZ *and* $w_{it} \times Ctrl \times$ FIXPUZ. The evidence supports our hypothesis: we observe a positive and significant interaction of w_{it} with FGE in the CTRL - BUYERPOWER condition $(0.041, p < 0.001)$, and an insignificant coefficient on the interaction of w_{it} with CTRL - BUYERPOWER $(-0.003, p = 0.624)$. In contrast, the interaction coefficients of FIXPUZ with w are positive and significant (respectively, $0.019, p < 0.001$ and $0.014, p = 0.016$). This indicates reciprocity concerns underpin gift exchange, whenever it is present in our data. Since the regression results in Table 5 already indicate some high-quality output is produced irrespective of price, and that inequality aversion likely accounts for such behavior, we summarize our next observation as follows.

Observation 8: *Inequality aversion primarily determines baseline levels of high-quality output produced, while reciprocity concerns primarily affect gift exchange.*

6 Conclusion

The findings from our experiment help us reconcile much of the conflicting evidence from lab and field experiments using a single experimental design. In addition to differences in the nature of the task between lab and field, most, if not all field experiments on gift exchange to date have focussed on low skill jobs in which there is an almost infinitely inelastic labor supply. Furthermore, workers in these settings have very little knowledge of the production function of the firm, which does not allow them either to compute the gains from trade, or to infer how that surplus is split between themselves and the firm as a function of their effort. In contrast, lab "workers" typically have full information about the profit function of the firm, which combined with financially costly effort, enables them to easily calculate the costs and benefits of reciprocal behavior.

We find it is the combination of competitive pressures and absence of information about payoffs in our experiment that leads to the absence of gift exchange. In markets where there is excess *supply* of labor, removing information about firms' production function leads to the collapse of gift exchange: firms are no longer able to signal their intentions to workers, as the latter cannot evaluate how kind a contract offer is, because the surplus from trade is unknown. As such, firms make very high output demands at low prices, which lead to no gift exchange. In contrast, in markets where there is excess *demand* for labor, the competition for workers leads firms to post contract offers specifying low quantities and high wages. This in turn reduces the cost of reciprocity, as it is easier for workers to fulfill these contracts with a high percentage of high quality output. Market structure therefore plays an important role in determining gift exchange in markets where labor output is contractible: if taken at face value, our findings indicate that the role of gift exchange may be very industry-specific. In reality, the production function of the firm is either unknown or too difficult to infer by its workers, so we are less likely to observe gift exchange in industries where there is excess labor supply.

By exploiting a richer environment, in which both sides of the market contract quantity and price, we are also able to separate cleanly the roles of inequality aversion and reciprocity in determining gift exchange, whenever it is present. We show that inequality aversion primarily determines the baseline proportion of high-quality output produced in by sellers, irrespective of the wages on offer, rather than gift exchange itself. In our experiment, the more high-quality output is produced, the bigger the surplus from trade will be. However, there will also be higher inequality in income between the two agents, since all the extra surplus goes directly to the buyer. When we regress the fraction of high-quality output on contracted output and price, the coefficients on the intercept and the quantity slopes are therefore a measure of the ideal distribution of income from the perspective of sellers for any given output level. The positive intercept terms and negative coefficients on output indicate that, for any given price, sellers are increasingly inequality averse as the contracted output increases. When we introduced individual-level estimates of inequality aversion, most intercept coefficients cease to be significant at standard levels, while not affecting most of the wage coefficients. This provides further evidence supporting the idea that inequality aversion affects the baseline level of high-quality output, but not gift exchange itself.

In line with this interpretation of behavior, the introduction of a pure redistribution tax resulted in a crowding out of gift exchange. If gift exchange is driven by inequality aversion, then we should have observed an increase in gift exchange in the presence of the tax: doing so decreases inequality between buyers and sellers, since it transfers some of the extra surplus from the former to the latter; it also decreases inequality between the four active players in the market and the fifth player who is unmatched. Therefore, any seller who displays inequality aversion, efficiency or maximin preferences should be more responsive to wage levels. Instead, we observe an increase in the intercept coefficient, which as argued above, is a proxy for distributional preferences.

Instead, through a simple experimental manipulation we observe that gift exchange is determined primarily by reciprocity motivations. We exploit the fact that since output is determined endogenously in our experiment, it can work a signal of buyers' intentions to engage in gift exchange. In particular, we would expect sellers to reciprocate more readily to higher wages when contracts demand low levels of output, because such sellers will find it easier to fulfill such contracts with high-quality output. However, if output is determined by the experimenter instead, the signaling value of the output level is lost, and we should observe a correlation between wages and high-quality output independently of whether the quantity specified is too high or not. We use an individual-level measure of ability to distinguish contracts which allow full gift exchange (i.e., 100% high-quality output) from those which do not. As predicted, when contracted output is endogenous, we observe a significant correlation between the proportion of high-quality output produced and wages in contracts which allow for full gift exchange, and no such correlation in contracts that specify too high a quantity. However, when quantity levels are determined exogenously, we observe a correlation between wages and high-quality output for both types of contracts.

Our results provide a causal link to some of the field-based findings of Krueger and Mas (2004), who study the relationship between labor strife in a tire manufacturer and product quality. Following the takeover of Firestone by Bridgestone in 1988, the new management proposed a new labor agreement with the workers, which included a change in the way workers supplied contractible quantity: production was moved from 8-hour shifts to 12-hour shifts which rotated between daytime and nighttime, as well as a reduction of benefits to hourly workers. The authors found there was a higher fault rate in the tires produced *after* the start of the labor negotiations but *before* the strike (which cannot therefore be attributed to staffing changes), than the tires produced before the negotiations began. While the authors are naturally cautious in the interpretation of their results, in light of our data, we view them as evidence that the new proposed contract by management signaled their intent to extract further surplus from workers, to which workers responded by producing lower quality (non-contractible) effort, which in turn caused higher fault rates in tires.

In short, by implementing a real-effort labor market in the laboratory, we are able to determine the mechanisms through which outcome-based and intentions-based fairness considerations operate in an environment in which buyer-seller interactions are short-lived and fully anonymous, thus not allowing for any reputation formation. While fairness preferences seem to affect the degree to which sellers are willing to transfer surplus to buyers through costly, non-monetary effort in order to achieve some optimal trade-off between efficiency and inequality, reciprocity operates on the basis of whether sellers reward truly kind offers, which are contracts that specify low amounts of output that involve lower risk of non-completion, should the seller choose to produce a high proportion of high-quality output.

