Journal of Economic Behavior and Organization 191 (2021) 714–737

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

When no bad deed goes punished: Relational contracting in Ghana and the UK☆

Elwyn Davies[a], Marcel Fafchamps[b,*]

[a] The World Bank, 1818 H Street NW, Washington DC 20433, USA
[b] Freeman Spogli Institute for International Studies, Stanford University, 616 Jane Stanford Way, Stanford CA 94305, USA

ARTICLE INFO

Article history:
Received 10 September 2020
Revised 28 June 2021
Accepted 17 September 2021
Available online 10 October 2021

JEL classification:
C71
D2
D86
E24
O1

Keywords:
Relational contracting
Labor incentives
Gift-exchange game
Punishment strategies
Ghana

ABSTRACT

Experiment evidence to date indicates that subjects follow a trigger strategy in finitely repeated games: they punish bad contractual performance by reducing future offers and the threat of punishment disciplines opportunistic breach. This behavior contradicts standard game theory predictions. We conduct a repeated prisoner's dilemma (PD) game experiment with university students in Ghana and the UK. The experiment is framed as an employment contract. Each period the employer makes a irrevocable wage offer to the worker who then chooses an effort level. UK subjects use a trigger strategy to discipline workers, in line with previous experiments: wage offers reward high effort and punish low effort in the past; this induces workers to choose high effort; and gains from trade are shared between workers and employers. We find no such evidence with Ghana subjects: employers seldom reduce wage offers after low effort and, if they do, workers respond by lowering effort; employer often reduce wages after high effort; and employers earn a zero payoff on average. Introducing competition or reputation does not significantly improve workers' effort. We conclude that the use of trigger strategies in repeated labor transactions is not a universally shared heuristic.

☆ We would like to thank Johannes Abeler, Alan Beggs, Douglas Bernheim, Stefano Caria, Johannes Haushofer, Rocco Macchiavello, Karine Marazyan, Fola Malomo, Muriel Niederle, Simon Quinn, Christopher Roth, Yogita Shamdasani, Hannah Uckat, Marc Witte as well as participants of the Oxford Firms and Development Research Group, the Stanford Behavioral and Experimental Lunch, the CSAE-iiG Conference, the Stanford Development Workshop, the CSAE Research Workshop, the Paris DIAL conference, the World Bank Poverty and Applied Micro Seminar, NEUDC, PacDev and SEEDEC for helpful comments. Many thanks to Václav Těhle and Martin Černý for excellent research assistance, and thanks to Andrew Kerr and Denise Gray for their assistance during trial sessions of this experiment. Furthermore, we would like to thank Moses Awoonor-Williams for his role in helping out with logistics, as well as our local team in the field, in particular Eric Agyekum and Bismark Owusu Tetteh Nortey. This research belongs to a wider study of Ghanaian entrepreneurship funded by the UK Department for International Development (DFID) [grant GB-1-113580] as part of the iiG, a research programme to study how to improve institutions for pro-poor growth in Africa and South-Asia. Elwyn Davies gratefully acknowledges additional funding from the Economic and Social Research Council (ESRC) [grant ES/J500112/1]. IRB approval for this research was obtained from the Central University Research Ethics Committee of the University of Oxford. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of DFID, the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

* Corresponding author.
E-mail addresses: edavies@worldbank.org (E. Davies), fafchamp@stanford.edu (M. Fafchamps).

https://doi.org/10.1016/j.jebo.2021.09.024
0167-2681/

1. Introduction

A growing body of evidence has shown that norms, preferences, and behavior differ across cultures (Cardenas and Carpenter, 2008; Herrmann et al., 2008; Henrich et al., 2010). The Global Preference Survey conducted by Falk et al. (2018) is a good example of recent work in this area. The authors elicit individual preferences in representative samples from 76 countries and find large differences both across and within countries. These findings are echoed in the work of Schulz et al. (2018a; 2018b) who document differences in social norms across cultures using survey questions or simple decision games, a point that is also the focus of Enke (2019). Fisman et al. (2014) similarly show large differences in attitudes towards redistributive justice between sub-populations of the US. Much less attention has been devoted to differences between subject pools in terms of *strategic* interaction in repeated games. This is unfortunate given that our understanding of institutions and contracts relies critically on assumptions regarding how human subjects behave in repeated interactions.

To fill this knowledge gap, we revisit a moderately complex game for which extensive behavioral evidence already exists for students in the developed Western world, namely the repeated contracting game between employer and employee of Brown et al. (2004; 2012). This game is interesting for several reasons. First, many economic and social transactions take the form of a repeated one-side or two-sided prisoner's dilemma (PD) game: e.g., permanent employment; relational contracts with suppliers; and social relationships of friendship and favor exchange. By framing the game as one of these familiar transactions, we can expect subjects to find suitable heuristics, minimizing the mistakes subjects make in strategically unfamiliar games.

Secondly, the theoretical properties of repeated PD games are well known. If the game is infinitely repeated or has an unknown duration, a trigger-strategy equilibrium exists in which players cooperate until one of them deviates, after which the other player stops cooperating. The trigger strategy is not a subgame perfect equilibrium when the number of periods is fixed (e.g., Fudenberg and Maskin, 1986; Fudenberg and Tirole, 1991). Yet, laboratory experiments have shown that subjects play one even in finitely repeated games (Andreoni and Miller, 1993) – perhaps because they intuitively reason as if the game had unknown duration. In particular, convincing evidence exists that the way students in Western universities play the employer-employee game is rationalizable as a trigger strategy: if a worker defects by choosing low effort, the employer reduces the next wage offer and, anticipating this, the worker exerts high effort in most rounds (e.g., Brown et al., 2004; Brown et al., 2012).

These findings have led to the belief that trigger strategies are a universal regularity of human behavior even in finitely repeated games. As a result, they lie at the core of economic thinking on how repeated interactions works: risk sharing (e.g., Coate and Ravallion, 1993, Kocherlakota, 1996; Bloch et al., 2008; Jackson and Rodriguez-Barraquer, 2012); relational contracting (e.g., MacLeod, 2007; Chakravarty and MacLeod, 2009; Malcomson, 2016); as well as much thinking on economic institutions (e.g., Greif, 1993) and trade (e.g., Rauch and Casella, 2003). The applicability of these concepts to developing countries has been documented, among others, by McMillan and Woodruff (1999), Fafchamps and Minten (2002), Fisman (2003), Macchiavello and Morjaria (2015), and Grossman and Startz (2020).

If, however, trigger strategies are not used by some modern subjects – including in games framed in a way familiar to them – this would cast doubt on the *universal* applicability of the above literature to all modern populations. To show that a behavioral principle is *not* universal, all that is needed is convincing evidence of *one* case in which it does not apply. To be convincing, the evidence has to use comparable subjects, be reproducible, and be internally consistent. Putting the universality of trigger strategies to the test is the object of this paper.

There are good reasons to suspect that reliance on a trigger strategy need not be universal. Theory teaches us that the trigger strategy equilibrium is only *one* of a multiplicity of equilibria in infinitely repeated PD games; non-cooperation is also an equilibrium (e.g., Fudenberg and Tirole, 1991). Furthermore, non-cooperation is the only subgame perfect equilibrium of finitely repeated PD games such as those played in the lab. The fact that trigger strategies have been observed in many experiments is, therefore, a puzzle: they are not subgame perfect in a finitely repeated game; and even if subjects reasons 'as if the duration of the game is unknown', this only makes trigger strategies feasible, not why they are selected. To make trigger strategies the equilibrium chosen by most subjects, they have to be supported by behavioral expectations, such as those activated by a focal point or norm that subjects naturally adhere to in a given population.

We should therefore expect variation in repeated game equilibria if behavioral expectations vary across populations. For instance, if workers do not increase effort after a reduction in wage, employers cannot hope to increase effort by reducing wages; and if employers do not offer a high wage after high effort, workers cannot hope to raise their wage by increasing effort. Since there is no a priori reason for such expectations *not* to arise naturally in some populations, it is suspicious that deviations from trigger strategy equilibria have not been observed before. As our paper demonstrates, we were perhaps looking in the wrong places.

To ensure that our results are comparable to the literature, we base our experiment on the well-known designs of Fehr et al. (1993) and Brown et al. (2004). Subjects are recruited among students in top universities, a commonly used subject pool selected to facilitate comparisons with the literature and ensure economic relevance.[1] This also guarantees that mistakes or poor understanding are unlikely to account for behavioral differences between subject pools. The experiment is

[1] Many studies of variation in human behavior have taken an anthropological approach by focusing on 'non-modern' subjects (e.g., Henrich et al., 2005; Gneezy et al., 2009). While often a source of deep insights, the findings from these studies may not be relevant for the modern world.

run in two countries: the United Kingdom; and Ghana, a middle income country in West Africa. These two countries are selected because they share the same language of higher education, which allows us to use the same experimental protocol and equipment. They have similar legal institutions but a different culture and history.

In the experiment, subjects are randomly assigned the role of worker or employer. This frame is chosen because it is familiar to our study populations.[2] The subjects interact in a principal-agent setting. An employer makes an offer to a worker, specifying a wage and requesting an effort level. The worker chooses to accept or reject this offer and, then, a level of effort. The employer cannot condition the wage on effort but the worker and employer interact repeatedly and the employer can adjust wage offers in subsequent periods. As in similar experiments from the literature (e.g., Fehr et al., 1993; Brown et al., 2004; Brown et al., 2012), there is no real task and effort is notional. We investigate whether employers rely on a trigger strategy to discipline shirking workers and whether worker reputation and competition across workers create additional incentives to exert effort. This is achieved by combining five treatments that vary market size, contractual completeness, and the information available to employers about past worker performance.

We find that students in the United Kingdom behave like other Western subjects with whom the gift exchange experiment has been run. But university students in Ghana behave differently: workers consistently provide low effort, leading to low aggregate efficiency; and employers pay high average wages and get a low payoff. These results are robust to the different treatments, and they were replicated in a second set of Ghana sessions.

The Ghana findings are surprising given the widely accepted view that cooperation naturally emerges in repeated games, irrespective of context. The difference arises because Ghana subjects do not follow a trigger strategy equilibrium. Workers whose high wage is lowered after providing low effort do not respond by increasing effort – on the contrary, they tend to reduce effort. This implies that Ghana employers are unable to incentivize workers by disciplining them and, as a result, they tend not to punish shirking workers. We also find that high effort is rewarded by subjects in the UK sessions, but not by those in Ghana. This means that workers cannot induce employers to pay them more by increasing effort. Together these two behaviors lead to an inability to sustain a high-wage-high-effort equilibrium. The contrast between the two subject pools remains when we introduce worker reputation and competition between subjects.

This paper contributes to a large literature on gift-exchange games in experimental economics. The first gift-exchange lab experiment was conducted by Fehr et al. (1993). It consisted of one-shot interactions between workers and employers who were randomly rematched at the end of each period. Gachter and Falk (2002) introduced a treatment of this game in which an employer repeatedly interacts with the same worker for ten periods. They find that repeated interaction functions as a disciplining device, in line with the predictions of Kreps et al. (1982) and Fehr and Schmidt (1999) on cooperation in finite games. Brown et al. (2004; 2012) introduced competition between multiple employers and workers and again find naturally emerging relational contracting: an employer who receives low effort either stops making offers to that worker, or reduce the offered wage; this in turn incentivize the worker to exert effort.

Gift-exchange experiments have since been replicated in numerous high income countries: Austria (e.g., Fehr et al., 1997); Germany (e.g., Abeler et al., 2010; Altmann et al., 2014; Falk et al., 2008; Fehr et al., 2007); Switzerland (e.g., Fehr and Falk, 1999; Falk et al., 2006); the Netherlands (e.g., van der et al., 2001); Portugal (e.g., Pereira et al., 2006); Spain (e.g., Brandts and Charness, 2004); the United Kingdom (e.g., Gachter et al., 2016); and the United States (e.g., Charness, 2004; Cooper and Lightle, 2013; Wu and Roe, 2007). All have found the same general pattern of high effort in return for high wages. Similar results have been found in experiments in former communist countries, such as Hungary (e.g., Falk et al., 1999) and Russia (e.g., Fehr et al., 2014). To the best of our knowledge, however, few gift-exchange experiments have been conducted in less developed countries[3] and none in sub-Saharan Africa. Our contribution to is to carefully document an educated subject population that significantly deviates from the pattern documented to date.

A few field and natural experiments have been conducted on gift exchange, some of them in developing countries. Gneezy and List (2006) investigate whether paying an above-market wage incentivizes workers to exert more effort. de Ree et al. (2018) do the same for a wage increase among school teachers in Indonesia. Both find a short-lived initial boost in productivity that is insufficient to compensate the employer. Foltz and Opoku-Agyemang (2015) find instead that raising the salaries of police officers in Ghana increased bribe solicitations at checkpoints instead of reducing them. In contrast, a reverse experiment by Kube et al. (2013) shows that a pay cut results in a durable fall in productivity. Since none of these experiments allow the employer to vary the wage depending on output, they are more about conditional reciprocity than punishment. In a similar vein, Jayaraman et al. (2016) study the effect of a base wage increase in India that did not affect the piece rate incentive. They find a temporary increase in output. Esteves-Sorenson (2018) compares the effect of a piece rate to that of a flat wage increase, while controlling for experimentally measured prosociality. The author finds no evidence of conditional reciprocity, but a response to incentives. Fafchamps and Davies (2021) compare effort provision under various incentive structures such as high or low fixed wage and a wage contingent on output. They find that workers in India and Africa are more likely than in the US to exert high effort irrespective of wage, while more subjects in the US respond to incentives. In none of these cases is the *future* wage purposefully reduced by the employer after low output as

[2] Most university students in both countries are destined to a career in wage employment. In addition, most college students in Ghana grew up with hired family help.

[3] Siang et al. (2011) conducted a bilateral gift exchange experiment in Malaysia.

punishment for low performance. Our contribution focuses on this important repeated game dimension that has received less attention in the comparative literature to date.

The paper also contributes to a broader literature about the role of labor incentives. Some studies suggest that recruitment through social networks increase effort (e.g., Bandiera et al., 2009). Others have shown that performance pay – i.e., incentives for high effort – can be demoralizing. For example, in a field experiment in India, Breza et al. (2018) find that effort fell after workers started receiving wages (weakly) based on their past effort. By setting incentives, employers de facto peg workers against each other, and the extra effort of one worker imposes a negative externality on others (e.g., Bandiera et al., 2005). Our findings complement this earlier research by showing that, in one developing country context, experimental subjects assigned the role of employer are unable to incentivize workers enough to make a profit.

