

Sonntag, Axel; Zizzo, Daniel John

Conference Paper

Accountability one step removed

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games IV, No. F01-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Sonntag, Axel; Zizzo, Daniel John (2017) : Accountability one step removed, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games IV, No. F01-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168235>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Accountability one step removed

Axel Sonntag^a, Daniel John Zizzo^b

^a Corresponding author, Vienna Center for Experimental Economics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, axel.sonntag@univie.ac.at

^b Behavioural and Experimental Northeast Cluster (BENC) and Newcastle University Business School, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom, daniel.zizzo@ncl.ac.uk

Abstract

In a real effort lab and online team production experiment we analyze individual effort under different conditions of individual accountability. In a repeated setting workers receive a distorted signal about their co-players' previous efforts. We vary the degree to which production can be directly traced back to a participant's individual or randomly drawn effort level. We find that individuals produce much less and the decline of contributions over time is significantly steeper under high as compared to low personal accountability. However, in an additional condition, we allow for endogenous accountability and observe the highest effort level. We conclude that accountability one step removed makes use of a deterrence effect of potential monitoring but without forcing subjects to learn about their under-performing peers, thus limiting the typical decline of contributions over time.

Keywords: team production, imperfect observability, information acquisition, online experiment

JEL Classification: C91, D82, M54

1. Introduction

Consider a team in a company, jointly working on a task over time. In one scenario, of low accountability, workers know about the output of each member of the team, but they do not know how much effort they have put in (and how much output was caused by luck or other factors beyond the control of the workers). In another scenario, of high accountability, they know whether the observed output of each the co-workers is due to chance or not. In this sense, co-workers are accountable for the effort they are putting in. In a third scenario, workers do not know about the source of the observed output and specifically whether it is due to effort, but they can easily find out by spending a very tiny symbolic (ϵ) price. There is still (high) accountability but this is *one step removed*, as workers need to actively choose to get the information. The ϵ price could be thought of as having to search the internet, write an email, or perhaps go to a central office and ask the secretary in person. This differs from the second scenario, which could be thought of (for example) as everyone simply getting an email with information about whether the observed output is due to effort or luck. We present a team production experiment that tests how team output varies over time under different accountability environments.

We find that, surprisingly, accountability one step removed is the most effective in eliciting effort, even though people do not actually choose to take advantage of the option to get the information. Conversely, exogenously set high accountability is the least effective in sustaining effort, as this unravels with time. We combine an initial lab session with an online part taking place over 3 weeks, so as to have a closer correspondence between our experimental setting and natural world teams which genuinely work over time, as opposed to within the maximum 1 or 2 hours timeframe of a standard lab experiment. In a lab setting, subjects are more likely to be affected by an activity bias, by which they will wish to do something rather than do nothing, unless artificial distractors (in the form of alternative tasks) are provided (Crump et al., 2013; Eckartz, 2014; Sitzia et al., 2015). Lab participants are also more likely to be affected by emotional responses, insofar as they have to make decisions just after they learn information relevant for their payoffs; even small decision making delays can make a difference in terms of emotional reactions (Kritikos and Bolle, 2004). Our setup can rely on natural distractors and on the sequencing of tasks over three weeks to avoid both issues, and allows us to more accurately study decision making over time. Furthermore, our team compositions are exogenously set, which is a natural choice for the work environment we are trying to model, and also allows better interpretability in terms of experimental design.

Our experiment compares two information-rich decision making environments.¹ There is clearly reason to believe that workers will behave differently if they feel that what they do is being observed.

In a team production setting, Mohnen et al. (2008) show that agents behave in an inequality-averse way if they know that they will receive information about the others' contributions at an interim stage. In a team production field experiment run on a UK based fruit farm, Bandiera et al. (2005) find that relative incentive schemes induce greater production, but only when workers can observe each other. In a money burning setting, agents are more likely to engage in destructive behavior if they can hide their move behind nature (Abbink and Herrmann, 2011).