References

[1] Abeler, J., Altmann, S, Kube, S. and Wibral, M. 2010. "Gift exchange and workers' fairness concerns: When equality is unfair." *Journal of the European Economic Association* 8(6), 1299-1324.

[2] Abeler, J., Falk, A., Goette, L., and Huffman, D. 2011. " Reference points and effort provision." *American Economic Review* 101, 470-492.

[3] Agell, J. and Lundborg, P. 1995. "Theories of pay and unemployment: Survey evidence from Swedish manufacturing firms. *Scandinavian Journal of Economics* 97, 295-307.

[4] Akerlof, G. 1982. "Labor contracts as partial gift exchange." *Quarterly Journal of Economics* 97, 543-569.

[5] Bellemare, C. and Shearer, B. 2009. "Gift giving and worker productivity: Evidence from a firm-level experiment." *Games and Economic Behavior* 67, 233-244.

[6] Bellemare, C. and Shearer, B. 2011. "On the relevance and composition of gifts within the firm: evidence from field experiments." *International Economic Review* 52(3), 855-882.

[7] Bewley, T.F. 1999. *Why Wages Don't Fall During a Recession.* Cambridge, MA: Harvard University Press.

[8] Blanco, M., Engelmann, D., Koch, A. and Normann, H-T. 2010. "Belief elicitation in experiments: Is there a hedging problem?" *Experimental Economics* 13(4), 412-438.

[9] Brandts, J. and Charness, G. 2004. “Do labour market conditions affect gift exchange? Some experimental evidence.” *Economic Journal* 114, 684-708.

[10] Charness, G. 2004. “Attribution and reciprocity in an experimental labor market.” *Journal of Labor Economics* 22(3), 665-688.

[11] Charness, G., Cobo-Reyes, R., Jimenez, N., Lacomba, J.A. and Lagos, F. 2012. “The hidden advantage of delegation: Pareto improvements in a gift exchange game.” *American Economic Review* 102(5), 2358-2379.

[12] Charness, G. and Kuhn, P. 2007. “Does Pay Inequality Affect Worker Effort? Experimental Evidence.” *Journal of Labor Economics* 25(4), 693-723.

[13] Charness, G. and Kuhn, P. 2011. “Lab labor: What can labor economists learn from the lab?” in Ashenfelter, O. and Card, D. (eds) *Handbook of Labor Economics* Vol 4a., 229-330.

[14] Charness, G., Fréchette, G. and Kagel, J.H. 2004. “How robust is laboratory gift exchange?” *Experimental Economics* 7, 189-205.

[15] Charness, G. and Rabin, M. 2002. “Understanding social preferences with simple tests.” *Quarterly Journal of Economics* 117, 817-869.

[16] Charness, G., Villeval, M.C. 2009. Cooperation, competition, and risk attitudes: An intergenerational field and laboratory experiment, American Economic Review, 99 (3), 956978.

[17] Falk, A. 2007. Gift-exchange in the field. *Econometrica* 75(5), 1501-1511.

[18] Fehr, E., Gächter, S. and Kirchsteiger, G. 1997. “Reciprocity as a contract enforcement device: Experimental evidence.” *Econometrica* 65(4), 833-860.

[19] Fehr, E., Kirchsteiger, G. and Riedl, A. 1993. “Does fairness prevent market clearing? An experimental investigation.” *Quarterly Journal of Economics* 108(2), 437-59.

[20] Fehr, E., Kirchsteiger, G. and Riedl, A. 1998. “Gift exchange and reciprocity in competitive experimental markets.” *European Economic Review* 42(1), 1-34.

[21] Fehr, E., Kirchler, E., Weichbold, A. and Gächter, S. 1998. "When social norms overpower competition: Gift exchange in experimental labor markets." *Journal of Labor Economics* 16, 324-351.

[22] Fischbacher, U. 2007. "Z-Tree - Zurich toolbox for readymade economic experiments." *Experimental Economics* 10, 171-178.

[23] Friedman, D. and Sunder, S. 1994. *Experimental Methods: A Primer for Economists* Cambridge University Press.

[24] Gächter, S. Nosenzo, D. and Sefton, M. 2010. "Peer effects in pro-social behavior: Social norms or social preferences?" *Journal of the European Economic Association* 11(3), 548-573.

[25] Gneezy, U. and List, J.A. 2006. "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments." *Econometrica* 74(5), 1365-1384.

[26] Greiner, B. 2004. "An online recruitment system for economic experiments." *MPRA Paper* 13513, University Library of Munich, Germany.

[27] Hannan, R. L., Kagel, J.H. and Moser, D.V. 2002. "Partial gift exchange in an experimental labor market: Impact of subject population." *Journal of Labor Economics* 20(4), 923-951.

[28] Holt, C.A., and Laury, S.K. 2002. "Risk aversion and incentive effects." *American Economic Review* 92, 1644-1655.

[29] Krueger, A. and Mas, A. 2004. "Strikes, scabs and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires." *Journal of Political Economy* 112(2), 253-289.

[30] Kube, S., Maréchal, M.A. and Puppe, C. 2012. "The currency of reciprocity: Gift exchange in the workplace." *American Economic Review* 102(4), 1644-1662.

[31] Maximiano, S., Sloof, R. and Sonnemans, J. 2007. "Gift exchange in a multi-worker firm." *Economic Journal* 117, 1025-1050.

[32] Niederle, M. and Vesterlund, L. 2007. "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics* 122(3), 1067-1101.

[33] Papke, L.E. and Wooldridge, J.M. 1996. "Econometric methods for fractional response variables with an application to 401(K) plan participation rates." *Journal of Applied Econometrics* 11, 619-632.

[34] Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data* Second Edition. Cambridge, MA: MIT Press.

A Proof of Proposition 1

First note that when agents are self-interested, sellers will always set $g = 0$, since this minimizes costs. Now consider the case where $S = 3 > B = 2$. Let's show that the constraint below is binding:

$$U_S(matched) \geq U_S(unmatched) \iff (1-\tau)\pi_S + T \geq T \iff (1-\tau)\pi_S \geq 0. \tag{A1}$$

We do so by constructing the following Lagrangian for the buyer's problem:

$$L = (1-\tau)\pi_B + T + \lambda(1-\tau)\pi_S. \tag{A2}$$

The resulting FOCs are:

$$-(1-\tau)q + (1-\tau)\lambda q = 0, \tag{A3}$$

$$(1-\tau)\left[\frac{a}{2\sqrt{q}} - w\right] + \lambda(1-\tau)[qw - c_l q] = 0, \tag{A4}$$

$$\lambda(1-\tau)(qw - c_l q) = 0. \tag{A5}$$

If $\lambda = 0$ then we have: $q = 0$, $w = \frac{a}{2\sqrt{q}}$ and $\pi_B = 0$. If $\lambda \neq 0$ we have $q^* = (\frac{a}{2c_l})^2, w^* = c_l$ and $\pi_B > 0$. The constraint is therefore binding.