The paper is structured as follows: Section 2 presents the experiment and predictions based on the theory. The research hypotheses are summarized in Section 3. In Section 4 we present the empirical results in the absence of competition and reputation; these are discussed in Section 5. Section 6 concludes. An extended series of robustness checks are detailed in Appendix B.

2. Experimental design

The experiment is a multi-period gift exchange game based on Brown et al. (2004) and the original gift-exchange game of Fehr et al. (1993). The game is framed as a labor contract in a principal-agent setting.[4] At the beginning of the experiment, participants are randomly assigned the role of worker or employer, a role they keep throughout the experiment. Each game is played for five periods, after which employers and workers are rematched for another game. Each participant plays four games of five periods. Instructions to experimental subjects are provided in Online Appendix A.

2.1. Base games

Each game involves either two or six players. The two player variant involves one worker and one employer (1-on-1 treatment). The six player variant has three workers and three employers (3-on-3 treatment). The sequencing of play is similar in the two variants, but the 3-on-3 treatment includes more steps.[5] The 1-on-1 treatment is similar to a bilateral gift exchange game. The sequence of moves is as follows:

- **Contracting**: At the beginning of first period $t = 1$, the employer makes a wage offer $w_t \geq 0$ and specifies a desired effort level $\widetilde{e}_t$; the worker then either accepts or rejects the offer; if the worker rejects the offer, the game moves to the next period and a new contracting stage begins. The employer can also decide not to offer any contract, or to offer a zero wage. In both cases, both employer and worker earn a zero payoff in that period.
  At the beginning of periods $t = 2$ to 5, the offer made by the employer is normally determined by the decisions made at the rehiring stage of the previous period – see below. If there was no rehiring stage, the contracting stage starts afresh as described above.
- **Effort choice:** If the worker accepts the offer, they then decide an effort level e_t. This effort costs $c(e)$ to the worker with $c'(e) \geq 0$ – additional effort is increasingly costly to the worker. Effort can take one of three possible values: low, medium, or high ($e \in \{L, M, H\}$). The employer collects a revenue $b(e)$ with $b'(e) > 0$ – effort benefits the employer. The payoffs to the employer $\pi_{E,t}$ and to the worker $\pi_{W,t}$ at time t are given by:

$$\pi_{E,t} = b(e_t) - w_t \tag{1}$$

$$\pi_{W,t} = w_t - c(e_t) \tag{2}$$

  Mutual gains are possible if $b(e) > c(e)$ for any e, which we impose throughout. We also select functions $c(.)$ and $b(.)$ such that $b(L) - c(L) < b(M) - c(M) < b(H) - c(H)$, i.e., high effort generates larger gains from trade. The main research question is whether these gains can be achieved in a sustainable and equitable manner.
- **Rehiring**: In this stage we elicit subjects' choices using a strategy method. Before moving to the next period, we ask the employer to make a contingent choice for contract renewal in the next period. At this stage the employer does not yet know the effort level chosen by the worker. For each possible effort choice $e_t \in \{L, M, H\}$ we ask the employer to specify a conditional wage offer $w_{t+1}(e_t)$ and desired effort level $\widetilde{e}_{t+1}(e_t)$. The purpose of this step is to verify that subjects intentionally pursue a trigger strategy, i.e., that they intend to punish a worker who has chosen an effort level lower than stipulated in the contract – i.e., $e_t < \widetilde{e}_t$. The subject is also given the choice not to make any offer.[6]

[4] During the Ghana pilot we experimented with a supplier-producer frame instead. Although play was fairly similar, subjects complained that they did not understand the game. This led us to switch to the employer-worker frame as it was easily understood by all subjects. This minimizes the likelihood that our results are due to mistakes and lack of familiar heuristics.

[5] The group of two or six players remains the same during the five periods of a game. Subjects are identified by the same randomly assigned letter through the game. This letter is reassigned each game.

[6] If the subject runs out of time before choosing a selection or before finalizing a set of conditional offers, it is considered as not having made an offer.

We also ask the worker to specify a reservation wage r_{t+1} below which the worker reject the contract. If the realized contract offer $w_{t+1} \geq r_{t+1}$, the worker is regarded as accepting the offer. The purpose of this step is to investigate whether the worker anticipates a lower offer if $e_t < \widetilde{e}_t$, i.e., whether the worker has internalized the possibility of retaliation by the employer.

Following a rehiring stage, the contracting stage of period $t+1$ is automatic if $w_{t+1} \geq r_{t+1}$: it implements the conditional offer $\{w_{t+1}(e_t), \widetilde{e}_{t+1}(e_t)\}$ that corresponds to the actual effort level e_t. The game then moves to the effort choice of the worker as above before moving to the next rehiring stage. If $w_{t+1} < r_{t+1}$, the conditional offer is not implemented and period $t+1$ starts with the offer stage as explained above: the employer can make a fresh offer and the worker can choose to accept or reject this offer, as before.

- **Rematching**: The above three steps are repeated in sequence until $t = 5$, at which point the game ends. Workers and employers are then rematched for a new game – often with a different treatment. More about this later.

To make the game easy to play, great care has been taken in designing a friendly and efficient interface for subjects. As the screenshots presented in the Online Appendix make clear, at no point in the game are subjects expected to make any calculations themselves. All actions are taken by clicking or moving cursors, and players are always informed of the consequences of their choices on their own payoff and that of the other player. At the end of a period, subjects are provided with a summary of the actions taken during the stage game and their associated payoffs. At the rehiring stage when employers and workers are asked to fill a choice schedule, they are provided with a clear description of the different actions and their consequences, and they are reminded that workers choose effort after accepting a particular wage – see the screenshots for details.

In the 3-on-3 treatment, the sequence of moves is similar to the 1-on-1 treatment, except that it allows for multiple actions by employers and workers.[7] In the contracting stage, employers first make offers to all workers without seeing the offers made by other employers, as in the original game. As for the 1-on-1 variant, great care has been applied to making the subject interface clear and easy, e.g., when making offers to multiple workers – see the screenshots in the Online Appendix for details. After all employers have made offers to workers, the offers of other employers are revealed to employers who have a chance to revise their offer. This is done to maximize competition between employers. Once employers have finalized their offers, one randomly selected (unmatched)[8] worker sees all the offers made to them, and can pick one. If the worker accepts one of the offers, the matched employer is removed, the second randomly selected worker sees the offers from the unmatched employers, and they can select one. The third worker then sees the offer made by the remaining unmatched employer or employers. This process of sequential matching mimics job search in a competitive market, whereby employers post jobs at a given wage while observing other job ads. Workers who reject all offers earn a null payoff for that round.

The effort choice stage of the 3-on-3 game is the same as in the 1-on-1 variant: the worker picks an effort level. The rehiring stage is also similar: the employer is asked to make wage offers for next period – one for each of the worker's (yet unrevealed) effort choice in the current period.[9] Workers are similarly asked to stipulate a reservation wage below which they accept the wage offer for next period. If the wage offer for the realized effort level is above the reservation wage of the worker, employer and worker remain matched for the next round. This mimics what happens in a permanent employment contract: the worker only goes back on the market if the continuation wage offered by the employer is below their reservation wage; otherwise employment continues. It follows that only employers who are unmatched at the beginning of the round make offers and unmatched workers see them. This feature again serves to induce more competition among employers.

The number of effort levels is limited to three to simplify strategy elicitation in the rehiring stage. The values of $c(e)$ and $b(e)$ are as follows. High effort costs the worker 6 points and gives the employer a benefit of 40 points. Medium effort costs the worker 2 points and give the employer 20 points. Low effort is costless to the worker but only gives the employer 5 points. High effort maximizes joint surplus but requires trust: offering a high wage exposes the employer to shirking (i.e., low effort) by the worker.

Our experimental design improves on earlier gift-exchange experiments in a number of ways. The rehiring stage is novel to our experiment. The strategy method allows to investigate whether subjects explicitly pursue conditional strategies: does the employer intentionally reduce the wage offer after low effort by the worker; does the worker intentionally accept a contract conditional on the wage offered? This aspect is important to test the existence of punishment strategies capable of deterring opportunistic behavior. We also allow subjects – under certain conditions – to revisit their strategy after the action of the other player has been revealed. The purpose of this aspect of the design is to test whether the intent to punish is self-commitment-proof, i.e., do players stick to their guns and carry through their punishment strategy, or do they cave in when the desired result fails to materialize. These two features of our experimental design will prove useful in the empirical analysis.

To elicit information about the conditional strategies of employers and workers, our design must depart from earlier experiments in other, less essential ways: since it is not possible to elicit conditional play in continuous time, contracting takes place in discrete stages, not continuously; we reduce the number of effort levels to three to reduce the number of

[7] The detailed order of play is given in Appendix A.

[8] See below.

[9] As in the 1-on-1 games, timing out is treated as not having made offers.

conditional play decisions for employers; and in the multiple workers/multiple employers treatment, we limit the number of workers and employers to three in order to simplify the range of conditional strategies subjects can choose from.

Although these changes are forced upon the experiment by the strategy method design, they also enhance the experiment in many ways. First, when played in continuous time, the experiment tends to reward technical ability, something that puts less experienced subjects at a disadvantage and may create artificial differences across subject pools. Continuous play may also distract subjects from adopting simple conditional strategies, e.g., punishment for low effort. Secondly, we do not introduce excess labor demand or supply in the multiple workers and employer treatment. The literature has shown that this complication is not required for relational contracting to emerge; it only affects the division of surplus (e.g., Brown et al., 2012). Third, we reduce the number of periods in each game from 15 or 20 to five so that we can subject participants to different treatments, to be detailed below. A within-subject design increases power and gives more opportunities for subjects to learn the value of conditional play.[10]

The last difference with earlier work is framing. Some gift-exchange experiments have sought to use a neutral language, describing work as a "good", employers as "buyers" and workers as "sellers". We tried these neutral terms in our Ghana pilot but they decrease the understanding of the game.[11] Understanding improved considerably by framing the experiment as an interaction between an employer and a worker, probably because doing so triggers heuristics more in line with the strategic structure of the game. There is evidence that framing is not a major concern for this type of game: in a gift-exchange game experiment with Munich students, Fehr et al. (2007) found that using a neutral frame or a labor market frame does not not produce different behavior.

While each of these changes, taken individually, should not have a dramatic effect on subjects' behavior, taken together they make the game simpler and more intuitive. This in turn should make subjects more likely to follow behavioral norms with which they are already familiar. This is important because we are not interested in how our subjects behave in a highly unusual and unintuitive setting. We want the experiment to reveal, in a way least contaminated by experimental artifacts, how the subjects naturally tend to behave in a labor relationship.

We are particularly keen to ascertain the generality of the trigger strategies documented in Western student populations. Many of the simplifications we have introduced make conditional play more salient, a feature that we deliberately set to reinforce. We nevertheless remain concerned that findings for Ghana subjects may be driven by design differences with earlier work. To address this concern, we repeat the experiment with university students in the United Kingdom. As we show below, our results with United Kingdom students are similar to earlier experiments. This provides reassurance that making conditional play more salient does not have the paradoxical consequence of making it less common.

2.2. Other treatments

In addition to the 1-to-1 and 3-to-3 treatments described above, we vary whether contract compliance is enforced (treatment C) or not (treatment E); and whether information about the past actions of workers is automatically shared among all employers (treatment S). Since the latter treatment only applies in the 3-to-3 treatment and is only relevant when the contract is not externally enforced, there is a total of five possible treatment configurations.

The control treatment is when the contract is externally enforced, which means that the worker can only provide the level of effort stipulated in the employer's offer. The 1-on-1 and 3-on-3 versions are denoted 1C and 3C, respectively. These treatments are essentially a modified version of an ultimatum game: the worker can only accept or reject the division of gains from trade proposed by the employer, and refusal yields a null payoff for both.

Treatments 1E and 3E are as described in the previous sub-section: after accepting a contract, the worker is free to choose any of the three effort levels. Treatment 1E is similar to a bilateral gift-exchange game with a fixed partner (e.g., Kirchler et al., 1996; Gachter and Falk, 2002). In contrast, the 3-on-3 treatment 3E allows competition between employers and workers. It is closest to the multilateral gift-exchange games conducted by Brown et al. (2004). Treatment 3ES only differs from 3E in that information about the past actions of each worker is available to all three employers. Treatment 3ES allows for a multilateral reputation mechanism, while 3E and 1E only allow for bilateral reputation/relational contracting.

Each participant plays four games of five periods.[12] This setup is designed to allow comparisons within and between subjects, and to facilitate the gradual introduction of more complicated treatments. These treatments allow us to compare the impact of imperfect enforcement, the role of competition (increasing the number of employers and workers), and the role of sharing information between employers. In treatment 1C and 3C the worker has to comply with the contractual level of effort. Comparing these treatments with treatments 1E and 3E estimates the impact of imperfect enforcement on effort choice. In treatments 3C, 3E and 3ES, there is competition between workers and between employers. Comparing these treatments with treatment 1C and 1E estimates the impact of having a larger market on wage offers and on effort. Finally,

[10] Whether five periods are sufficient for repeated game reasoning to kick in is an issue we examine in detail in the empirical section. There we also discuss a follow-up experiment in which the number of periods was increased without affecting our main findings (Davies and Fafchamps, 2017).

[11] This is evidence by subjects' answers to comprehension questions asked at the end of the pilot. In particular, subjects found counterintuitive that the buyer makes a take-it-or-leave-it price offer because it contradicts what they observe in everyday purchases where the price is set by the seller.

[12] The distribution of subjects into different treatments and the seven treatment sequences are shown in Table A1 and A2 in the Online Appendix, respectively.

comparing treatments 3E and 3ES tests whether information on past effort results in a reputational equilibrium in which employers offer higher wages to workers who have supplied higher effort to other employers in the past.[13]

2.3. Implementation

The participants to the study were recruited among students from colleges and universities in Accra, Ghana, and Oxford, United Kingdom.[14] Because not all subjects have the citizenship of the country in which the experiments are implemented, we refer to them as 'Ghana subjects' and 'UK subjects' rather than as Ghanaian or British.