Bag and Pepito (2012) provide a theoretical account for the effect of transparency in a two-player public goods game. If inputs are complements, they find a positive effect of transparency on equilibrium effort, particularly because additional information eliminates inferior equilibria. However, if inputs are substitutes, this effect ceases to exist. As in Corgnet et al. (2014), our experiment removes any efficiency maximization motive by having perfectly substitutable inputs and no social returns from contribution. This allows us to control for one possible reason for the effectiveness of transparency. It follows previous research on team production with zero (e.g. Dijk et al., 2001; Vranceanu et al., 2014) or limited (e.g. Cason and Khan, 1999) social returns.²

We are not aware of previous public good or team production experiments, which look at endogenous accountability. Rustagi et al. (2010) come closest. In the context of a very interesting study on forest resource management in Ethiopia, they find that the choice of a costly monitoring technology is linked to higher cooperation. That said, cooperative types are more likely to choose costly monitoring, which makes it hard to identify the pure observability

¹ It bears a relation to, but obviously differs from, experiments that have compared information with no information (Sell and Wilson, 1991), verified the effect of monitoring (Cason and Khan, 1999; Nalbantian and Schotter, 1997), considered different ways of presenting the information (Jones and Mckee, 2004) or changed whether last period's or the current contribution is provided to agents (Nikiforakis, 2010).

² It can be thought of as perhaps most closely modelling combinations of unskilled work as in Bandiera et al.'s (2005) fruit farm. While our rationale for this choice is primarily in terms of experimental control, we note that Carpenter et al. (2009) find no evidence for differences in behavior with low and high social returns from contribution.

effect, or the effect of having a choice, in the lack of controls where there is no choice and where there is either high or low accountability.

Why could accountability one step removed be helpful? Under a self-interest benchmark, obviously, it would not: as long as the marginal cost of effort is higher than the return on it, no effort should be made. Even if this assumption is not met, what others do should still remain irrelevant. But, even leaving aside efficiency maximization, accountability may be helpful as argued by Mohnen et al. (2008) and Bandiera et al. (2005). That said, one of the most common findings of repeated public good contribution experiments is the unraveling of cooperation as the experiment progresses (Andreoni, 1988) – an unraveling largely driven by (either social preferences based or strategic) conditional cooperators (Burlando and Guala, 2005; Fehr and Gächter, 2000; Fischbacher et al., 2001; Fischbacher and Gächter, 2010). Neugebauer et al. (2009) found the sharpest progressive decline in contributions with their information treatment. Intuitively, a conditional cooperator may be more likely to retaliate if he or she knows that the low production of the co-worker is due to the co-worker's low effort rather than to nature.

If, under accountability one step removed, workers seek the information, we should expect the same benefits but also the same potential for unraveling as we would under full accountability. We suggest, however, that workers may not seek the information if it is one step removed. This could be due to a number of reasons, two of them being simply sticking to defaults (e.g. Johnson and Goldstein, 2003; Madrian and Shea, 2001) or strategic information aversion as postulated by Huck et al. (2015) in an individual choice real effort experiment - workers may not be seeking information that may hurt them if they had it.³ This would then imply a situation in which workers under accountability one step removed do not work differently than if there was no accountability, and so do not suffer from the same level of cooperation unraveling that is possible under full accountability. Knowing that others have an option to get information might however be sufficient to induce perceived accountability and therefore greater cooperation – what we label a *Damocles effect*.

Section 2 contains the experimental design and hypotheses. Section 3 presents the results, which show the highest level of effort under accountability one step removed. Section 4 discusses the results and concludes.

³ We discuss possible reasons further in section 2.4.

2. Experimental Design and Hypotheses

2.1. Experimental setting

Subjects and procedures

The experiment was run at the University of East Anglia, UK. Subjects were invited from the CBESS subject pool using ORSEE (Greiner, 2004).⁴ In the lab, randomly selected into groups of four, subjects received paper-based instructions which were also read aloud to them. After an extensive check for understanding, which all subjects needed to pass, they were trained to use their personal IDs to log on to an online system and practiced completing a few sample tasks, to familiarize them with the online environment, the screen layout and the mechanics of task completion.⁵ After completing four tasks, subjects could leave and were provided with a summary sheet of paper containing their personal ID and a schedule for nine working days.⁶ The lab sessions took place in close time proximity. In the following week subjects could work on online real effort tasks (details below) for a total of nine working days that were scheduled for every Monday, Wednesday and Friday for a period of three weeks. A working day started at 8am and ended at 8pm. During these working hours subjects were entirely free to complete up to 20 tasks, but we opted for real effort tasks so that completing the task would take more than a quick decision on a single screens.⁷ They were entirely flexible with respect to when and how many tasks they wanted to complete (if at all). After three weeks, at the end of the online part of the experiment, they were reminded by email that they needed to come to the lab once more to collect their payments in cash and also to complete a very brief pre-payment online questionnaire.⁸