Let's now consider the case where $S = 2 < B = 3$. Let's show that the constraint below is binding:

$$U_B(matched) \geq U_B(unmatched) \iff (1-\tau)\pi_B + T \geq T \iff (1-\tau)\pi_B \geq 0. \tag{A6}$$

We do so by constructing the following Lagrangian for the seller's problem:

$$L = (1-\tau)\pi_S + T + \lambda(1-\tau)\pi_B. \tag{A7}$$

The resulting FOCs are:

$$(1-\tau)q - (1-\tau)\lambda q = 0, \tag{A8}$$

$$(1-\tau)[qw - c_l q] + \lambda(1-\tau)\left[\frac{a}{2\sqrt{q}} - w\right] = 0, \tag{A9}$$

$$\lambda(1-\tau)(a\sqrt{q} - wq) = 0. \tag{A10}$$

If $\lambda = 0$ then we have: $q = 0$, $w = c_l$ and $\pi_S = 0$. If $\lambda \neq 0$ we have $q^* = (\frac{a}{2c_l})^2, w^* = 2c_l$ and $\pi_S > 0$. The constraint is therefore binding.

We now turn to the three sub-cases. Statement 1 follows from the fact that in both market structures $\frac{\partial q^*}{\partial c_l} < 0$ and $\frac{\partial w^*}{\partial c_l} > 0$. Statement 2 follows from the fact that sellers' choices when $S > B$ do not depend on knowledge of the firm's payoff. Statement 3 follows from the fact that in both market structures the expressions for q^* and w^* do not depend on τ.

B Instruction Sets — Main Experiment

The instruction sets all followed the same exact structure. The only differences were (a) the sentence describing the number of buyers and sellers, (b) the screenshots accompanying the experimental instructions and (c) the payoff section. To economize on space we will present the common version of the text, and the different payoff sections for all the treatments in succession where they would originally appeared in the text. The screenshots refer to the 2 Buyer, 3 Seller treatments.

Instructions

Welcome to our experiment. Please do not communicate with the other participants in the room. If you have any questions, please raise your hand, and we will take your question in private.

Please read these instructions carefully. Through your decisions and the decisions of others you may stand to gain a significant amount of money.

This experiment is divided in two parts. Your payment for the experiment will be the sum of payments from both parts.

The payoffs in the experiment are in Experimental Currency Units (ECU). 500 ECU are worth £1. Once the experiment is over, your payoffs in ECU will be converted to pounds and paid in cash.

Part 1

[SellerPower: In this part of the experiment, there are 3 buyers and 2 sellers, who interact in a market for puzzles.]

[BuyerPower: In this part of the experiment, there are 2 buyers and 3 sellers, who interact in a market for puzzles.]

You will be assigned the role of [**buyer/seller**] at the beginning of the experiment and you will retain that role throughout the experiment. In other words, if you are a buyer in round 1, you will always be a buyer for all the remaining rounds. Likewise, if you are a seller in round 1, you will always be a seller.

A sellers payoff depends on how many puzzles he sells.

A buyers payoff depends on how many puzzles he gets minus the price he pays the seller for those puzzles. There are two puzzles in the economy, type-A and type-B puzzles. The buyers valuation of type-A and type-B puzzles are given in the payoff sheet that accompanies this instruction set.

For example, 1 type-B puzzle and 0 type-A puzzles, is worth 173ECU to the buyer. 1 type-A puzzle and 0 type-B puzzles is worth 100 ECU. 4 type-B puzzles and 1 type-A puzzle are worth 361 ECU; 4 type-A puzzles and 1 type-B puzzle are worth ECU 265 ECU. In other words, type-B puzzles are more valuable to buyers than type-A puzzles.

Type-A puzzles are simple addition operations, for example: 7+5. Type-B puzzles are addition operations with a subtraction: 6+3-8.

To obtain puzzles, each buyer must post a contract offer. This contract offer specifies the number of puzzles the buyer would like to get and a price per puzzle to be paid to the seller. The maximum number of puzzles a buyer can ask for is 15.

For example, a buyer could ask for 10 puzzles at 26 ECU per puzzle. However, the buyer CANNOT specify which type of puzzles he would like.

Once a seller agrees to that contract, he must choose how many puzzles of each type he wants to solve. In our example if a seller accepts the contract offer, that seller can choose to solve 10 type-A puzzles and no type-B puzzles; or 10 type-B and no type-A puzzles; or any other combination of type-A and type-B puzzles.

The payoff to the seller will depend on whether he completes the contract. If the seller completes the contract, his payoff is equal to the price per puzzle times the agreed number of puzzles. In our example that would be 10×26 ECU = 260 ECU.

If the seller cannot complete the agreed number of puzzles, he will be penalised. His payoff will be equal to the number of puzzles he completed times HALF the agreed price. In our example, if the seller could only correctly solve 7 puzzles, his payoff would be equal to $7 \times (26/2)$ ECU= 91ECU.

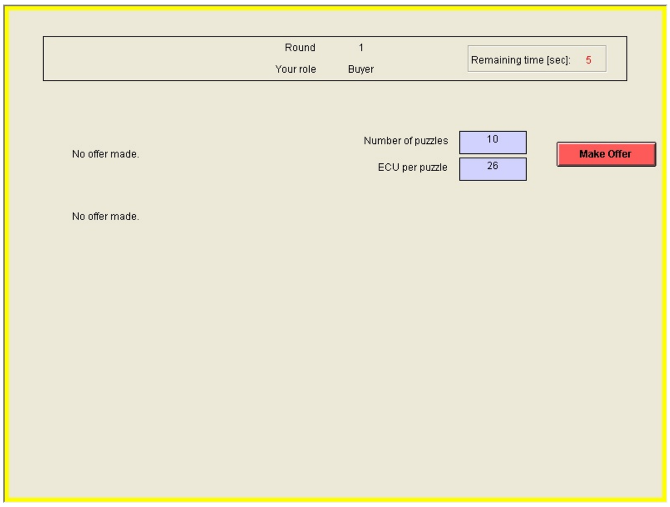
Puzzle trading stage

In the market for puzzles, buyers first must post how many puzzles they wish to buy, and how much they are willing to pay for each puzzle. They have 15 seconds to do this. If a buyer takes too long to post a contract, they will have to wait until the next opportunity.