In Ghana a total 16 sessions were held, with 18 to 20 participants each and a total of 304 participants. In the UK we held 13 sessions, with 192 participants in total. Sessions lasted between 1.5 and 2 hours. The points earned during the session were converted to Ghana cedis or British pounds at the end of each session, with an exchange rate of 0.05 Ghana cedis and 0.03 British pound for every point. Including the show-up fee, average earnings are 32 Cedis (about 10 British pounds) in Ghana and 18 pounds in the UK.[15]

For the experiment, we developed our own tablet-based mobile lab, LabBox. This platform can operate completely independently from the electricity mains and the existing IT network. The experiment runs on 7-inch Android tablets with a custom-built app. This app collects user input and communicates with a LabBox server using a wireless connection. Each session starts with a 15 minute instruction on how to use the touch screen, followed by an extensive demonstration of how the game is played. The experiment is entirely conducted in English, which is the language of instruction in the higher education system of both countries. To make sure that participants are always fully cognizant of the payoff implications of their actions, we provide visual on-screen aids that display to participants the prospective earnings of the choices they are about to make (such as making a job offer or setting an effort level). Screenshots are provided in the Online Appendix for illustration. This is to ensure that differences in behavior between subjects are not driven by differences in their ability to calculate payoffs or memorize game rules. Subjects also play three practice rounds before the experiment starts.

The experimental sessions in Ghana were held in September 2013 in the central Osu neighborhood of Accra. The UK sessions took place at the Oxford CESS lab in November 2015 and between January and May 2016. These sessions were preceded by an extensive pilot held in Ghana in April 2013 and involving 4 sessions with 48 students and 20 small entrepreneurs. This pilot served to test the visual interface used in the experiment and to refine the experimental design. As a result of the pilot, changes were introduced to make the game easier to understand. These improvements were, among other things, intended at facilitating the emergence of a trigger strategy. They did not change the main findings of the paper, which are also present in the pilot, both with student subjects and with small entrepreneurs.

3. Hypotheses

Our main interest is whether subjects in both Ghana and the UK follow a trigger strategy similar to what has been described in the literature for this game. If UK subjects are found not to follow such strategy, we will not be able to rule out that the changes we made to the experimental design caused the change in behavior. If, however, the behavior of UK subjects conforms to the existing literature, they should achieve a high level of effort by following a trigger strategy.

Our main alternative hypothesis is that Ghana subjects pick another equilibrium of the repeated game. One strong candidate suggested by the theory of finitely repeated games is the Nash equilibrium of the stage game. Since wage offers in a round are final, the workers' best response is to choose low effort irrespective of wage and, anticipating this, the employers' best action is to offer a low wage. The static Nash equilibrium is thus a low-effort-low-wage equilibrium.

For a trigger strategy to emerge, workers have to expect to be punished for choosing low effort when paid a high wage. Employers must similarly understand that workers are reading changes in their wage offers as rewarding or punishing a past level of effort – and act accordingly. If neither workers nor employers *expect* to see a trigger strategy, the static Nash equilibrium is a likely outcome.

To investigate whether subjects follow a trigger strategy, we combine a standard reduced-form analysis – e.g., documenting effort choices in different treatments and over different rounds – with a more detailed analysis of behavioral play. While we do not observe subjects' expectations, we do observe employers' conditional wage offers – which we elicit through the strategy method. We can therefore test whether employers condition wage offers on past effort. We also test whether, after providing an effort level lower than requested, workers expect a punishment and thus set a low reservation wage, and whether they increase effort after being punished.

[13] In the United Kingdom, we only conduct the three treatment sequences that are most relevant to demonstrate the comparability of our findings with the literature.The purpose of the UK sessions is to test whether differences in findings between our Ghana and earlier experiments can be ascribed to variation in design. To achieve this, we only need to replicate in the UK the effort choice treatments with exactly the same design as in Ghana. See Table A2 for a treatment summary.

[14] Our UK subjects include students from the University of Oxford students and other universities in the Oxford area – notably Oxford Brookes University, a former polytechnic school.

[15] During recruitment, participants were told that, regardless of their earnings, they would receive a minimum payment of 10 Cedis. To prevent employers from receiving a negative payoff at the end of the experiment, subjects were all given an initial endowment for each game (70 points for workers and 60 points for employers). One Ghana subject given the role of employer ended the four games with a negative earning. This subject was given the minimum advertised payment of 10 Cedis.

We examine other possible behavioral theories as well. One is conditional reciprocity by workers, which means choosing high effort when offered a high wage.[16] In equilibrium, this behavior is observationally similar to the trigger strategy: if employers know workers are conditional reciprocators, we should also observe high-wage-high-effort in all rounds and games. The conditional reciprocity equilibrium nonetheless differs from a trigger strategy equilibrium in one important way: workers are never punished for choosing low effort – since offering a low wage would be counter-productive, i.e., it would 'cause' low effort. This can be tested by examining the *conditional* wage offers made by respondents:[17] in a trigger strategy, employers respond to a lower than requested effort with a low wage offer in the next period; but if employers expect conditional reciprocity by workers, they offer a high wage, irrespective of past effort.

Another possibility we consider is that employers expect workers to be intrinsically motivated and provide effort irrespective of the wage offered. In this case, the optimal strategy is to offer a low wage and receive whatever effort level the worker provides, which can result in a low-wage-high-effort equilibrium. This again can be tested by examining conditional wage offers: in a trigger strategy equilibrium, a high effort is followed by a high wage offer; in an intrinsically motivated worker equilibrium, the wage offer is always low and does not depend on past effort.

The above considerations can be generalized to a world with multiple worker types: conditional reciprocators; intrinsically motivated; and selfish-rational (Davies and Fafchamps, 2021). In this world, the employers' best strategy depends on their beliefs about the distribution of worker types in the population. If they believe that the proportion of conditional reciprocators is high but some are selfish-rational, the optimal strategy is to start by offering a high wage. If the worker provides low effort, the employer permanently switches to a low wage since the worker is revealed to be selfish-rational. If the worker provides high effort and beliefs about the proportion of intrinsically motivated workers is low, the employer continue offering a high wage. If the employer believes the proportion of intrinsically motivated workers to be high enough among non-selfish-rational workers, however, the employer switches to offering a low wage. If the worker provides high effort, this means that the worker is selfish-rational and the employer continues with a low wage. If the worker provides low effort, this means that the worker is a conditional reciprocator, and the employer switches back to high wage.

Regarding offer acceptances, we first note that selfish-rational workers should accept all non-zero wage offers since they can always guarantee themselves a non-negative payoff by choosing low effort.[18] Conditional reciprocators may nonetheless reject contracts with a low wage offer and a medium or high requested effort because they intend to respect the contract or because they find unfair contracts offensive. Selfish workers may nonetheless mimic this behavior to induce the employer to raise the wage and to extract a higher payoff from low effort later. Increasing the wage after an offer rejection is thus a risky strategy for employers if the proportion of selfish workers in the population is high.

We look at the experimental data for evidence that a majority of subjects follow any of these strategies. If the subjects are found not to follow a trigger strategy, we look for evidence of other possible beliefs and strategies. We also investigate the extent of conditional reciprocity by assigning subjects a level of conditional altruism based on their behavior and we compare the result across out two subject populations.

4. Empirical results

We now present the results from the experiment. All regressions pool Ghana and UK subjects but allow all coefficients to differ between the two study populations. Errors are always clustered at the session level. Since we have less than 30 sessions in total, we use a wild cluster bootstrap to obtain p-values (e.g., Cameron et al., 2008). Standard errors are not reported since the wild bootstrap does not produce them.

We start by focusing on three main hypotheses: (1) do employers offer wages higher than what is typically observed in finitely repeated games such as ours; (2) do workers reciprocate conditionally by exerting high effort when receiving a high wage; and (3) do employers reciprocate conditionally by offering a high wage following high effort. The first hypothesis relates to a large literature showing that experimental subjects placed in a finitely repeated game are capable of improving on the Nash equilibrium of the stage game, i.e., low effort and low wage. The other two hypotheses come from the literature on relational contracting: by conditioning high wage on high effort and vice versa, players can establish an incentive structure that sustains cooperation.

Before delving into the analysis proper, it is useful to take a look at Table 1. It provides a summary of average play for all treatments in the United Kingdom and Ghana. The Table shows the average offered wage, the share of accepted individual offers, compliance with demanded effort, and average earnings. We find little difference across the two experimental populations in terms of wage offers, share of accepted offers, and worker payoff. But in all treatments where workers choose their effort level (i.e., 1E, 3E and 3ES), there is a difference in terms of effort compliance, and the employer's average payoff

[16] It is also conceivable that employers also behave as conditional reciprocators, i.e., by reducing the wage of workers who shirk. If workers expect employers to behave in this way, it would discipline them – much in the same way as a trigger strategy. There are two conceptual difficulties with this strategy: (1) after observing shirking, employers are conflicted between reciprocating by lowering the wage or incentivizing future effort by offering a high wage; (2) if employers opt for the first option, the continuation equilibrium is low-wage-low-effort, making cooperation not part of a trembling-hand equilibrium. We ignore this possibility here, but discuss it below when considering heterogeneity in worker types.

[17] Since these conditional offers are off the equilibrium path in both the trigger strategy and the conditional reciprocity strategy, they would not be observed without resorting to the strategy method.

[18] In the trigger strategy, workers must 'take their punishment' by accepting a low-wage-high-effort contract to redeem themselves and get a high wage offer later. Hence we should not observe offer refusal either if subjects follow a trigger strategy.

Table 1
Average wage, share of accepted offers, compliance, and earnings.

Treatment		Ghana sessions					UK sessions				
		Average wage offer	Share of accepted offers	Average compliance	Average employer payoff	Average worker payoff	Average wage offer	Share of accepted offers	Average compliance	Average employer payoff	Average worker payoff
1C	Game 1–4	20.0	78%	100%	12.8	16.2	18.6	78%	100%	17.0	14.3
1E	Game 2–4	14.0	87%	41%	0.0	13.1	13.2	81%	60%	6.7	12.3
3C	Game 2–4	20.1	30%	100%	12.2	17.0	n.a.	n.a.	n.a.	n.a.	n.a.
3E	Game 3–4	13.4	30%	43%	0.3	13.3	13.1	30%	66%	7.6	12.7
3ES	Game 4	12.3	31%	50%	1.5	12.3	13.2	29%	76%	9.4	12.7

Note: The above averages pool observations from all games with the same treatment. Compliance is a dummy equal to one if the worker chose the requested level of effort or higher. For treatments 3C, 3E and 3ES, the share of accepted offers is the proportion of offers that were accepted, regardless of which worker it was made to. It is smaller than in treatments 1C and 1E because employers can make offers to up to three workers but only one can be taken up. Treatment 3C was not played in the UK sessions.

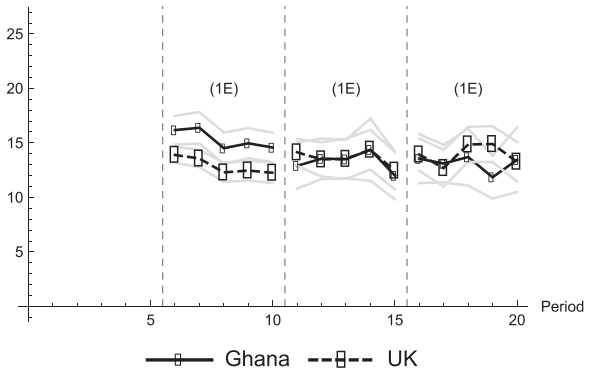

Fig. 1. Average wage offers in treatment 1E for Ghana and the UK. *Note:* Each five-period block represents a separate game. Only treatment 1E games are shown here (see Table A2). They are never played in the first five periods. The 95% confidence interval is shown in grey.

is close to zero in Ghana and much lower than in the UK sessions. What drives this difference is the focus of the rest of our analysis.

4.1. Contract offers

We start by investigating our first hypothesis, namely, that employers offer a wage above the subgame perfect equilibrium of finitely repeated games, which is 0 or 1 point in treatment 1E. Fig. 1 shows, for each of the three 1E games, the average wage offer in each periods for both Ghana and the UK. In game 2 the average wage is 14.9 points in Ghana and 12.9 in the UK, with a slight downward trend across the five periods. This drops in games 3 and 4 to an average offer of 12.6 and 12.8 points in Ghana and 13.6 and 13.9 in the UK, with little noticeable trend over time. The differences between the Ghana and UK sessions are mostly non-significant.[19]

Average offers are significantly higher than the Nash equilibrium of the stage game. This finding is in line with earlier bilateral gift-exchange experiments (e.g., Kirchler et al., 1996; Fehr et al., 1998; Gachter and Fehr, 2001). We also find no drop in wage offers in the last period of each game, as is sometimes found in finitely repeated games. This suggests that the games' short duration is unlikely to drive our results. Finally we note that average offers are higher than the employer's revenue with low effort, which is 5 points. Hence unless the worker chooses high or medium effort, the employer suffers a net loss: $b(e) - w < 0$ in this case. A high wage may induce a worker to reciprocate with high effort, but it leaves the employer vulnerable if reciprocation does not occur. This feature is at the heart of trust games and gift-exchange games.

The averages reported so far pool data from all periods and therefore partially incorporate the employer's response to the worker's choices. In contrast, the wage offered in period 1 cannot, by construction, depend on past worker effort and is more informative of the employer's initial expected effort. We do not, however, find different results and none of the differences between the UK and Ghana sessions are statistically significant.

In the UK the distribution of wage offers in treatment 1E is multimodal, with peaks around 3, 11 and 23 points. These levels roughly correspond to equal payoffs for employer and worker when the worker chooses low, medium or high effort,

[19] A t-test for game 2 yields a p-value of 0.066. For games 3 and 4 the corresponding p-values are 0.878 and 0.738. All p-values are corrected for clustering at the individual level.