⁴ The subject pool of the Centre for Behavioural and Experimental Social Science (CBESS) contains mainly university students. The sample used for this experiment was well balanced between treatments regarding typical demographic dimensions: the average age was 24.0 years (median: 23.0), 35.7% were male and 13.9% had an economics major.

⁵ The full set of instructions is provided in Appendix A3.

⁶ An example of the information provided on this separate sheet of paper can be found in appendix A.3.3.

⁷ Since this experiment aims to explain team production behavior in workplace environments, we consider using a real effort task to be more appropriate than using induced effort decisions.

⁸ About 12% of subjects forgot to fill-in the final questionnaire before collecting their payment. These subjects were offered the opportunity to complete this questionnaire on a computer provided on site, just before collecting their payments.

Online real-effort task

There exists a huge range of real effort tasks that have been successfully implemented in economics experiments. For our experiment the task needed (i) to be easy to understand, (ii) not to rely on mathematical, logical or language ability, (iii) to require an intermediate amount of time to be completed after a short learning curve at the beginning, and (iv) to take approximately the same amount of time (as a proxy for effort costs) for every task completed, after the trial phase. The chosen task, which is very similar to the one used by Erkal et al. (2011), largely meets these requirements. Participants saw a table of 26 letters and 26 numbers. All table columns were ordered by letters from A to Z and each letter was assigned a unique number (in the same column and below the corresponding letter). We asked subjects to find and enter the numbers corresponding to a sequence of ten letters. They could provide the answers by simply clicking on a dropdown menu below each letter to select the correct number for each letter separately (see Figure 1).⁹

Figure 1: Example task

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
17	21	23	20	19	1	15	12	16	13	2	22	8	26	18	11	6	24	10	4	9	5	7	25	14	3

C	B	M	G	H	H	T	Z	X	R
...

Notes: The table with 26 letters and their corresponding numbers (top) and the drop-down menus to provide the answers (bottom). Both the letter-number combinations as well as the selected sequence of letters were randomized once (before the experiment) and this random order then was kept constant across subjects (i.e. all subjects faced the same tasks in the same order).

2.2. Games and profit maximizing behavior

Each subject could complete up to 20 tasks per working day. In all treatments, subjects knew from the very beginning that there was a 50% probability that a (uniformly distributed) random number from 0 to 20 would be recorded instead of their actual number of tasks. Let x_i be the number of tasks subject i completed successfully and c_i be the constant marginal costs of completing one task for subject i . There is a 50% probability that $x_i c_i$ becomes a worthless investment for subject i as this number is replaced with a random number. Each subject's

⁹ Alternatively, subjects could also speed up the selection of the correct number for each letter by first selecting the appropriate drop-down menu by clicking on it and subsequently typing the number.

realization of the random draw was independent of other subjects' draws. All individually recorded numbers (stemming from either the actual effort or a random draw) were summed up and split equally across all group members. Applying an exchange rate of 1 task = £1, and assuming risk-neutrality, subject i 's payoff can be written as

$$E[\pi_i] = \frac{1}{4} \sum_{i=1}^4 \left\{ \frac{1}{2} x_i + \frac{1}{2} E[U(0,20)] \right\} - x_i c_i.$$

From the above we see that a selfish pay-off maximizer should choose his/her effort level independently of any other group members and would produce an effort of 20 (full effort) if his/her marginal costs of completing one task were less than £0.125. If the marginal costs were higher than £0.125, the narrowly selfish individual would exert zero effort. He or she would be indifferent between all effort levels if $c_i = \text{£}0.125$.¹⁰ Note that we would not like to emphasize this particular threshold or derive behavioral predictions for the experiment. Nevertheless, the payoff function demonstrates the two main differences of our team production task and classical public goods games. Firstly, whereas in the public goods game the selfish payoff maximizer would be best off, if she would not contribute any of her endowment to the public good (and thereby keep her full endowment as her payoff), this is not a dominant strategy in our team production task. Whether or not someone should work, solely depends on her perceived unit cost of production. Second, in contrast to a standard PGG, not contributing to the common project would not result in a positive payoff equal to the endowment, but in a zero payoff, since all income is generated by producing the common good itself.