For example, buyer A may offer a contract for 10 puzzles at 26 ECU per puzzle, and buyer B may offer a contract for 5 puzzles at 13 ECU per puzzle.

Once a buyer posts his contract, that contract is visible to all sellers for 15 seconds. If a seller wishes to accept a contract, he should click on the Accept button next to the contract offer.

If a seller does not want to accept a contract, he should let the 15-second timer run out.

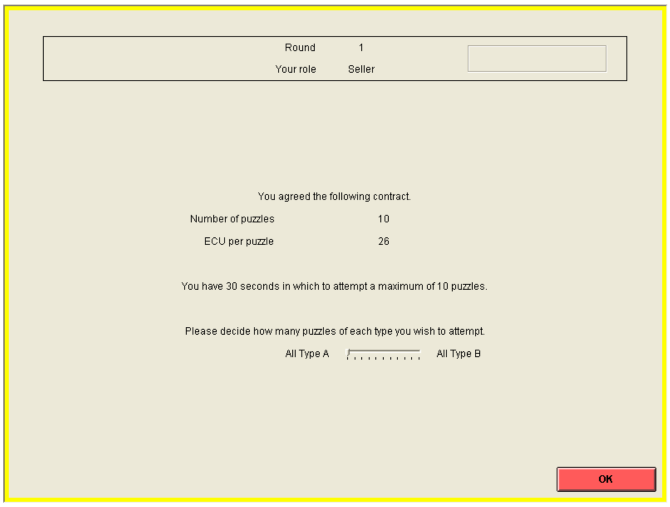


Once a seller accepts an offer, both he and the buyer who posted that offer leave the market. This means that the buyer cannot revise his offer and the seller cannot accept other offers.

If after 15 seconds there are still unaccepted offers, the relevant buyer can revise his contract and re-post it for another 15 seconds.

This will go on for a total of 2 minutes, or until all buyers have had their contracts taken up by sellers. A buyer who does not have his contract accepted after 2 minutes gets a payoff of zero (0) for the round. Also, a seller who does not have a contract will also get a payoff of zero (0) for the round.

Puzzle-solving stage



In the second stage, the sellers must

1. Decide how many puzzles of each type to solve
2. Correctly solve the puzzles.

For a given contract, sellers must decide how many type-A and type-B puzzles to solve, they must use the slider bar. If they keep the slider bar to the left, that means they will only solve type-A puzzles. By sliding the bar to the right, they increase the proportion of type-B puzzles they wish to solve. If the slider goes all they way to the right, they will solve type-B puzzles only.

After having decided how many puzzles of each type to solve, sellers must correctly solve the agreed number of puzzles.

Sellers have **30 seconds** to solve the agreed number of puzzles.

Payoff stage (CTRL treatments, Buyer and Seller)

If the seller correctly solves all puzzles, the payoffs are as agreed by the contract.

The payoff to the seller is therefore 10×26ECU $= 260$ ECU.

The payoff to the buyer will depend on how many puzzles of each type the seller correctly solved. If the seller correctly solved 5 type-A and 5 type-B, the payoff to the buyer is 447 - 260 = 187 ECU.

If the seller could only finish a part of the agreed puzzles, his payoff is equal to the number of correctly solved puzzles times HALF of the price. However, the payoff to the buyer is still the payoff from the solved puzzles minus the agreed price.

As per our example, suppose the seller could only correctly solve 5 type-A and 2 type-B puzzles.

- Payoff to the seller = $7 \times (26/2) = 91$ ECU
- Payoff to the buyer = 332 - 260 ECU = 72 ECU.

Payoff stage (ASYM treatments, Buyer)

If the seller correctly solves all puzzles, the payoffs are as agreed by the contract.

The payoff to the seller is therefore 10×26ECU = 260 ECU.

The payoff to the buyer will depend on how many puzzles of each type the seller correctly solved. If the seller correctly solved 5 type-A and 5 type-B, the payoff to the buyer is 447 - 260 = 187 ECU.

If the seller could only finish a part of the agreed puzzles, his payoff is equal to the number of correctly solved puzzles times HALF of the price. However, the payoff to the buyer is still the payoff from the solved puzzles minus the agreed price.

As per our example, suppose the seller could only correctly solve 5 type-A and 2 type-B puzzles.

- Payoff to the seller = $7\text{-} \times (26/2) = 91$ ECU
- Payoff to the buyer = 332 - 260 ECU = 72 ECU.

Payoff stage (ASYM treatments, Seller)

If the seller correctly solves all puzzles, the payoffs are as agreed by the contract.

The payoff to the seller is therefore 10×26ECU = 260 ECU.

The payoff to the buyer will depend on how many puzzles of each type the seller correctly solved. The more type B puzzles the seller solves, the higher the payoff to the buyer.

If the seller could only finish a part of the agreed puzzles, his payoff is equal to the number of correctly solved puzzles times HALF of the price. However, the payoff to the buyer is still the payoff from the solved puzzles minus the agreed price.

As per our example, suppose the seller could only correctly solve 5 type-A and 2 type-B puzzles.

- Payoff to the seller = $7 \times (26/2) = 91$ ECU
- Payoff to the buyer = ??? - 260 ECU.

Payoff stage (TAX treatments, Buyer and Seller)

If the seller correctly solves all puzzles, the payoffs are as agreed by the contract.

The payoff to the seller is therefore 10×26ECU = 260 ECU.

The payoff to the buyer will depend on how many puzzles of each type the seller correctly solved. If the seller correctly solved 5 type-A and 5 type-B, the payoff to the buyer is 447 - 260 = 187 ECU.

If the seller could only finish a part of the agreed puzzles, his payoff is equal to the number of correctly solved puzzles times HALF of the price. However, the payoff to the buyer is still the payoff from the solved puzzles minus the agreed price.

As per our example, suppose the seller could only correctly solve 5 type-A and 2 type-B puzzles.

- Payoff to the seller = $7 \times (26/2) = 91$ ECU
- Payoff to the buyer = 332 - 260 ECU = 72 ECU.

Tax calculations

Once earnings are calculated, a tax of 10% will be taken out of all players earnings. The amount collected will then be redistributed equally among the five players.

This means that in the first example the seller would pay 26 ECU in tax and the buyer would pay 18.7 ECU.

In the second example, the seller would pay 9.1 ECU in tax and the buyer would pay 7.2 ECU in tax.