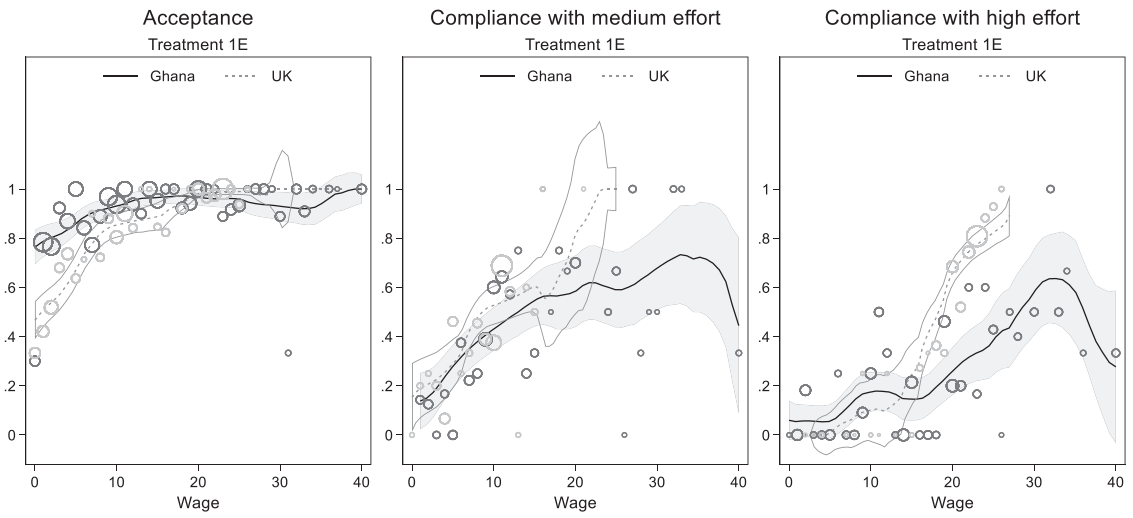

Fig. 2. Non-parametric regression of acceptance and compliance on wage offered. *Note:* The Figure is constructed by combining observations on all periods from all games played under Treatment 1E. Two non-parametric regressions are presented, one for the Ghana sessions in black and one for the UK sessions in grey. Both regression lines are accompanied by their 95% confidence interval. Circles represent observations, with the size of the circle representing the number of observations. Ghana circles appears in black and UK in gray. Compliance is defined as providing an effort level equal or above that requested by the employer.

respectively. In Ghana the distribution of wage offers is more spread out than in the UK. Non-parametric tests of equality of distribution nonetheless show that these differences are mostly non-significant.[20]

We also find that average wage offers in treatment 1E are lower than in treatment 1C when workers cannot choose effort: wage offers are on average 19.8 in Ghana and 18.6 in the UK for treatment 1C, compared to 14.9 in Ghana and 12.9 in the UK for treatment 1E. These differences between treatment 1C and 1E are significant for both countries, with p-values smaller than 0.001.[21] This is in line with Brown et al. (2004)'s findings, which the authors attribute to contractual incompleteness.

Next we examine the effort levels requested by employers in treatment 1E. For 51% of the offers in both Ghana and the UK, employers demand high effort. A substantial fraction of employers nonetheless request low effort: 12.3% of offers in Ghana and 14.2% in the UK. In most cases this occurs in combination with a low wage offer, and indicates a lack of trust in the worker. Although low wage/low effort is the Nash equilibrium of the stage game, the employer could have requested high effort since the worker can adjust effort downwards anyway. The data shows that UK workers tend to reject low wage/high effort offers, even though the requested effort is not binding.[22] To examine these patterns more in detail, we turn to the choices made by workers.

4.2. Acceptance and effort choice

Next we test our second hypothesis, i.e., conditional reciprocity: do workers reciprocate a high wage with high effort, and do they respond to a low wage offer either by rejecting the offer or applying low effort. We start with acceptances and then turn to effort choices. To recall, offer rejection is not consistent with a trigger strategy.

Across all periods and games, the proportion of rejected offers is 12.3% in Ghana and 19.5% in the UK. Fig. 2 displays non-parametric regressions of acceptance and compliance rates on the wage offered. We see that workers are more likely to reject low wage offers than high wage offers: offers of five points or less are rejected 23.6% of the time by Ghana subjects and 45.7% of the time by UK subjects – a difference that is statistically significant. In Table 2 we present the results of a linear probability model of acceptance and compliance on the wage offered. They confirms that the relation between wage offer and acceptance is positive and that it is stronger for UK than Ghana subjects: a wage increase of one point increases the acceptance probability by 0.65 percentage points in Ghana and by a significantly higher 2.53 (= 0.65 + 1.88) percentage

[20] Both the Kolmogorov-Smirnov and the Mann-Whitney U rank-sum tests fail to reject the null hypothesis of equal distributions (e.g., in game 2, the Kolmogorov-Smirnov test gives $p = 0.393$ and the U-test gives $p = 0.234$, with $Z = -1.189$). The tests are conducted with unmatched data pairs. The offer is averaged across the five periods for each employer such that each employer counts as one observation. Wherever appropriate, the tests are two-sided and exact t statistics are used.

[21] In Online Appendix D we calculate the effect of imperfect enforcement using within-subject, between-subject, and difference-in-difference approaches. Most tests confirm that the difference in wage offers between treatment 1E and 1C is statistically significant.

[22] For wage offers of five points or less, UK workers reject 64.5% of the offers asking for high effort, but 51.4% and 28.6% of the offers asking for low and medium effort, respectively. We find no such differences in Ghana: rejection rates for low wage offers asking for low, medium and high effort are 24.0%, 21.2% and 25.0%, respectively.

Table 2
Workers' actions (Treatment 1E only).

	(1)		(2)	
	Medium or high effort requested			
Dependent variable:	Acceptance of wage offer		Compliance with requested effort	
Wage offered	0.0065*	($p = 0.093$)	0.0148**	($p = 0.046$)
High effort requested	−0.0971*	($p = 0.094$)	−0.2520**	($p = 0.016$)
Wage × High effort requested	0.0038	($p = 0.371$)	−0.0032	($p = 0.584$)
UK dummy interacted with:				
Wage offered	0.0188**	($p = 0.028$)	0.0229*	($p = 0.060$)
High effort requested	−0.1570	($p = 0.144$)	−0.0885	($p = 0.650$)
Wage × High effort requested	−0.0003	($p = 0.979$)	0.0002	($p = 0.987$)
Worker fixed effects	Yes		Yes	
Period dummies	Yes		Yes	
Period dummies × UK	Yes		Yes	
Intercept	0.6970		0.2400	
Number of observations	1145		1023	
Number of clusters	20		20	
Adjusted *R*-squared	0.29		0.44	

Note: Each regression is a linear probability model, pooling the Ghana and UK sessions and only using observations from treatment 1E. Only observations for which medium or high effort is requested by the employer are included. The dependent variables are: a dummy equal to 1 if the offer was accepted (column 1); and a dummy equal to 1 if the worker provided the requested level of effort or higher (column 2). The number of observations is lower in column 2 since compliance is only observed if the offer was accepted. The UK dummy is equal to 1 in the UK sessions, and 0 in the Ghana sessions. The inclusion of worker fixed effects controls for individual differences in the likelihood of accepting or complying. The above coefficients should therefore be interpreted as workers' response to contractual variation during the experiment. Period dummies are included for each of the periods from 6 to 20 (Treatment 1E is not played in the first game). Reported *p*-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

points in the UK. This confirms that high wage offers are more likely to be accepted, a result in line with conditional reciprocity at the offer stage.

Next we turn to effort levels. To recall, non-compliance should not arise in a trigger strategy equilibrium. But it could be observed if workers are 'testing' whether their employer retaliates for not providing the requested effort. We find considerable non-compliance in both Ghana and the UK in treatment 1E. But, as shown in Table 1, compliance is higher among UK than Ghana subjects: in the Ghana sessions, compliance levels are around 40% throughout, with no evidence that subjects are learning to discipline workers over the course of the experiment; in the UK they rise steadily from 56% to 72% between games 2 and 4, suggesting increased coordination on the efficient effort level. This increase in compliance is consistent with the idea that, having tested employers' response to non-compliance, workers faced a fall in wage and, in response, they decide to comply – i.e., they fall in step with a trigger strategy.

To summarize, we find evidence of conditional reciprocity at the offer stage for workers. This positive correlation between wage offer and effort choice corresponds to findings from some earlier studies. We similarly find a positive relationship between wage and effort compliance – in the non-parametric regressions of Figure 2 as well as in the linear probability model of Table 2.[23] According to the latter, a one point wage increase is associated with a 1.48 percentage point increase in the probability of compliance in Ghana, and 3.77 (= 1.48 + 2.29) percentage points in the UK. Both increases are significantly different from zero at the 5% level and the difference between the UK and Ghana sessions is significant at the 10% level.

4.3. Revision of wage offers by employers

Having found evidence of conditional reciprocity in the behavior of workers, we now look for evidence of the use of a trigger strategy by employers. To this effect, we ask whether employers condition their wage offers on past effort levels, i.e., reward high effort with a high wage offer next period, punish low effort with a low wage offer, or make no offer at all. As argued in models of relational contracting with a trigger strategy, such behavior creates an incentive for workers to choose high effort today if they expect it to be rewarded with a wage increase in future periods.

We first focus on wage revision in the period following a wage offer above the median of 15 points. For these cases, Table 3 shows the employer's response in treatment 1E to low, medium and high effort. Panel A only includes wages offered in the second period, conditional on effort; Panel B pools data from periods 2 to 5.

[23] As can be seen in Fig. 2, the relationship between wage and compliance has an inverted U-shape. This result is only driven by a few observations above 35 points, and is mainly driven by low effort choices following offers of 40 points. Such high wage offers are difficult to rationalize since, even with high effort, the employer's payoff is zero.

Table 3
Response of employers to the past effort of workers paid an above-median wage (Treatment 1E only).

	Ghana sessions				UK sessions			
	Worker effort in period $t-1$ for above median wage in $t-1$				Worker effort in period $t-1$ for above median wage in $t-1$			
	Low	Medium	High	All	Low	Medium	High	All
Panel A. Response in period 2								
Share of observations in period $t-1=1$	37.0%	39.1%	23.9%	100.0%	10.1%	33.3%	56.5%	100.0%
Employer's response in period $t=2$:								
Decrease the wage offered by more than 2 points	41.2%	38.9%	54.6%	43.5%	71.4%	60.9%	10.3%	33.3%
Offer the same wage +/- 2 points	29.4%	38.9%	36.4%	34.8%	0.0%	26.1%	84.6%	56.5%
Increase the wage offered by more than 2 points	29.4%	16.7%	9.1%	19.6%	14.3%	8.7%	5.1%	730.0%
No offer	0.0%	5.6%	0.0%	2.2%	14.3%	4.4%	0.0%	2.9%
Panel B. Average of responses in periods 2 to 5								
Share of observations in period $t-1$	41.3%	32.3%	26.4%	100.0%	11.0%	20.9%	68.1%	100.0%
Employer's response in period:								
Decrease the wage offered by more than 2 points	50.7%	29.6%	45.5%	42.5%	70.0%	70.2%	6.5%	26.7%
Offer the same wage +/- 2 points	24.6%	46.3%	47.7%	37.7%	10.0%	19.3%	88.2%	65.2%
Increase the wage offered by more than 2 points	21.7%	20.4%	4.6%	16.8%	13.3%	8.8%	4.3%	6.2%
No offer	2.9%	3.7%	2.3%	3.0%	6.7%	1.8%	1.1%	1.8%

Note: The first row of each Panel gives the horizontal breakdown of effort levels in the previous period. The four other rows give the vertical breakdown of offers made, conditional on the worker's effort level in the previous period. Only observations following above median wages are included in the Table. Observations on offers rejected in period $t-1$ are not included since they have no recorded effort. Only observations for Treatment 1E are included.

The Table highlights two main differences between UK and Ghana subjects. First, Ghana subjects are less likely than UK subjects to decrease their final wage offer following low effort: 41% of Ghana subjects lower their wage offer in period 2 following low effort in period 1; the corresponding figure for UK subjects is 71%.[24] When we pool periods 2 to 5, these figures are respectively 51% and 70%. Second, after high effort, UK subjects are more likely to keep their wage offer unchanged: 85% of UK employers offer the same wage compared to 36% among Ghana subjects. If anything, Ghana employers are more likely to lower their wage offer after high effort: 55% lower their wage offer in period 2 after high effort, compared to 10% in the UK. These findings are broadly in line with a trigger strategy equilibrium for the UK subjects. In contrast, the behavior of Ghana subjects is more consistent with employers viewing workers as intrinsically motivated and seeking to take advantage of that.

These findings are confirmed when using regression analysis (see Table 4). Among UK subjects, compliance with a high effort request is associated with a 6.26 (= 5.65 + 0.61) points increase in wage offer, a result that is significant at the 1% level. In Ghana, the corresponding coefficient is 0.61 and is not statistically significant. The difference between the two subject pools is 5.65 and is significant at the 1% level.

Taken together, these results confirm that UK subjects are more likely to lower their wage offer following low effort and to keep their wage offer unchanged after high effort. Neither of these behavioral patterns is present among Ghana subjects. The fact that Ghana subjects do not naturally adopt a punishment-and-reward trigger strategy to discipline workers stands in a stark contrast with other experiments conducted in developed countries, including our own replication in the UK.

Could it be that Ghana employers initially intend to punish low effort but cave in when the worker refuses a low wage offer? To throw light on this question, we use the fact that, before employers observe the effort level selected by the worker in round t (i.e., low, medium, or high), they state whether they would reemploy the worker and, in that case, stipulate a period $t+1$ wage offer for each of these effort levels. This is the so-called strategy method.

Answers, shown in Table 5, confirm the stark difference between Ghana and UK subjects. In the UK, most employers follow a deliberate trigger strategy: planned re-employment frequencies and wage offers increase steadily in effort level in all three treatments 1E, 3E and 3ES. In contrast, Ghana subjects are much less likely to stipulate a conditional wage offer: less than 20% of them are willing to specify a wage even after high effort, compared to 60–70% of UK subjects. Furthermore, when Ghana subjects do make a conditional wage offer, the wage offered varies much less by effort than the offers made by UK subjects: average offers in Ghana range from 11 to 13 units after low effort, to 13–19 units after high effort, compared with UK offers ranging from 2 to 4 units for low effort to 19–20 units for high effort. We also note that a larger proportion of Ghana subjects (73.2% vs. 33.1% for low effort) fail to make conditional offers because they 'time out' on the question – either due to lack of interest or because they find the decision difficult to make. Either way, this confirms that conditioning the wage on past effort is not an idea familiar to a large fraction of subjects in the Ghana sessions.

To confirm the statistical significance of these differences, we report in Table 6 an employer fixed-effect regression of wage offers on hypothetical effort levels for each of the three treatments in Ghana and the UK. Results for Ghana show some evidence of conditional offers only in treatment 1E – when there is no competition with other employers. In contrast,

[24] To account for the use of a touch screen, we follow Table 3 in adopting a 2 point difference for an offer to be regarded as the "same" wage. The results do not change qualitatively when no margin or a margin of 1 point are used.

Table 4
Wage offered by employer (Treatment 1E only).

Dependent variable: Wage offer in period t	(1) Periods 2–5 only		(2) Periods 2–5 only	
Rejection by worker in period $t-1$	0.378	($p=0.871$)	0.519	($p=0.836$)
Compliance by worker in period $t-1$	0.606	($p=0.430$)		
Compliance by worker in period $t-1$ (high effort requested)			1.538	($p=0.484$)
Compliance by worker in period $t-1$ (medium effort requested)			1.420	($p=0.144$)
UK dummy interacted with:				
Rejection by worker in period $t-1$	2.109	($p=0.294$)	2.126	($p=0.305$)
Compliance by worker in period $t-1$	5.654***	($p=0.000$)		
Compliance by worker in period $t-1$ (high effort requested)			8.676***	($p=0.004$)
Compliance by worker in period $t-1$ (medium effort requested)			3.486*	($p=0.055$)
Employer fixed effects	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes
Period dummies × UK	Yes	Yes	Yes	Yes
Intercept	13.26	(se = 0.637)	12.82	(se = 0.611)
N. Observations	1055		0.095	
Number of clusters	20		20	
Adjusted R-squared	1055		0.160	

Note: The Table shows two OLS regressions on the pooled Ghana and UK sessions. Only observations from periods 2 to 5 are used since, in period 1, by construction, the explanatory variables do not exist. The dependent variable is the wage offer made in period t. Rejection is an indicator variable equal to one if the worker rejected the offer in the previous period, and 0 otherwise. Compliance is an indicator variable equal to one if the worker chose an effort level equal to or higher than the requested level, and 0 otherwise. The regression in column (2) splits compliance into one dummy equal to one if high effort was requested (0 otherwise), and another dummy equal to one if medium effort was requested (0 otherwise). Only wage offers in period 2–5 are included. Reported p-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Only observations for Treatment 1E are used in this Table.

Table 5
Conditional wage offers at the rehiring stage.

	Ghana sessions				UK sessions			
	If the effort level of the worker was:				If the effort level of the worker was:			
	Low	Medium	High	N.obs.	Low	Medium	High	N.obs.
Panel A. Treatment 1E								
% of employers who re-offer a wage to the worker	6.4%	11.5%	18.1%	409	10.8%	36.7%	61.3%	499
Mean offered wage	13.0	16.9	19.4		2.7	10.9	19.4	
% of offers above reservation wage of worker	5.6%	10.5%	5.3%		0.0%	15.9%	51.1%	
Panel B. Treatment 3E								
% of employers who re-offer a wage to the worker	11.2%	12.9%	19.3%	518	10.9%	38.1%	69.6%	339
Mean offered wage	11.7	12.0	13.8	3.8		11.1	19.6	
% of offers above reservation wage of worker	13.3%	5.6%	16.7%		42.8%	24.1%	52.1%	
Panel C. Treatment 3ES								
% of employers who re-offer a wage to the worker	16.2%	19.2%	19.2%	198	13.2%	24.3%	66.9%	136
Mean offered wage	11.2	13.8	13.2		2.7	10.3	20.0	
% of offers above reservation wage of worker	5.8%	25.0%	33.3%		30.7%	6.7%	54.0%	

Note: At the rehiring stage of a period in which their offer was accepted, employers are given the chance to stipulate a conditional offer schedule to the same worker, before knowing the actual effort level chosen by the worker (strategy method). Workers are simultaneously asked to set a conditional reservation wage above which they automatically accept a combination of wage offer and requested effort. The Table presents summary statistics on the conditional wage offers made by employers. The number of observations corresponds to the number of cases in which employers could select a conditional wage for the next period – no information was collected in the last period of a game, or when no offer by the employer was accepted in the previous period. There are fewer observations for Treatment 3ES because it was only introduced in game 4. All variables refer to the same worker as the one who last worked for the employer. The mean offered wage is the average wage offered, conditional on re-offering. Offers that are equal or higher than the reservation wage stipulated by the worker are automatically implemented at the beginning of the next period.

UK subjects show strong evidence of conditional play in all three treatments. Furthermore, even in treatment 1E where Ghana subjects do condition on effort, the range of offers they make is narrower than those of UK subjects. From this we conclude that Ghana subjects assigned the role of employer are more reluctant to condition wage on effort than their UK counterparts. Put differently, it is not the case that Ghana subjects intend to punish low effort but subsequently cave in when the worker demands a high wage; rather, they show little a priori desire to punish workers for low effort.

What about workers? Do they anticipate being punished for low effort and thus are more likely to subsequently accept a low offer – i.e., to 'take their punishment'? To examine this issue, we rely on the fact that, after choosing an effort level, workers are asked to stipulate a reservation wage above which an employer's offer is automatically accepted.

Table 5 reports the automatic acceptance rate for offers made by employers, conditional on effort. We see that, in Ghana, the acceptance rate is low for all effort levels: even though Ghana employers offer higher wages after low effort than

Table 6
Regression of conditional wage offers at the rehiring stage.

Dependent variable:	(1)		(2)		(3)	
Wage offered for period $t+1$ conditional on effort at period t						
High effort	7.789**	($p = 0.024$)	0.296	($p = 0.367$)	−0.496	($p = 0.925$)
Medium effort	4.555**	($p = 0.047$)	0.285	($p = 0.560$)	0.913	($p = 0.542$)
UK dummy interacted with:						
High effort	6.437*	($p = 0.078$)	15.160***	($p = 0.000$)	9.806**	($p = 0.027$)
Medium effort	2.627	($p = 0.155$)	8.553***	($p = 0.000$)	3.396	($p = 0.246$)
Employer fixed effect	Yes		Yes		Yes	
Intercept	6.079	(se = 0.815)	6.736	(se = 0.609)	10.350	(se = 1.758)
N. Observations	690		627		250	
Number of clusters	20		23		13	
Adjusted *R*-squared	0.664		0.720		0.795	

Note: The Table shows two OLS regressions on the pooled Ghana and UK sessions. The dependent variable is the conditional wage offer made by employers at the rehiring stage, before the effort level chosen by the worker is revealed. The number of observations is less for Treatment 3ES because this treatment was only introduced in Game 4. Since respondents can set up to three different wage offers, one for each effort level, there can be up to three separate observations per conditional offer. But if no offer by the employer was accepted in the previous period, these observations are missing. The constant term can be interpreted as the average wage offer conditional on low effort. The high effort and medium effort coefficients represent the additional wage offered conditional on high or medium effort, respectively. Reported *p*-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

UK subjects, there is no evidence that these wages are more likely to be accepted. In fact, in treatments 3E and 3ES the opposite is true: UK employers make very low offers, but these offers are above the worker's reservation wage in 30–43% of the cases – suggesting that workers anticipate being penalized for low effort. In Ghana, the frequencies are much lower at 6–13%. Among UK subjects, we also observe that workers accept the majority of conditional wage offers made after high effort – suggesting some kind of convergence towards a mutually acceptable remuneration for high effort. The same is not true for Ghana subjects: workers overwhelmingly set a reservation wage above the conditional offers made after high effort. Workers seem intent on receiving a wage that is high and does not depend on their effort level – with no evidence of convergence towards a mutually acceptable wage level.

4.4. Response of workers

Next we look for evidence that workers understand and expect punishment for non-compliance in the form of a reduced wage. Gift-exchange experiments are often understood as equilibria of a repeated game with punishment strategies: workers expect to receive lower wage offers if they shirk, and this induces them to choose high effort.

We did not collect information about expectations of punishment among our two subject pools, so we cannot test whether subjects who a priori expect to be punished comply more. But we can investigate whether experimental subjects who receive a lower wage offer after shirking shirk less later in the experiment. If they do, this would suggest a strategic understanding of the game consistent with a punishment equilibrium: workers start by putting employers to the test; if they are punished, they revise their expectations and improve their behavior; if they are not, they continue to shirk.

To investigate this possibility, we test whether subjects who have been punished for low effort in Game 2 (the first game with a free effort level) are less likely to shirk in Games 3 and 4. Conditional on shirking in the previous period, we define punishment as either being offered no contract or being offered a contract with a reduced wage. We limit our attention to treatments in which the worker can choose the effort level.

We first look at the frequency distribution of the number of shirking episodes in Games 2, 3 and 4.[25] We see that, as noted before, shirking is much more prevalent in Ghana. Furthermore, shirking falls across games in the UK: from 1.76 to 1.13 episodes between Game 2 to Game 4. But it remains essentially constant in Ghana: 2.77 in Game 2 and 2.71 in Game 4. These differences are strongly statistically significant. Similarly, the frequency of punishment is roughly constant in Ghana – 1.2 to 1.4 episodes per game – and falling in the UK – from 0.97 to 0.67 episodes from Game 2 to Game 4.[26] On average, Ghana subjects do not reduce shirking over time in spite of a reasonably high frequency of punishment.

To verify whether this holds at the individual level, we regress shirking in Games 3 and 4 on the number of times the subject was punished in Game 2, conditioning on the number of times the subject shirked in Game 2. We estimate the coefficients separately for Ghana and UK subjects. A UK subject dummy is included to control for differences in the unconditional propensity to shirk. Estimation results are presented in Table 7. We see that subjects who shirked in Game 2 are more likely to shirk again in Games 3 and 4. This is true in both countries but the effect is stronger for UK subjects, suggesting the presence of persistent shirkers. The effect is less pronounced in Ghana where subjects are less persistent in

[25] See Table A3 in the Online Appendix for details. Shirking is defined as exerting less effort than required by the contract or choosing low effort in response to a high wage level (i.e., above 23, which would equate the subjects payoffs if the worker chooses high effort). Similar results obtain if we define shirking as exerting less effort than required in the contract.

[26] In Game 2 the difference between countries is not statistically significant, but it becomes so in Games 3 and 4.

Table 7
Responsiveness of worker's compliance to past punishment.

(1)	Number of times	
Number of times the worker did not comply in Game 2:		
UK subjects	1.015**	($p = 0.031$)
Ghana subjects	0.397**	($p = 0.046$)
Number of time the worker was punished for non-compliance in Game 2:		
UK subjects	−0.169	($p = 0.781$)
Ghana subjects	0.674**	($p = 0.014$)
UK subject dummy	−2.760***	($p = 0.001$)
Intercept	3.959	(se = 0.451)
N. Observations	165	
Number of clusters	20	
Adjusted *R*-squared	0.445	

Note: The Table shows a linear probability regression on the pooled Ghana and UK sessions. Each observation corresponds to a different subject assigned the role of worker in Treatments 1E, 3E and 3ES. The dependent variable is the number of times a worker provided less than the requested effort level in Games 3 and 4 combined (from 0 to 10). The number of times the worker shirked in Game 2 is the number of times (from 0 to 5) that the worker provided less than the required level of effort in game 2 (the first game under treatment 1E). The number of times the worker was punished is the number of times (from 0 to 4) that the employer reduced the wage offer after the worker shirked. Reported *p*-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

their behavior – but more likely to shirk on average. UK shirkers tend to shirk less in Games 3 and 4 if they have been punished for shirking in Game 2, but the effect is not statistically significant. In contrast, we see that Ghana subjects who were punished in Game 2 are much more likely to shirk again than those who were not punished. In other words, punishing workers in the Ghana sessions does not have a deterrent effect on shirking later on.

These findings confirms that workers in the two subject pools differ in the way they conceive the strategic nature of the game: on average workers in the UK sessions comply more except for a subset of persistent shirkers; workers in the Ghana sessions respond negatively to punishment, i.e., they are even less likely to comply with the requested effort if punished in the past. Based on this, it is easier to understand why employers in the Ghana sessions do not punish, since doing so fails to discipline workers and, in fact, makes matters worse.

4.5. Consequences on effort and earnings

The absence of a punishment strategy in Ghana has consequences on worker effort, payoffs, and efficiency. Looking at the evolution of effort across periods, in the UK we find few transitions away from high effort: of the workers who chose high effort in one period, 74% choose it again in the following period. The corresponding figure for Ghana is 49%, implying that Ghana subjects assigned the role of employer are less able than UK subjects to maintain high effort provision by their workers. This is possibly related to the fact that Ghana employers often reduce their wage offer after high effort – perhaps seeking to extract surplus from a worker they assume to be intrinsically motivated. We also observe, as in other experiments, a reduction in effort in the last period of the game. This fall is strongest among UK subjects and it brings the UK share of low effort up to Ghana levels, i.e., 44% versus 47%. This is still well below the 100% low effort that would be predicted by the theory of finitely repeated games.

Differences in average effort levels between the two subject pools have dramatic consequences for the earnings of employers and workers in our experiment. As was shown in Table 1, under treatment 1C (perfect enforcement) the earnings of workers and employers are nearly equal, especially in later games. Imperfect enforcement (treatment 1E) significantly reduces employers' average earnings: between treatment 1C in game 1 and treatment 1E in game 2, employer earnings fall from 12 to 0.6 points in Ghana and from 17 to 6 points in the UK. These differences between games 1 and 2 are highly significant.[27] The difference in employer earnings between UK and Ghana subjects in treatment 1E is also significant. This difference even increases in later games: in Ghana, employer's earnings fall further in games 3 and 4 (to −1.0 and −0.6 points, respectively), while they increase slightly in the UK (to 6.3 and 9.7 points, respectively). These differences between UK and Ghana subjects are all statistically significant.

In contrast, we do not observe significant differences in worker's earnings, both between treatments 1E and 1C, and between UK and Ghana subjects. Workers' earnings are slightly lower in 1E than in 1C in both subject pools, but this difference is not significant. In other words, it is Ghana employers who are "paying the price" of lower effort.

[27] See Table A4 to A6 in the Online Appendix for details.

4.6. Interpretation

Combining the results on the behavior of employers and workers, we conclude that subjects in the two countries follow different equilibria of the repeated game. Subjects in the UK operate, by and large, along a trigger strategy equilibrium – albeit with some noise. Subjects in Ghana, in contrast, operate in an equilibrium in which (1) employers cannot hope to increase effort by reducing wages since workers do not increase effort after a reduction in wage upon non-compliance; and (2) workers cannot hope to increase their wage by offering high effort since employers do not increase or maintain high wages after high effort. This equilibrium is not, however, the Nash equilibrium of the static PD game: in that equilibrium, employers would never offer a wage above 4,[28] which is much lower than the average wage of 14 offered by Ghana subjects (see Table 1, treatment 1E). The net effect is a large loss of efficiency due to the low effort chosen by workers.

In Section 3 we discussed other possible behavioral strategies that employers and workers may play. If employers believe that all workers are intrinsically motivated and always choose the same level of effort (i.e., high, medium, or low), they should offer the same low wage as in the static PD game. As just discussed, this prediction is rejected by the data. But we do find evidence that a large fraction of Ghana employers reduce their wage offer after high effort – a choice that can only be interpreted as an effort to capture a larger share of the surplus from high or medium effort workers believed to be non-responsive to wages. This behavioral strategy, however, fails to explain why high wages are offered in the first place.

Another strategy we discussed in Section 3 is conditional reciprocity by workers, who reciprocate by conditioning their effort level to the wage offered. A pure conditional reciprocity model is unable to account for our findings, since it predict much higher average effort by Ghana subjects, contrary to our findings. Can a hybrid model do better at accounting for the behavior of Ghana subjects?