The amount each player receives back would depend on the earnings of the other 3 players in the market. The higher the earnings of the other players, the more a player receives back. The higher your earnings, the more the other players receive back.

For example, if the total amount taxed from the 5 players equals 50 ECU, then each player receives 10 ECU back.

If the total amount taxed equals 75 ECU, then each player receives 15 ECU back.

Trial Stage

To help you understand how the puzzle-solving stage works, you will be asked to solve a number of puzzles. Firstly you will have to solve a number of type-A puzzles. You will have two attempts to do so, and each attempt lasts 30 seconds. Then you will have to solve type-B puzzles. Like before, you will have 2 attempts to solve these puzzles. Each attempt will last 30 seconds.

You will receive £0.20 for each puzzle you correctly solve in the trial stage.

Duration of Part 1 and Payoffs

There will be at least 20 periods in part 1 of the experiment. Once the 20th period is over, the computer will throw a virtual dice. If the dice rolls a 6 the experiment stops; otherwise there will be another period.

You payoff will be equal to the earnings in the trial stage, plus the sum of ECU you accumulate in part 1 of the experiment.

Before the experiment starts, we would like you to complete a small quiz, which comes with this instruction set in a separate sheet. This quiz is there to ensure that no aspect of the instructions is unclear and it will have no bearing on your payment. We will come round to check your answers and take any questions you may have.

Number of Type-A puzzles \ Number of Type-B puzzles	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	173	245	300	346	387	424	458	490	520	548	574	600	624	648	671
1	100	200	265	316	361	400	436	469	500	529	557	583	608	632	656	
2	141	224	283	332	374	412	447	480	510	539	566	592	616	640		
3	173	245	300	346	387	424	458	490	520	548	574	600	624			
4	200	265	316	361	400	436	469	500	529	557	583	608				
5	224	283	332	374	412	447	480	510	539	566	592					
6	245	300	346	387	424	458	490	520	548	574						
7	265	316	361	400	436	469	500	529	557							
8	283	332	374	412	447	480	510	539								
9	300	346	387	424	458	490	520									
10	316	361	400	436	469	500										
11	332	374	412	447	480											
12	346	387	424	458												
13	361	400	436													
14	374	412														
15	387															

Figure 4: Payoff Table for Buyers

Quiz

This quiz is just to ensure that no aspect of the instructions is unclear. It has no impact on your earnings.

Question 1: What is the maximum number of puzzles a buyer can ask for?

Question 2: Can a buyer request a particular number of A-type and B-type puzzles? .

Question 3: Tick the correct answer. Suppose you are a seller and you see contracts on offer.

a) What should you do if you do not want to accept the contract?

1. Allow the clock to run down
2. Click Accept on one of the offers

b) What should you do if you want to accept the contract?

1. Allow the clock to run down
2. Click Accept on the relevant offer

Question 4: Suppose a seller accepted a contract. However, a more appealing contract just came up. Can the seller renege on the old contract accept the new one?

Question 5: Can a seller decide what type of puzzles to solve?

Question 6: How long (in seconds) does a seller have to solve the agreed number of puzzles?

Question 7: Suppose that the Puzzle trading stage ends and 1 buyer and 2 sellers did not agree a contract.

a) What is the payoff to the sellers?

b) What is the payoff to the buyer?

Question 8: Suppose a seller accepts a contract requesting 7 puzzles at 8 ECU per puzzle.

a) If the seller successfully solves 7 A-type puzzles, the final payoffs for the seller will be [ ] ECU and the final payoff for the buyer will be [ ] ECU.

b) If the seller successfully solves 4 A-type puzzles and 3 B-type puzzles, the final payoffs for the seller will be [ ] ECU and the final payoff for the buyer will be [ ] ECU.

c) If the seller successfully solves 7 B-type puzzles, the final payoffs for the seller will be [ ] ECU and the final payoff for the buyer will be [ ] ECU.

d) If the seller can only solve 4 A-type puzzles and 1 B-type puzzle, the final payoffs for the seller will be [ ] ECU and the final payoff for the buyer will be [ ] ECU.

Question 9 (Tax treatment only):

In part a) of Question 8, how much would the seller pay in tax? [ ] ECU. How much would the buyer pay in tax? [ ] ECU. If the total amount paid in tax by the 5 people in their market was 200 ECU, how much would the seller get back? [ ] ECU. How much would the buyer? [ ] ECU.

In part b) of Question 8, how much would the seller pay in tax? [ ] ECU. How much would the buyer pay in tax? [ ] ECU. If the total amount paid in tax by the 5 people in their market was 200 ECU, how much would the seller get back? [ ] ECU. How much would the buyer? [ ] ECU.

In part c) of Question 8, how much would the seller pay in tax? [ ] ECU. How much would the buyer pay in tax? [ ] ECU. If the total amount paid in tax by the 5 people in their market was 200 ECU, how much would the seller get back? [ ] ECU. How much would the buyer? [ ] ECU.

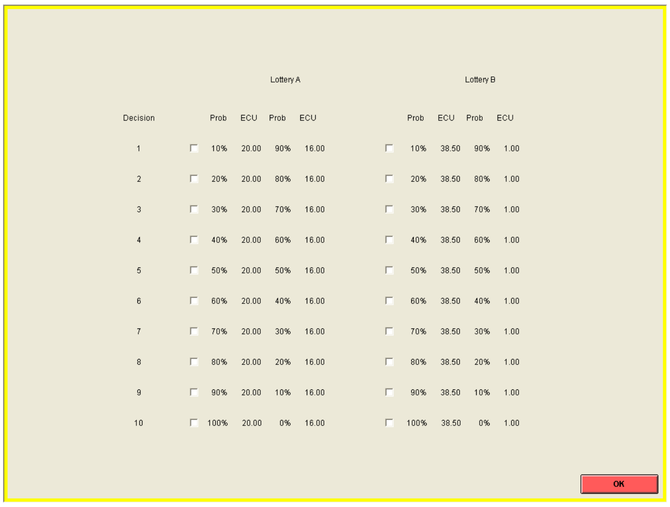
C Instruction Sets - Follow Up Experiment

Welcome to our experiment. Please read this document carefully: your payment in this experiment will depend on your decisions and the decisions of others. As such, understanding the rules of the experiment is very important. There will be two parts to the experiment. You will get a separate payoff from each part of the experiment. Once the experiment is over, we will add up the payoffs from the two parts, and pay you in cash. Throughout the experiment we will use an Experimental Currency Unit (ECU); 4 ECU are worth £1.