In Section 3, we noted that if employers believe that some workers always choose low effort irrespective of the wage, but a large enough fraction of them reciprocate, it is optimal to start by offering a high wage to identify conditional reciprocators. If Ghana subjects have a lower proportion of conditional reciprocators than UK subjects, this may explain the low effort levels observed in Ghana.

To investigate this possibility, we fit to the effort choices of workers a structural model of distributional preferences that captures the extent to which they reciprocate to a high wage with high effort. The details of the procedure are presented in the Online Appendix. In both study populations, we find that a sizable proportion of workers act in a way that reduces the earnings gap with the employer, i.e., they reciprocate high wage offers with high effort. This proportion is slightly larger among UK subjects, but not sufficiently large to explain the difference in effort level observed in the data. Furthermore, at least part of this difference is because UK workers respond more to wage incentives – not because they have altruistic preferences that induce them to reciprocate.

We also find that employers respond differently to worker heterogeneity in reciprocity. UK employers act more strategically in setting wages than their Ghana counterparts: if a worker is shown to reciprocate less, employers reduce their wage offers. In contrast, Ghana employers matched with a less reciprocating worker on average keep offering higher wages – and thus making losses – until the end of the game. At the same time, as noted above, many employers in the Ghana sessions reduce the wage after high effort, thereby undermining cooperation from reciprocating workers. These features explain much of difference in average employer payoff between the two countries.

5. Competition and reputation

So far we have only considered 1-on-1 games. Can the situation be improved by introducing competition among workers and employers? Competition may increase punishment if rejected workers are replaced and remain unemployed – the so-called disciplining effect of unemployment (e.g., Shapiro and Stiglitz, 1984). If true, this could increase compliance among Ghana subjects. However, competition for workers among employers may also interfere with the employer ability to punish non-compliance: it could bid up the wage above what would constitute a sufficient punishment; and it could allow non-compliant workers to find alternative employment, hence avoiding punishment. Which of the two possible effects dominates is an empirical question, already partly explored by Brown et al. (2004; 2012) in the context of unbalanced markets.

To investigate the effect of competition in our setting, we compare perfect and imperfect enforcement treatments 1C and 1E to their 3-to-3 counterparts 3C and 3E. We start by comparing wages across 1-on-1 and 3-on-3 treatments. Since employers make the first move, competition for workers between employers may increase wage offers. We indeed find some evidence in Ghana that offers are higher in 3C compared to 1C. This is in particular beneficial to workers who capture a larger share of the surplus (e.g., Davies and Fafchamps, 2016).

Turning to imperfect enforcement treatments 1E and 3E, we find no evidence that wage offers differ between them. For example, in game 3 when treatment 3E is first introduced, the average wage offer is 14.3 points in Ghana and 12.9 in the UK, while in game 3 of treatment 1E the average is 12.6 in Ghana and 13.6 in the UK. These small differences between

[28] Offering a wage of 1 gives the low effort worker a payoff of 1 – just enough to accept the offer. Giving 5 or more to a low effort worker gives the employer a payoff ≤ 0, and is thus not rational.

Table 8
Wages and past interactions (Treatment 3E only).

Dependent variable:	(1)	(2)		(3)		(4)		(5)	
Wage offered in period:	Period 1	Period 2		Period 3		Period 4		Period 5	
1 past interaction		−0.26	(0.635)	−0.13	(0.853)	0.59	(0.364)	−1.36	(0.084)
2 past interactions				−0.40	(0.604)	1.28	(0.189)	−0.39	(0.690)
3 past interactions						−1.32	(0.576)	−1.15	(0.211)
4 past interactions								−1.19	(0.406)
UK dummy interacted with:									
1 past interaction		−0.93	(0.398)	−1.84	(0.283)	−1.42	(0.500)	−0.37	(0.817)
2 past interactions				3.13**	(0.031)	−1.31	(0.606)	−0.51	(0.757)
3 past interactions						10.26**	(0.026)	6.00***	(0.001)
4 past interactions								7.64**	(0.043)
Wage without past interactions (Ghana)	15.18	14.38		13.92		12.47		13.40	
Wage without past interactions (UK)	14.23	13.69		13.43		11.02		11.11	
Employer fixed effects	Yes	Yes		Yes		Yes		Yes	
N. Observations	685	593		586		554		561	
Number of clusters	23	23		23		23		23	
Adjusted *R*-squared	0.753	0.656		0.649		0.634		0.697	

Note: The Table reports the coefficient estimates of an employer fixed-effect regression in which the dependent variable is the wage offered in a particular period – going from period 1 to period 5. The regressors represent the number of times the worker accepted an offer from the employer in earlier periods of the game. Observations from the Ghana and UK sessions are pooled, but interaction terms with a UK dummy are included. Since there is a separate intercept for each employer, we report in the Table the average intercept for the Ghana and UK subjects; it is the average wage offer when there are no past interactions between the employer and worker. Reported *p*-values, in parentheses, are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

treatments are not statistically significant.[29] We do, however, find a significant difference in offer levels between the UK and Ghana in treatment 3E: average wage offers are lower among UK subjects and non-parametric tests reject the null hypothesis of equal offer distribution in Ghana and the UK.[30]

Next we examine the data for evidence that, in treatment 3E, employers offer higher wages to individual workers who have provided high effort to them in the past: since employers are competing over workers, they may seek to hold on to those who provided them with high effort in the past. We first look for evidence of higher wages in repeated interactions. Table 8 shows how wage offers vary with the number of past interactions with a worker, regardless of past effort level. In the UK, wage offers are increasing in the number of past interactions, suggesting that repeated contracting is associated with more beneficial exchange for employers. For example, in period 5, the wage offer is 4.85 points higher when an employer has interacted with a worker for three periods. This is in line with previous findings.[31] In contrast, among Ghana subjects we find no significant correlation between wage offers and the number of past interactions.

A similar contrast is obtained if we control for whether the worker complied with the requested effort level in the past. Columns (1) and (3) of Table 9 present fixed effects regressions of the offered wage on whether the employer contracted with this worker in the past period and whether the worker chose the requested effort level or higher. We find a significant coefficient for past compliance among UK subjects, but not in Ghana. In the UK, past compliance is associated with a 7.5 points increase in wage offer in the subsequent period. The corresponding coefficient for Ghana is small (1.0 point) and not significant at a 10% level. These findings confirm that our UK subjects behave in a way similar as other Western subjects. But Ghana subjects behave differently: they do not reward workers for good behavior and they do not seek to retain them by increasing their wage over time.

These findings suggest that the two populations differ in the kind of heuristics they bring to contractual situations: punishing and rewarding workers for low and high effort is not something that comes naturally to college students in Ghana, while it does for students in the UK and, from the literature, in other developed economies. Why is this the case?

One possibility is that Ghana subjects are handicapped by the lack of a reputation mechanism to discipline workers. Indeed, for non-compliant workers to be punished by *not* being offered alternative employment, it is necessary that other employers observe why a worker was not rehired by a previous employer. Making that information available may help Ghana subjects coordinate on punishing non-compliant workers by refusing to hire them.

[29] This is true for within-subject, between-subject or difference-in-difference comparisons – see Online Appendix Tables A8, A9 and A10 for details. There is also no noticeable difference in the dispersion of wage offers in either of the two study populations, and non-parametric tests similarly find no significant difference in the offer distribution between the two treatments.

[30] The Kolmogorov-Smirnov test rejects the null for both game 3 ($p = 0.093$) and 4 ($p = 0.000$). The rank-sum test similarly rejects the null for game 4 ($Z = 2.90$, $p = 0.004$), but not for game 3 ($Z = -0.986$, $p = 0.324$).

[31] In the experiment of Brown et al. (2004), employers could make public offers to all workers as well as private offers to individual workers. They found that private offers were on average higher than public offers and that employers tailored their wage offers to past effort levels. In our design all offers are to a specific worker, but employers can make different offers to different workers. We find similar behavior in spite of the difference in design.

Table 9
Wage offers (Treatments 3E and 3ES only).

Dependent variable:	(1)		(2)	
Wage offered by employer i in period t	Treatment 3E only		Treatment 3E and 3ES	
Contracted with employer at $t-1$	−0.596	($p=0.406$)	−0.393	($p=0.605$)
Contracted with another employer at $t-1$	−0.640	($p=0.338$)	−0.444	($p=0.539$)
Complied with employer's requested effort at $t-1$	0.474	($p=0.431$)	0.367	($p=0.604$)
Complied with another employer's requested effort at $t-1$	0.636	($p=0.172$)	0.493	($p=0.300$)
Treatment 3ES dummy			0.766	($p=0.627$)
Treatment 3ES dummy interacted with:				
Contracted with employer at $t-1$			0.0140	($p=0.984$)
Contracted with another employer at $t-1$			−0.179	($p=0.857$)
Complied with employer's requested effort at $t-1$			0.117	($p=0.907$)
Complied with another employer's requested effort at $t-1$			−0.140	($p=0.908$)
UK dummy interacted with:				
Contracted with employer at $t-1$	−2.474**	($p=0.035$)	−2.899**	($p=0.030$)
Contracted with another employer at $t-1$	1.165	($p=0.282$)	0.962	($p=0.409$)
Complied with employer's requested effort at $t-1$	7.038***	($p=0.000$)	7.483***	($p=0.001$)
Complied with another employer's requested effort at $t-1$	−0.656	($p=0.465$)	−0.390	($p=0.636$)
Treatment 3ES dummy			1.318	($p=0.549$)
UK dummy × Treatment 3ES dummy interacted with:				
Contracted with employer at $t-1$			1.604	($p=0.261$)
Contracted with another employer at $t-1$			−0.746	($p=0.604$)
Complied with employer's requested effort at $t-1$			−0.089	($p=0.952$)
Complied with another employer's requested effort at $t-1$			4.200***	($p=0.005$)
Employer fixed effects	Yes		Yes	
Period dummies × UK dummy	Yes		Yes	
Intercept	14.03	(0.419)	13.75	(0.460)
N. Observations	2294		3179	
Number of clusters	23		23	
Adjusted R-squared	0.633		0.642	

Note: The Table presents coefficient estimates of a linear regression using as dependent variable the wage offered by the responding employer in period t. All regressors are indicator variables and relate to period $t-1$, i.e., the period before the offer. The regressor 'Contracted with employer at $t-1$' is a dummy variable equal to 1 if the worker accepted the offer of the employer in the previous period. Observations from the Ghana and UK sessions are pooled, but interaction terms with a UK dummy are included. Column (2) pools Treatment 3E with Treatment 3ES to test the effect of reputation. Reported p-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Testing this hypothesis is the object of treatment 3ES which, within each game, introduces public information about the past effort choices of each worker, irrespective of which employer hired them. The existing literature has found that introducing reputation in this way does act as an additional (albeit weak) incentive for workers to provide high effort in subjects populations from developed countries. Is this also true in our Ghana sample?

We first compare average wage offers between treatments 3E and 3ES. We find some evidence that reputation increases wages in the UK, as in earlier experiments. No such effect is observed in Ghana, however. This is shown in Table 10 that compares 3ES to 3E for a number of outcome variables, using within subject and difference-in-difference approaches.[32] For Ghana, the within-subject estimator shows a significant drop in wage as a result of information sharing, but in the diff-in-diff regression where we control for a time trend, the effect disappears. In contrast, for the UK the within-subject, diff-in-diff and fixed effects results all show a positive and significant effect of information sharing on wage offers.

We also find that UK subjects modulate their wage offer on the reputational information provided on past effort. We find no such evidence for Ghana subjects – thereby rejecting the hypothesis that Ghana subjects needed a reputation mechanism in order to deter non-compliance. This is shown in columns 2 and 4 of Table 9: compliance with another's employer's requested effort is associated with a significant wage increase of 4 points in the UK, while in Ghana the corresponding coefficient is less than half a point and is not significant. Ghana subjects do not appear to rely on a reputation mechanism to incentivize workers. We also find that UK subjects reward workers more for complying with their own requested effort than that of another employer: compliance to oneself is associated with a wage increase of 7.9 points, almost double that for compliance with another employer's requested effort. This could be because they value compliance more when it benefits them – and wish to reward it more. Another possibility is that employers have more detailed information for own workers – they know the wage they paid, not the wage paid by other employers. Consequently, they are in a better position to assess who was 'at fault' for non-compliance: a worker may not be expected to provide high effort when the wage is unreasonably low. None of these patterns is present among subjects in the Ghana sessions.

[32] See also Online Appendix Tables A4-A6 for the difference-in-difference and fixed effects comparisons.

Table 10
The effect of the reputation treatment 3ES on the main experimental outcomes (within-subject analysis).

	Ghana sessions		UK sessions	
	(1)		(2)	
Dependent variable is:	3ES vs. 3E		3ES vs. 3E	
Offered wage	−2.067***	($p = 0.000$)	1.678***	($p = 0.006$)
Acceptance of contract offer	−0.009	($p = 0.551$)	0.019	($p = 0.491$)
Compliance with requested effort	0.038	($p = 0.237$)	0.114**	($p = 0.020$)
Joint surplus	0.189	($p = 0.904$)	3.718***	($p = 0.001$)
Employer's earnings	1.870	($p = 0.269$)	2.723***	($p = 0.002$)
Worker's earnings	−1.663**	($p = 0.044$)	1.280**	($p = 0.039$)

Note: Each cell of the Table corresponds to a separate regression. The reported coefficient is the treatment effect of 3ES relative to 3E. Each row corresponds to a different dependent variable, defined as in earlier Tables. The regressions are subject-fixed-effect linear regressions using only those subjects included in both treatments 3E and 3ES; it compares outcomes within subjects across games played at different times in the same session. Each regression includes a dummy equal to 1 if the subject is assigned to treatment 3ES and 0 otherwise, plus game order dummies. Observations from games other than 3E and 3ES are omitted throughout. Reported p-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11
Replication analysis with Ghana entrepreneurs in addition to students: Wage offer (Treatment 1E – Ghana only).

Dependent variable is:	Wage offer in period t	
Worker provided high effort in period $t-1$	1.729***	($p = 0.000$)
Worker rejected the offer in period $t-1$	1.702***	($p = 0.000$)
Entrepreneur dummy interacted with:		
Worker provided high effort in period $t-1$	1.795	($p = 0.107$)
Worker rejected the offer in period $t-1$	−0.896	($p = 0.343$)
Employer fixed effects	Yes	
Period dummies	Yes	
Intercept	15.20	(se = 0.263)
N. Observations	2785	
N. Clusters	23	
Adjusted R-squared	0.608	

Note: The Table presents coefficient estimates from a OLS regression with employer fixed effects, using observation from a replication experiment conducted several months later with university students and entrepreneurs in Ghana. The dependent variable is the wage offered by the subject assigned the role of employer in the experiment. The Entrepreneur dummy is equal to 1 if the subject is an entrepreneur, and 0 if the subject is a college student. By construction, only wage offers in period 2–5 are included. Reported p-values are based on errors clustered at the session level using a wild cluster bootstrap – which by construction does not report standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

So far we have examined whether employers reward workers for compliance with other employers' effort requests. But do reputational incentives induce workers to comply more – i.e., is reputation effective in deterring shirking? We do not find strong evidence that this is the case. From Table 1 we see that the rate of effort compliance in Ghana is higher (50%) in treatment 3ES than in treatment 3E (43%). This arises even though employers do *not* modulate wage offers based on past compliance with other employers. As shown in Table 10, however, this difference is not statistically significant.

In contrast, in the UK where employers do vary wage offers to reflect compliance with others, the compliance rate in treatment 3ES in game 4 is identical to treatment 3E in game 4 – 76% in both cases.[33] Similar findings are obtained if we control for the wage offered.[34] Taken together, these results suggest that multilateral reputation is not a stronger disciplining mechanism than bilateral reputation/relational contracting in the UK sessions, even though it leads to higher wage offers. Why do employers offer more if the non-compliance risk has not changed? One possibility is that circulating information on workers allows employers to draw inference about their type (e.g., selfish-rational or conditional cooperators), thereby

[33] This is a between-subject comparison. The within-subject treatment effect estimate for compliance is positive: compliance of workers in treatment 3ES in game 4 is 11 percentage points higher than the compliance rate of these workers in treatment 3E in game 3.

[34] Table A7 in the Online Appendix presents a linear probability model of acceptance and compliance as a function of wage offer and being in treatment 3ES. As before we find that the wage coefficient is positive and significant in both Ghana and the UK. But the coefficient on treatment 3ES is mostly non-significant.

Table 12
Predicting behavior across country samples.

	Punishment by employer	Worker effort	Compliance with contract	Placebo test
Actual difference in behavior between subjects in the Ghana and UK sessions	−0.688	−0.723	−0.789	0.112
p-value	*0.000*	*0.000*	*0.000*	*0.441*
Predicted difference using a pooled regression: random forest	−0.492	−0.600	−0.628	0.101
p-value	*0.000*	*0.000*	*0.000*	*0.258*
Predicted difference using a pooled regression: OLS	−0.220	−0.375	−0.370	0.081
p-value	*0.000*	*0.000*	*0.000*	*0.026*
Predicted difference based on out-of-sample predictions: random forest	1.092	0.631	0.741	0.060
p-value	*0.000*	*0.000*	*0.000*	*0.269*
Predicted difference based on out-of-sample predictions: OLS	1.007	0.613	0.559	0.126
p-value	*0.000*	*0.000*	*0.000*	*0.470*

Notes: The first two rows report a *t*-test of the actual difference in the mean value of the behavior between subjects in the Ghana and UK sessions. A negative value means that Ghana subjects are less likely to engage in the listed behavior, i.e., less likely to punish shirking workers, less likely to supply high effort, and less likely to comply with the high effort requested by the employer. The fourth column does the same for a randomly generated variable that serves as placebo test. Next we test whether coefficients from a pooled regression can predict behavioral differences between the two samples. Since there are differences across samples in the average values of the regressors, we expect in-sample predictions to mechanically capture differences in behavior. The results show that they do, whether predictions are based on a random forest algorithm or on OLS. We then compare out-of-sample predictions for both models. In this comparison, a model is fitted to the UK subjects and used to predict the behavior of Ghana subjects, while another model is fitted to Ghana subjects and used to predict the behavior of UK subjects. We compare the out-of-sample predictions to each other. If differences in observed characteristics predict differences in behavior, the sign and significance of the difference in predictions will be similar for in-sample and out-of-sample predictions. If the sign is reversed, it means that the ability to predict is strongly rejected. A positive value means that, based on their observed characteristics, Ghana subjects are predicted to be more likely to engage in the behavior than UK subjects.

reducing the risk of non-compliance perceived by employers. This, in turn, increases the employers' expected gain, which they share with workers by offering higher wages.

6. Conclusion

Experiments with repeated games of gift exchange have provided evidence of cooperating behavior consistent with trigger strategy equilibria. Our results in the UK sessions support this as well. We do, however, find markedly different behavior among college students in Ghana where subjects assigned the role of worker tend to choose a low level of effort even after receiving a high wage offer. This contrasts with UK subjects, who are more likely to choose high effort if offered a high wage. A similar contrast is found in the behavior of subjects assigned the role of employer. In the UK sessions, employers tend to offer a high wage following high effort, and to reduce their wage offer otherwise. Ghana employers, on the other hand do not penalize low effort or reward workers for good behavior. In particular, they often persist in making high wage offers in spite of repeated low effort by the worker. Similar results, shown in Table 11, were found with Ghanaian entrepreneurs (in addition to students) in a companion study by Davies and Fafchamps (2017). These behaviors among Ghana subjects are not consistent with trigger strategy equilibria, and many actions taken by workers (e.g., low effort when paid a high wage) and employers (e.g., reducing the wage after high effort) can similarly not be reconciled with conditional reciprocity. As a result, efficiency is lower in the Ghana sessions and employers earn much less than workers.

We show that competition among workers and among employers does not induce more punishment and higher effort in Ghana and hence does not eliminate the difference between subject pools. Another possibility is to introduce non-monetary incentives – i.e., by allowing employers to praise or criticize workers based on their effort choice. This avenue is explored in Davies and Fafchamps (2017), but it does not lead to a significant increase in effort either. Finally, we find that employer subjects in the UK offer higher wages to workers with a good reputation, in line with the result of Charness and Kuhn (2011). But no such effect is observed in the Ghana sessions. We also demonstrate, in the robustness discussion in Appendix Section B, that the difference in behavior between the two subject pools cannot be explained by differences in subject characteristics.

What could account for our findings? One possibility – that is often raised when using subjects from developing countries – is that subjects do not understand the rules of the game. In anticipation of this, we took great care in designing the player interface in an intuitive manner. The simple wording of the screens is well within the language comprehension of Ghana and UK college students who are all taught in English. We also simplified the design to reduce possible sources of confusion. In particular, we do not require subjects to make any numerical calculation: all payoffs are instantaneously displayed on their screen for any effort choice. Furthermore, at the rehiring stage we ask employers to make conditional offers based on past effort. This encourages employers to think strategically about wage offers. Finally, subjects play three practice rounds before the experiment starts. Notwithstanding all these precautions, we looked for evidence that lack of understanding caused our results. Each subject plays the stage game twenty times. If misunderstanding was the problem, we should find evidence of learning over the course of the experiment. But we find no evidence in the analysis that Ghana subjects improve the way

they play over the course of a session – in contrast to UK subjects, whose compliance with requested effort levels increase between games 2 and 4.[35]

Based all this evidence and the fact that the game is quite easy to play, we have no doubt that Ghana and UK subjects understood the *rules* of a game. The evidence nonetheless indicate that they had a different understanding of its *strategic* implications – i.e., they apply different heuristics. Support for this interpretation is found in the comments that subjects were encouraged to volunteer at the end of the session. The comments made by Ghana subjects assigned the role of employer in interviews after the experiment confirm they clearly understood that, by providing low effort in response to high wages, workers were causing losses for them. But some of the employers indicated that this was caused by the experimental conditions, without acknowledging that they could have done something about it themselves. This further confirms that they understood the rules of the game, but perceived its strategic implications differently. They are also in line with a field experiment with Ghanaian data entry workers conducted by Bandiera and Fischer (2015), who show that introducing performance-based pay did not increase effort.[36] Our study complements these findings by showing that punishments and rewards are not used by Ghana subjects put in the role of employer.

Having ruled out trigger strategies and pure conditional reciprocity, what can account for the mix of behaviors we observe in the Ghana sessions? One way of rationalizing what we observe is to assume that employers believe they face a multiplicity of worker types (e.g., Davies and Fafchamps, 2021): selfish-rational types – who cannot be incentivized; conditional reciprocators – who provide high effort if paid well; and intrinsically motivated workers – who provide effort irrespective of wage. Nearly half of Ghana employers reduce the wage offer after high effort, a behavior consistent with a discovery strategy aimed at identifying workers who provide high effort even for low pay. Similarly, the fact that Ghana employers repeatedly offer a high wage to workers without ever punishing them suggests they believe that most workers are conditional reciprocators. What remains unexplained is that they *continue* to offer high wage even after non-compliance. One conjecture is that employers believe workers will *eventually* reciprocate once they have received enough. This hope is not fully realized in the experiment – on average, Ghana subjects assigned the role of employer collect little or no payoff from their participation to the games. But it suggests a strong prior belief in conditional reciprocity. More research is needed on this topic.

Declaration of Competing Interest

1. Funding for this study was provided by the UK Department for International Development (DFID) as part of iiG, a research programme to study how to improve institutions for pro-poor growth in Africa and South-Asia. This is acknowledged in the paper itself.

2. Over the last three years, Marcel Fafchamps has received research grants from: the Department for International Development (DfID-UK), including from IZADfID and PEDL-DfID; the International Growth Centre (IGC); the Economic and Social Research Council (ESRC-UK); Stanford University; and the World Bank. Marcel Fafchamps has worked as consultant for the World Bank and for Oxford University. Elwyn Davies acknowledges funding from the Economic and Social Research Council (ESRC), grant number ES/J500112/1, provided through the University of Oxford Social Sciences Doctoral Training Centre (DTC). Prior to his employment at the World Bank (from September 2017 onward), Elwyn worked as a consultant for the World Bank's Trade & Competitiveness Global Practice.

3. Elwyn Davies does not hold any concurrent positions besides his employment at the World Bank. Marcel Fafchamps is Editor of Economic Development and Cultural Change (University of Chicago Press).

4. No close relative of Marcel Fafchamps or Elwyn Davies works as paid or unpaid member of a relevant non-profit or profit-making organization or entity.

5. No other party had the right to review the article prior to its circulation.

6. IRB approval has been sought and obtained from the University of Oxford Central University Research Ethics Committee (CUREC) for both the Ghana and the UK experiments (reference numbers are respectively Econ DREC1213/0016 and ECONCIA14-047).

Appendix A. Detailed order of play in the 3-on-3 game

In the 3-on-3 treatment, the sequence of moves is similar to the 1-on-1 treatment, except that it allows for multiple actions by employers and workers. The detailed order of play is as follows:

- **Contracting:** At the beginning of period $t = 1$, employers and workers contract with each other in a virtual marketplace. Each of the workers is listed with their identification number clearly visible. This stage consists of three steps:
  - First, each employer j makes offers to each individual worker i. An offer by employer j to worker i specifies the payment that the employer will make to the worker w_{ijt} and the effort level $\widetilde{e}_{ijt}$ desired from the worker. The employer can also decide not to make an offer to a particular worker. Employers make these choices without seeing the

[35] The response time limits were identical in all the sessions, making the length of the sessions comparable.

[36] In this field experiment, participants were recruited for a data entry position and randomly assigned a contract with either a flat pay, an individual piece rate, or a group piece rate based on the number of keystrokes. Unlike in our study, interactions were not repeated and incentives were set exogenously through randomization, not allowing relational contracting with employers.

choices made by other employers. At this stage, choices are private, i.e., they are not yet seen by workers and other employers.
  - Second, when all employers have finished selecting offers to all three workers, the selected offers are revealed to all three employers. Having seen the offers of the other employers, each employer then has one chance to revise his offers to each of the three workers. All these initial offers are not shown to the workers.
  - Third, when all employers have finished revising their initial offers, workers are allowed to see the three offers made to them. This is done in a randomly determined sequential order. One of the three workers is selected at random; that worker sees all the offers made to them; the worker either accepts one of them or none. It is then the next worker's turn, and so on. If a worker rejects all offers or no offer was made, the worker receives a zero payoff for that period. Once an offer by employer j is accepted by a worker i, no subsequent worker can accept an offer from employer j. This ensures that each worker has at most one employer and that each employer has at most one worker.
- **Effort choice:** If a worker has accepted an offer, they then decide an effort level e_t. The rest is as in the one-on-one treatment.
- **Rehiring.** Before moving to the next period, we ask each employer i matched with a worker i to choose a contract offer for next period. As in the one-on-one treatment, this contract $\{w_{ijt+1}(e_{it}), \widetilde{e}_{ijt+1}(e_{it})\}$ is contingent on the effort level of worker i. We also ask worker i to specify a reservation wage r_{ijt+1} for employer j.
  The game then moves to the contracting stage of period $t+1$. If employer j was matched with worker i at period t, the contingent offer $\{w_{ijt+1}(e_{it}), \widetilde{e}_{ijt+1}(e_{it})\}$ is automatically made to that worker. If worker i also stipulates a reservation wage r_{ijt+1} below $w_{ijt+1}(e_{it})$, the offer is deemed accepted and the employer-worker pair is removed from set of subjects yet to be matched. The purpose of this construct is to allow employer and worker to form a long-term bond, free of the vagaries of the randomized order in which workers accept employer offers. If the worker's stipulated reservation wage is higher than the offered wage, the offer is deemed rejected. All unmatched employers then make offers to the unmatched workers, as described in the contracting stage above.
- **Rematching**: The above three steps are repeated in sequence until $t = 5$, at which point the game ends. Workers and employers are then rematched for a new game, possibly with a different treatment. In total, each subject plays four different games of five periods. Their precise sequence is discussed more in detail below.

Appendix B. Robustness analysis

Since the Ghana findings differ from previous evidence from developed countries, we check their robustness to other specifications and we subject them to out-of-sample validation.

We start by estimating alternative specifications of the Ghana regression of wage offer on past compliance.[37] In the first specification we add the lagged value of the wage. This controls for the possibility that low effort in the previous period was seen by employers as a response to low wage. Adding this control does not change our main finding. Next we estimate a model in which the dependent variable is the change in wage relative to the last period – rather than the wage level. This again does not change the magnitude or significance of the coefficient of past compliance among Ghana subjects. We also regress the offered wage on past effort and whether the chosen effort was a positive or negative "surprise" (i.e., higher or lower than demanded).[38] We find that both Ghana and UK employers reward high effort with a high wage offer. But the coefficient is significantly higher for UK subjects. The response to negative surprises also differs. In the UK, a negative surprise lowers the offered wage of −2.1 points; in Ghana it raises the offered wage by 1.7 points – a difference that is statistically significant.