Part 1

In this part of the experiment, we will ask you to make a number of decisions. Your decisions in this part of the experiment will affect your own payoff, but not the payoff of other participants.

The decisions we will ask you to make will be choices between pairs of lotteries. The table below illustrates the task.



	Lottery A				Lottery B			
Decision	Prob	ECU	Prob	ECU	Prob	ECU	Prob	ECU
1	10%	20.00	90%	16.00	10%	38.50	90%	1.00
2	20%	20.00	80%	16.00	20%	38.50	80%	1.00
3	30%	20.00	70%	16.00	30%	38.50	70%	1.00
4	40%	20.00	60%	16.00	40%	38.50	60%	1.00
5	50%	20.00	50%	16.00	50%	38.50	50%	1.00
6	60%	20.00	40%	16.00	60%	38.50	40%	1.00
7	70%	20.00	30%	16.00	70%	38.50	30%	1.00
8	80%	20.00	20%	16.00	80%	38.50	20%	1.00
9	90%	20.00	10%	16.00	90%	38.50	10%	1.00
10	100%	20.00	0%	16.00	100%	38.50	0%	1.00

Each lottery can have two outcomes. The columns named Prob indicate the probability of each outcome of the draw, while the ECU columns indicate the value of the outcome.

For example, in Decision #1, picking Lottery A means you have a 10% chance of getting 20.00 ECU and a 90% chance of getting 16.00 ECU; picking Lottery B means you have a 10% chance of getting 38.50 ECU and a 90% chance of getting 1.00 ECU.

Option A	Option B
(6.34, 4.00)	(5.50, 5.50)

You must pick between Lottery A and Lottery B for all choices presented to you. If you look carefully, as you move down the table the probability associated with the high payoff rises for both Lottery A and Lottery B.

In fact, once you get to Decision 10, you know that the high-value outcome for each lottery will happen with certainty. Your choice is therefore between 20 ECU for sure if you pick Lottery A or 38.5 ECU for sure if you pick Lottery B.

In short, you will have to make 10 choices, one for each Decision row. You may choose A for some decision rows and B for other rows this includes picking always Lottery A, or always Lottery B should you wish to do so. You will do these choices on the computer screen in a few moments.

When you finish, the computer will pick one of the ten Decisions you make at random and play it out. The outcome of the lottery you picked will determine your payoff for this part of the experiment. If you have any questions, please raise your hand.

Part 2

In this part of the experiment, we will ask you to make a number of decisions. Your decisions in this part of the experiment will affect both your payoff as well as the payoff of one other person in the room.

The decisions we will ask you to do will be choices between how to distribute payoffs between yourself and another person.

Take the following hypothetical example outlined in the table below. Each option is a pair of numbers in parenthesis. The first number is your payoff in ECUs and the second number is the other persons payoff in ECUs.

Hence, if you pick option A you will get 6.34 ECU and the other person will get 4 ECU. If you pick option B, both you and the other person will get 5.5 ECU.

You will make a number of these decisions on the screen. Once you complete all decisions, the computer will select one at random, and play out your choice.

Since everyone in the room is facing the same choices as you, this means that you get a payoff from two decisions:

- The decision you made in which the other person had no active part
- The decision someone else made in which you had no active part

Importantly, these two people will NOT be the same. That is, the person whose payoff is affected by your choices will not be the person whose choices affect your payoff.

Please have another read through these instructions. If you have any questions, please raise your hand.

D Auxiliary Regressions

In this section, we present the estimation results from the alternative specifications for the econometric analysis of seller behavior. We use as an alternative to the Tobit model the fractional regression model proposed by Papke and Wooldridge (1996), which takes the following form for the first set of regressions.

$$g_{it} = \frac{\exp\left[(1 + Dq_{it} + Dw_{it})\beta_1 + X_{it}\beta_2\right]}{1 + \exp\left[(1 + Dq_{it} + Dw_{it})\beta_1 + X_{it}\beta_2\right]} + \varepsilon_{it} \tag{6}$$

We reiterate that since we cannot work with a proper panel, and we have repeated observations for all i, we cluster standard errors at the market level. We begin by looking at the results from the restricted model in column (A1). We note some differences in the significance levels of the coefficients on the intercept dummies between the fractional Logit model and the Tobit model: Ctrl × BuyerPower is still positive but no longer significant, while FixPuz × BuyerPower is negative but also no longer significant. In contrast Tax × SellerPower is positive and now significant. In terms of the quantity interactions,

DV: g	(A3)		(A4)	
w × Ctrl × BuyerPower	0.034**	(0.017)	0.029	(0.019)
w × Asym × BuyerPower	-0.009	(0.011)	-0.024	(0.026)
w × Tax × BuyerPower	0.002	(0.008)	0.008	(0.020)
w × FixPuz × BuyerPower	0.007	(0.008)	0.041	(0.027)
w × SliRgt × BuyerPower	0.027*	(0.015)	-0.028	(0.040)
w × Ctrl × SellerPower	0.019*	(0.010)	0.035**	(0.015)
w × Asym × SellerPower	0.050**	(0.020)	0.057***	(0.016)
w × Tax × SellerPower	0.042***	(0.008)	0.016	(0.017)
q × Ctrl × BuyerPower	-0.167***	(0.052)	-0.187***	(0.055)
q × Asym × BuyerPower	-0.047	(0.077)	-0.020	(0.189)
q × Tax × BuyerPower	-0.216***	(0.024)	-0.218***	(0.046)
q × SliRgt × BuyerPower	0.029	(0.068)	-0.210**	(0.106)
q × Ctrl × SellerPower	-0.172*	(0.089)	-0.223***	(0.061)
q × Asym × SellerPower	-0.290***	(0.087)	-0.381***	(0.097)
q × Tax × SellerPower	-0.238**	(0.101)	-0.282***	(0.105)
Ctrl × BuyerPower	0.594	(0.559)	0.442	(1.131)
Asym × BuyerPower	-1.023	(0.763)	-2.100	(2.300)
Tax × BuyerPower	1.759***	(0.292)	1.222	(1.152)
FixPuz × BuyerPower	-1.324***	(0.296)	-2.693**	(1.101)
SliRgt × BuyerPower	-0.677	(0.770)	0.501	(1.438)
Ctrl × SellerPower	0.701	(1.295)	0.304	(1.225)
Asym × SellerPower	0.398	(0.859)	0.901	(1.310)
Tax × SellerPower	0.720	(1.001)	0.841	(1.521)
Ability_i			0.008	(0.068)
DIA_i			1.627	(1.260)
AIA_i			1.370**	(0.655)
Risk aversion_i			0.034	(0.092)
Male_i			-0.037	(0.143)
Period	-0.005	(0.007)	-0.002	(0.008)
AIC	0.944		0.849	
N	2,380		1,512	

Market level clustered standard errors in parenthesis.

***, **, *: statistical significance at 1%, 5% and 10% level.

Table 7: Logit fractional regression model estimates of determinants of high quality output – equation (4)

we observe no sign changes, but the coefficient on q × Ctrl × SellerPower is no longer significant.

Looking at the unrestricted model, we see a similar pattern, in which the signs of all coefficients are the same in the fractional Logit model as in the Tobit model, with slight differences in the significance level of some coefficients, much like the restricted model. The only differences between the two models are in the significance level of w × FixPuz × BuyerPower, Tax × BuyerPower (significant in the Tobit model but not in the fractional Logit model), and FixPuz × BuyerPower (significant in the fractional Logit model but not in the Tobit model). In short, the conclusions we derive from the Tobit model analysis are unchanged.

As per the main text, we also consider an econometric model in which the relationship between high quality output and prices depends on the contracted level of output. The specification of the model in this case is given by:

$$g_{it} = \frac{\exp[(1 + w_{it} + w_{it}FGE)\, D\beta_1 + X_{it}\beta_2]}{1 + \exp[(1 + w_{it} + w_{it}FGE)\, D\beta_1 + X_{it}\beta_2]} + \varepsilon_{it} \tag{7}$$

Comparing the Tobit estimation results of the restricted model to those from the fractional Logit model, the only difference is that the intercept dummies are all significant under the fractional regression model. All other coefficients in the model retain their significance level in both specifications. Looking at the unrestricted model, we still observe some differences in the significance level on the intercept dummies. We also observe two minor changes in significance: in the fractional Logit model, the coefficient on FGE × w × Tax × SellerPower is marginally significant ($p = 0.076$), and the coefficient on w × Tax × BuyerPower ceases to be significant. In short, the results of our analysis are not affected by our choice of estimator.

E Measuring inequality aversion

We measured participant attitudes to fairness using a series of mini-dictator games proposed by Charness and Rabin (2002). In each game, a subject had to pick between two options, each of which assigned a monetary payoff to himself and another player. The other player

DV: g	(A5)		(A6)	
FGE × w × Ctrl × BuyerPower	0.088***	(0.015)	0.103***	(0.025)
FGE × w × Asym × BuyerPower	-0.004	(0.024)	-0.093*	(0.051)
FGE × w × Tax × BuyerPower	0.014	(0.010)	0.003	(0.018)
FGE × w × FixPuz × BuyerPower	0.031	(0.029)	0.052***	(0.016)
FGE × w × SliRgt × BuyerPower	0.005	(0.006)	-0.017	(0.023)
FGE × w × Ctrl × SellerPower	0.060***	(0.018)	0.062***	(0.017)
FGE × w × Asym × SellerPower	0.048**	(0.022)	0.061**	(0.029)
FGE × w × Tax × SellerPower	0.033*	(0.019)	0.050*	(0.028)
w × Ctrl × BuyerPower	-0.001	(0.015)	-0.005	(0.023)
w × Asym × BuyerPower	-0.003	(0.011)	0.019	(0.021)
w × Tax × BuyerPower	0.005	(0.013)	0.018	(0.012)
w × FixPuz × BuyerPower	-0.010	(0.016)	0.022*	(0.011)
w × SliRgt × BuyerPower	0.027	(0.018)	-0.001	(0.026)
w × Ctrl × SellerPower	-0.013	(0.017)	0.009	(0.016)
w × Asym × SellerPower	0.025	(0.033)	0.020	(0.030)
w × Tax × SellerPower	0.027	(0.017)	-0.001	(0.029)
Ctrl × BuyerPower	-1.303***	(0.181)	-1.881**	(0.743)
Asym × BuyerPower	-1.586***	(0.326)	-2.369***	(0.892)
Tax × BuyerPower	-0.547**	(0.220)	-1.131*	(0.671)
FixPuz × BuyerPower	-1.372***	(0.223)	-3.042***	(0.960)
SliRgt × BuyerPower	-1.005***	(0.269)	-2.107***	(0.710)
Ctrl × SellerPower	-1.290***	(0.293)	-2.367***	(0.820)
Asym × SellerPower	-2.420***	(0.335)	-2.782***	(0.582)
Tax × SellerPower	-1.587***	(0.284)	-2.072**	(0.807)
DIA_i			1.628	(1.281)
AIA_i			1.218**	(0.611)
Risk aversion_i			0.062	(0.087)
Male_i			-0.065	(0.126)
Period	-0.007	(0.007)	-0.005	(0.009)
AIC	0.962		0.868	
N	2,380		1,512	

Market level clustered standard errors in parenthesis.

***, **: statistical significance at 1% and 5% level.

Table 8: Logit fractional regression model estimates of determinants of high quality output – equation (5)

did not have any active role in the game. In order to collect data from all participants as dictators, we set up a matching scheme in which every participant received two payoffs: one from his choices as a dictator, and the other as a recipient (i.e., when another participant was the dictator). We ensured that the counterparts for each participant in either case were different, to avoid issues of reciprocity – this was clearly stated in the game instructions.

Every participant had to make 20 such choices. Only one of them would be selected (at random) to be played out. Table 9 outlines the choices faced by participants. The computer presented the choices to participants in random order, to prevent ordering effects. Our model of inequality aversion is that of Charness and Rabin (2002), in which the utility function of a participant playing a two-player game takes the form:

$$U_i(x_i, x_j) = x_i - r \cdot \rho(x_i - x_j) - s \cdot \sigma(x_j - x_i) \tag{8}$$

where x_i is the dictator's payoff; x_j is the recipients payoff; $r = 1$ if $x_i > x_j$ and 0 otherwise; $s = 1$ if $x_i < x_j$ and 0 otherwise. Finally, ρ and σ respectively measure the degree to which individuals are averse to disadvantageous inequality and advantageous inequality in payoffs.

We use participant decisions in each of the 20 choices to find an estimate of ρ and σ for each individual. Choosing one option defines an inequality in (ρ, σ) space. For example, selecting A in choice 20 means $4 - (12 - 4)\rho > 3.52 - (3.52 - 1.76)\sigma$. Each player made decisions that defined 20 inequalities that together defined a feasible set of values on the $\rho - \sigma$ plane. Within the feasible set, we then conservatively chose the closest point to the origin (0,0) as the (ρ, σ) pair representing that player's preferences.