Next we verify whether our Ghana results replicate to other sessions with similar or different subjects. To this effect we compare our findings to those of an experiment run in Ghana by Davies and Fafchamps (2017) and designed to test whether allowing employers to send messages of praise or criticism to workers alleviates the incentive issues outlined here. The basic design of the experiment is a simplified version of ours: it reduces effort levels from three to two and makes high effort the default contractual setting.[39] If anything, these changes should make rewards and punishment even more salient. In total 31 sessions were held with a total of 559 students from the same Ghana universities, to which are added 61 entrepreneurs recruited from small and medium-size enterprises.

In Table 11 we use data from those sessions to replicate our regression of wage offer on past compliance. Results are slightly more in line with the literature: past compliance is now associated with a statistically significant 2 point increase in wage offer. The magnitude of the coefficient, however, is very small relative to the employer's loss from low effort, which ranges between 25 and 30 points. For entrepreneur subjects, the estimated coefficient is 3.5, suggesting a stronger willingness to condition wage offers on past compliance. But the pooled regression in column (3) suggests that the difference is not statistically significant. More importantly, estimates from the replicated Ghana experiment remain well below what we find for UK subjects: in Table 4, an equivalent regression yields a wage increase of 10.2 points for compliance with high effort among UK subjects.

[37] Detailed results are given in Table A8 in the Online Appendix.

[38] See Table A9 in the Online Appendix.

[39] The payoff function is the same as ours, except that in some sessions the employer's earning from low effort is raised from 5 to 10 or 15 points.

As a final robustness check, we examine whether differences in shirking behavior between the two subject pools could have been predicted from differences in observable characteristics such as gender, age, having a parent entrepreneur, and answers to the Big 5 questions.[40] To answer this question, we conduct the following thought experiment: suppose we modify the composition of the UK sample to match the subject mix in the Ghana sample. Will we observe the same behavior as the Ghana sample? If we do, this means that behavioral differences between the two subject pools can be attributed to differences in sample composition. If we do not, this means that there exist a systematic difference in behavior between the subject pools that cannot be accounted for by average differences in observed individual characteristics across the two samples. In that case, we cannot say for sure what causes this difference in average behavior, but differences in behavioral expectations is one possibility that we cannot rule out a priori.

We estimate a random forest machine-learning algorithm on each of the subject pools separately and we test its predictive performance on the other subject pool – and vice versa. For comparison purposes, we do the same using OLS. We then tests whether the predicted difference in average behavior between the Ghana and UK subjects is statistically significant and has the same sign as the difference observed in the data itself.

For this approach to be convincing, we need to take care of two possible concerns. First, if survey characteristics are unable to predict behavioral differences between countries when we pool data from the two samples, then we should not be surprised that they do not predict them when we allow regression coefficients to vary across the two countries. To take care of this concern, we first test whether the behavior predicted by the pooled regressions is significantly different between the two subject pools. Second, tests based on regression predictions are affected by prediction error and overfitting bias, whether in sample or out of sample. To correct for this, we use randomization inference to simulate the distribution of each test under the null that regressors are independent of the dependent variable and thus have no true predictive power.[41]

Results are shown in Table 12 for three behavioral indicators: whether an employer punishes a shirking worker; whether a worker exert high effort; and whether a worker complies with the high effort requirement of the contract. We also report results for a randomly assigned behavior that serves as placebo. The first row of the Table gives the actual difference in average behavior between the two study samples. We also provide the p-value of a simple t-test of equality of means. The negative differences reconfirm that Ghana subjects are less likely to punish shirking, supply high effort, and comply with contract. Next we test whether coefficients from a pooled regression can predict behavioral differences between the two samples. To the extent that there are differences in regressor means across the two populations, in-sample predictions of the pooled regression will mechanically capture behavioral differences. The results confirms that indeed they do, whether predictions are based on a random forest algorithm or on OLS.

The test of interest is when we compare *out-of-sample* predictions for both models. In this comparison, a model is fitted to the UK data and used to predict the behavior of Ghana subjects, while another model is fitted to Ghana subjects and used to predict the behavior of UK subjects. We compare the out-of-sample predictions to each other. If differences in observed characteristics predict differences in behavior, the sign and significance of the difference in predictions will be similar for in-sample and out-of-sample predictions. If the sign is reversed, it means that the ability to predict is strongly rejected. The reported positive differences for the three listed behavior mean that, based on their observed characteristics, Ghana subjects are predicted to be *more* likely to engage in these behaviors than UK subjects. In contrast, the placebo test show no significant differences in out-of-sample predictions, confirming the statistical validity of the procedure. These results imply that differences in observables between the two subject pools are unable to account for the behavioral variation in our study.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.09.024.

References

Abeler, J., Altmann, S., Kube, S., Wibral, M., 2010. Gift-exchange and worker's fairness concerns: when equality is unfair. J. Eur. Econ. Assoc. 8, 1299–1324.
Altmann, S., Falk, A., Grunewald, A., Huffman, D., 2014. Contractual incompleteness,unemployment, and labour market segmentation. Rev. Econ. Stud. 81, 30–56.
Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. Econ. J. 103, 570–585.
Bandiera, O., Barankay, I., Rasul, I., 2005. Social preferences and the response to incentives: evidence from personnel data. Q. J. Econ. 120, 917–962.
Bandiera, O., Barankay, I., Rasul, I., 2009. Social connections and incentives in the workplace: evidence from personnel data. Econometrica 77, 1047–1094.
Bandiera, O., & Fischer, G. (2015). Can "Good" HR Practices be Exported? Evidence from a Field Experiment in Ghana. IGC Working Paper, No. F-6021-GHA-1.
Bloch, F., Genicot, G., Ray, D., 2008. Informal insurance in social networks. J. Econ. Theory 143 (1), 36–58.
Brandts, J., Charness, G., 2004. Do labour market conditions affect gift exchange? Some experimental evidence. Econ. J. 114, 684–708.
Breza, E., Kaur, S., Shamdasani, Y., 2018. The morale effects of pay inequality. Q. J. Econ. 133 (2), 611–663.
Brown, M., Falk, A., Fehr, E., 2004. Relational contracts and the nature of market interactions. Econometrica 72, 747–780.

[40] See Table A10 in the Online Appendix.

[41] This is achieved by constructing counterfactual samples in which the dependent variable y_i is permuted across observations so that, say, the permuted $\tilde{y}_i$ for observation i happens to be y_j with $j \neq i$, while the vector of regressors X_i remains unchanged. Random permutation ensures that, in expectation, $E[X_i \tilde{y}_i] = 0$ across all replications. For each permuted sample we calculate the various test statistics that we are interested in. We then compare the actual t-test statistic to the histogram of simulated t-statistics under the null to obtain their two-way equivalent p-value.

Brown, M., Falk, A., Fehr, E., 2012. Competition and relational contracts: the role of unemployment as a disciplinary device. J. Eur. Econ. Assoc. 10, 887–907.
Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90 (3) 414–247.
Cardenas, J.C., Carpenter, J., 2008. Behavioural development economics: lessons from field labs in the developing world. J. Dev. Stud. 44, 311–338.
Chakravarty, S., MacLeod, W.B., 2009. Contracting in the shadow of the law. RAND J. Econ. 40 (3), 533–557.
Charness, G., 2004. Attribution and reciprocity in an experimental labor market. J. Labor Econ. 22, 665–688.
Charness, G., Kuhn, P., 2011. Lab labor: what can labor economists learn from the lab? In: Handbook of Labor Economics, vol. 4. Elsevier Inc., pp. 229–330.
Coate, S., Ravallion, M., 1993. Reciprocity without commitment: characterization and performance of informal insurance arrangements. J. Dev. Econ. 40 (1), 1–24.
Cooper, D.J., Lightle, J.P., 2013. The gift of advice: communication in a bilateral gift exchange game. Exp. Econ. 16, 443–477.
Davies, E., Fafchamps, M., 2016. Who gains from competition? The ultimatum game in a labour market setting in Ghana. In: Social Economics (CESifo Conference Volume). MIT Press. Chap. 3
Davies, E., Fafchamps, M., 2017. Pledging, Praising, Shaming: Experimental Labour Markets in Ghana. IZA Discussion Paper.
Davies, E., Fafchamps, M., 2021. Material Incentives and Effort Choice: Evidence from an Online Experiment Across Countries. Stanford University (mimeo).
van der, H., Eline, C.M., Nelissen, J.H.M., Potters, J.J.M., Verbon, H.A.A., 2001. Simple and complex gift exchange in the laboratory. Econ. Inq. 39 (2), 280–297.
Enke, B., 2019. Kinship, cooperation, and the evolution of moral systems. Q. J. Econ. 134 (2), 953–1019.
Esteves-Sorenson, C., 2018. Gift exchange in the workplace: addressing the conflicting evidence with a careful test. Manage. Sci. 64 (9), 4365–4388.
Fafchamps, M., Minten, B., 2002. Returns to social network capital among traders. Oxf. Econ. Pap. 54, 173–206.
Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018. Global evidence on economic preferences. Q. J. Econ. 133 (4), 1645–1692.
Falk, A., Fehr, E., Zehnder, C., 2006. Fairness perceptions and reservation wages — the behavioral effects of minimum wage laws. Q. J. Econ. 121 (4), 1347–1381.
Falk, A., Gachter, S., Kovacs, J., 1999. Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. J. Econ. Psychol. 20, 251–284.
Falk, A., Huffman, D., MacLeod, W.B., 2008. Institutions and Contract Enforcement. IZA Discussion Paper.
Fehr, E., Falk, A., 1999. Wage rigidity in a competitive incomplete contract market. J. Polit. Economy 107, 106–134.
Fehr, E., Gachter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica 65, 833–860.
Fehr, E., Kirchler, E., Weichbold, A., Gachter, S., 1998. When social norms overpower competition: gift exchange in experimental labor markets. J. Labor Econ. 16, 324–351.
Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. Q. J. Econ. 108, 437–459.
Fehr, E., Klein, A., Schmidt, K.M., 2007. Fairness and contract design. Econometrica 75, 121–154.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114, 817–868.
Fehr, E., Tougareva, E., Fischbacher, U., 2014. Do high stakes and competition undermine fair behaviour? Evidence from Russia. J. Econ. Behav. Organ. 108, 354–363.
Fisman, R., Jakiela, P., Kariv, S., 2014. The Distributional Preferences of Americans. NBER Working Paper 20145. National Bureau of Economic Research.
Fisman, R.J., 2003. Ethnic ties and the provision of credit: relationship-level evidence from african firms. B.E. J. Econ. Anal. Policy De Gruyter 3 (1), 1–21.
Foltz, J., Opoku-Agyemang, K., 2015. Do Higher Salaries Lower Petty Corruption? A Policy Experiment on West Africa Highways. University of Wisconsin–Madison (mimeograph).
Fudenberg, D., Maskin, E., 1986. The folk theorem in repeated games with discounting or with incomplete information. Econometrica 54, 533–554.
Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Boston.
Gachter, S., Falk, A., 2002. Reputation and reciprocity: consequences for the labour relation. Scand. J. Econ. 104, 1–27.
Gachter, S., Fehr, E., 2001. Fairness in the Labour Market – A Survey of Experimental Results. Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets.
Gachter, S., Huang, L., Sefton, M., 2016. Combining “real effort” with induced effort costs: the ball-catching task. Exp. Econ. 19, 687–712.
Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. Econometrica 77 (5), 1637–1664.
Gneezy, U., List, J.A., 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica 74 (5), 1365–1384.
Greif, A., 1993. Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition. Am. Econ. Rev. 83 (3), 525–548.
Grossman, S., Startz, M., 2020. Reputation and relational contracting: experimental evidence on contracting frictions in West African trade. Dartmouth College (mimeo).
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Gintis, E.F.H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., Tracer, D., 2005. “Economic Man” in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. Cambridge University Press, Cambridge.
Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? Behav. Brain Sci. 33 (61–83), 83–135. Discussion
Herrmann, B., Thöni, C., Gachter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Jackson, M., Rodriguez-Barraquer, T., 2012. Social capital and social quilts: network patterns of favor exchange. Am. Econ. Rev. 102 (5), 1857–1897.
Jayaraman, R., Ray, D., De Vericourt, F., 2016. Anatomy of a contract change. Am. Econ. Rev. 106 (2), 316–358.
Kirchler, E., Fehr, E., R, E., 1996. Social exchange in the labor market: reciprocity and trust versus egoistic money maximization. J. Econ. Psychol. 17 (3), 313–341.
Kocherlakota, N., 1996. Implications of efficient risk sharing without commitment. Rev. Econ. Stud. 63 (4), 595–609.
Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely dilemma repeated prisoners” Dilemma. J. Econ. Theory 252, 245–252.
Kube, S., Marechal, M.A., Puppe, C., 2013. Do wage cuts damage work morale? Evidence from a natural field experiment. J. Eur. Econ. Assoc. 11 (4), 853–870.
Macchiavello, R., Morjaria, A., 2015. The value of relational contracts : evidence from a supply shock to Kenyan flower exporters. Am. Econ. Rev. 105, 2911–2945.
MacLeod, W.B., 2007. Reputations, relationships, and contract enforcement. J. Econ. Lit. 45 (3), 595–628.
Malcomson, J.M., 2016. Relational incentive contracts with persistent private information. Econometrica 84, 317–346.
McMillan, J., Woodruff, C., 1999. Interfirm relationships and informal credit in Vietnam. Q. J. Econ. 114, 1285–1320.
Pereira, P.T., Silva, N., Silva, J.A.e., 2006. Positive and negative reciprocity in the labor market. J. Econ. Behav. Organ. 59, 406–422.
Rauch, J.E., Casella, A., 2003. Overcoming informational barriers to international resource allocation: prices and ties. Econ. J. 113 (484), 21–42.
de Ree, J., Muralidharan, K., Pradhan, M., Rogers, H., 2018. Double for nothing? The effects of unconditional teacher salary increases in Indonesia. Q. J. Econ. 133 (2), 923–1039.
Schulz, J., Bahrami-Rad, D., Beauchamp, J., Henrich, J., 2018a. The Origins of WEIRD Psychology. Harvard University (mimeograph).
Schulz, J., Edmond, A., Dsouza, S., Kim, R., Henrich, J., Shariff, A., Bonnefon, J., Rahwan, I., 2018b. The moral machine experiment. Nature 563, 59–64.
Shapiro, C., Stiglitz, J., 1984. Equilibrium unemployment as a worker discipline device. Am. Econ. Rev. 74 (3), 433–444.
Siang, K., Ch'ng, T.R., Waichman, I., 2011. On the role of social wage comparisons in gift-exchange experiments. Econ. Lett. 112 (1), 75–78.
Wu, S.Y., Roe, B., 2007. Contract enforcement, social efficiency, and distribution: some experimental evidence. Am. J. Agric. Econ. 89, 243–258.