However, we found that players often made inconsistent decisions, leading to inequalities that could not be simultaneously satisfied. This created an empty feasible set, so that no (ρ, σ) could be found. Since this was most likely due to error on the part of the participant (e.g. failure to correctly judge the outcomes of a decision) we developed a work-around method for removing inconsistent choices.

This method worked by ranking each of the 20 decisions made by a participant for consistency. Thus when the 20 inequalities (defined by participant decisions) could not be simultaneously satisfied (i.e., the feasible set was empty), we sequentially removed the "least consistent" inequalities in turn, until a non-empty feasible set was generated. We

Choice	Option A	Option B
1	(6.34, 4.00)	(5.50, 5.50)
2	(7.50, 0.00)	(7.29, 4.00)
3	(1.08, 1.25)	(1.00, 10.00)
4	(7.25, 0.00)	(5.33, 3.90)
5	(8.00, 0.00)	(5.33, 3.90)
6	(7.50, 0.00)	(5.36, 3.90)
7	(7.50, 10.00)	(4.00, 5.75)
8	(4.50, 0.00)	(3.56, 4.44)
9	(4.50, 0.00)	(3.69, 4.31)
10	(7.00, 2.00)	(5.12, 6.22)
11	(8.00, 0.00)	(2.16, 5.84)
12	(2.24, 5.76)	(0.00, 8.00)
13	(3.94, 3.94)	(3.75, 10.00)
14	(3.96, 3.98)	(3.75, 10.00)
15	(7.28, 1.82)	(5.00, 5.00)
16	(4.50, 9.00)	(3.30, 4.00)
17	(5.50, 5.50)	(4.24, 3.98)
18	(5.50, 5.50)	(4.63, 3.96)
19	(7.50, 7.50)	(7.04, 1.76)
20	(4.00, 12.00)	(3.52, 1.76)

Table 9: Mini-dictator games. The first number in ordered pair is the dictator payoff, the second number is the recipient payoff.

determined the "least consistent" inequalities by ranking each inequality in terms of how many internally consistent subsets of the total set of inequalities it belonged to, i.e., how many consistent pairs, how many consistent triples, quadruples, 5-tuples, etc. When no unique least consistent inequality could be found (multiple least consistent inequalities were sometimes found due to symmetries) then we chose one of the least consistent inequalities at random for removal.

This process followed the following algorithm:

```
START
0: Test the consistency of the total remaining set of inequalities.
1: IF     (the inequality set is inconsistent)
   THEN   Go to 2.
   ELSE   Go to END.
2: Set N=2.
3: Calculate how many consistent N-tuples each inequality belongs to.
4: IF     (there is a unique least consistent inequality)
   THEN   Remove this inequality.
          Go to 0.
   ELSE
   {
          IF    (N<20)
          THEN  Set N=N+1.
                Go to 3.
          ELSE  There is no unique least consistent inequality.
                Remove one of the least consistent inequalities at random.
                Go to 0.
   }
END
```

After running the algorithm, the remaining set of inequalities is consistent, i.e., they can all be simultaneously satisfied. Then we found (ρ, σ) as before. We disregarded for the purposes of analysis any individual who had more than two inconsistent choices. In our econometric analysis, we denote ρ as AIA_i and σ as DIA_i.

F Measuring attitudes to financial risk

We measured attitudes to financial risk using the method proposed by Holt and Laury (2002). This method consists on presenting a set of binary lotteries as per Table 10. This table consists of ten decisions between two lotteries (Option A and Option B). In each decision, participants are faced with one lottery (Option A) with two intermediate payoffs (i.e., 20.00 and 16.00) and one lottery (Option B) with one high payoff (i.e., 38.50) and one low payoff (i.e., 1.00). In other words, option A is "safe" and option B is "risky".

As the penultimate column on Table 10 shows, in choice 1, option A has a higher expected value than option B. As we move down the list, the probability attached to the high payoff in both options increases by a factor of 10%, which means option B becomes more attractive than option A in terms of expected value. From choice 5 onwards, option B has a higher expected value than option A. When faced with choice 10, all participants should choose option B, since the high payoff in both options is certain — in that sense, choice 10 works as a "rationality check". In other words, either participants should always pick option B in all 10 choices, or they should switch from option A to option B at some point. That switch point will determine their risk aversion.

Formally, this method assumes a utility function of the form $u(x) = \frac{x^{1-r}}{1-r}$, for $x > 0$. This functional form implies constant relative risk aversion r for outcome x. When $r < 0$ a subject is risk seeking, when $r = 0$ a subject is risk-neutral, and when $r > 0$ a subject is risk averse.

As above, a choice between option A and option B defines an inequality of the form $\frac{20^{1-r}}{1-r}p + \frac{16^{1-r}}{1-r}(1-p) \gtrless \frac{38.50^{1-r}}{1-r}p + \frac{1^{1-r}}{1-r}(1-p)$. Picking Option A in Choice t and picking Option B in Choice $t+1$ therefore defines an interval of r which satisfies the two respective inequalities. The last column on Table 10 gives the ranges of r implied by a participant choosing B at a given choice. We model risk aversion as an integer between 1 and 10, corresponding to the switching point in the sequence of gambles. We excluded any subject

	Option A				Option B					
Choice	Prob	Payoff	Prob	Payoff	Prob	Payoff	Prob	Payoff	E(A-B)	r
1	10%	20.00	90%	16.00	10%	38.50	90%	1.00	11.7	(-∞, -0.95)
2	20%	20.00	80%	16.00	20%	38.50	80%	1.00	8.30	(-0.95, -0.49)
3	30%	20.00	70%	16.00	30%	38.50	70%	1.00	5.00	(-0.49, -0.15)
4	40%	20.00	60%	16.00	40%	38.50	60%	1.00	1.60	(-0.15, 0.15)
5	50%	20.00	50%	16.00	50%	38.50	50%	1.00	-1.80	(0.15, 0.41)
6	60%	20.00	40%	16.00	60%	38.50	40%	1.00	-5.10	(0.41, 0.68)
7	70%	20.00	30%	16.00	70%	38.50	30%	1.00	-8.50	(0.68, 0.97)
8	80%	20.00	20%	16.00	80%	38.50	20%	1.00	-11.80	(0.97, 1.37)
9	90%	20.00	10%	16.00	90%	38.50	10%	1.00	-15.20	(1.37, +∞)
10	100%	20.00	0%	16.00	100%	38.50	90%	1.00	-18.50	

Table 10: Risk aversion elicitation choices.

who never switched, since in the last lottery pair, Option B dominates Option A.