2.3. *Experimental Treatments*

Our experiment consists of three treatments and we collected data for 15 independent observations each (see Table 1). All treatments had the same payoff structure and in all treatments all players learned the number of recorded tasks of each player of the same group. Hence, the only dimension that was manipulated across treatments was information about the source of the other players' recorded number of tasks in previous periods.

¹⁰ Whereas the prediction for the selfish pay-off maximizer is not affected by her co-workers' effort levels, the appendix outlines benchmarks under Fehr and Schmidt (1999) and Charness and Rabin (2002) preferences, showing how, for sufficiently low effort costs, the effort of a worker should be a positive function of the effort of her co-workers.

Table 1: Number of subjects and observations, by treatment

	Subjects	Independent observations
High Accountability (HA)	60	15
Low Accountability (LA)	60	15
Endogenous Accountability (EA)	60	15

Note: All groups consisted of four subjects each

High accountability (HA)

Recall that the recorded number of tasks could be the result of the actual number of tasks completed or a random integer from zero to twenty, with 50% probability each. In the HA treatment, at the beginning of a period, all subjects not only learned the previous period's recorded number of all subjects in their group, but also the true source of each number.

Low accountability (LA)

In the LA treatment, subjects only learn the previous period's recorded number of all group members. Although they know whether their own recorded number represents their actual number of tasks or is the result of a random draw, it is never revealed to them whether the source of the recorded numbers of the other subjects of their group were their actual efforts or whether these numbers were the result of random draws.

Endogenous accountability (EA)

Whereas in the LA and the HA treatments subjects never learned the true source of their co-players' recorded numbers or were forced to learn this information, respectively, in the EA treatment, subjects could choose whether or not they wanted to receive this information at a tiny cost of 1 penny. Subjects knew that, if they chose to learn this information, no one else would know they had done so, thus eliminating strategic signaling as a motive for (not) receiving information (as instead in Falk and Kosfeld 2006; Fehr and Rockenbach 2003). This is a key element of difference of our setup from related previous research.

In the EA treatment subjects could indicate their wish to learn the true sources of their co-players' recorded numbers, by simply ticking a box beneath the task as displayed on the computer screen (see Appendix A5 for a screenshot). The instructions explained that any choice made would be carried over to all new task screens within a working day and that any decision made on a previous task's screen could be changed by simply ticking or unchecking the

box on the current screen. Only the most recent choice was made binding, either after subjects finished completing all 20 tasks of a working day or when the working day ended before they had finished all 20 tasks.¹¹

2.4. Hypotheses

Hypothesis 1a (H1a): The number of completed tasks will be higher under high accountability (HA treatment) than under low accountability (LA treatment).

As discussed in the introduction, accountability could potentially support greater contributions (Bandiera et al., 2005; Mohnen et al., 2008).¹² We expect information about whether a recorded number is the result of the other subjects' real effort or a random draw, to affect contributions positively.

Hypothesis 1b (H1b): Production will unravel faster in the HA than in the LA treatment.

If workers can observe the true source of their co-workers' output, even identifying a single free-rider might, over time, accelerate spoiling a potentially cooperative climate and lead to greater unraveling of cooperation in the HA treatment than in the LA treatment. Identifying more and more contributions of low effort over time that can be specifically blamed on co-workers, will start a vicious downward cycle of declining cooperation, as observed in many experimental investigations of social dilemma games.

Hypothesis 2a (H2a): If subjects have the option to receive information about the other subjects' true source of contribution (EA treatment) they will acquire this information, thereby exploiting the potential benefits of high accountability (see H1a).

It typically matters to subjects whether output is the results of a subject's costly effort or luck (Cappelen et al., 2013). Furthermore, conditional cooperators are expected to take the intentions and/or real efforts of other subjects into account when making decisions about their own contribution levels (e.g. Fischbacher and Gächter 2010, see introduction). Accordingly, as

¹¹ Note that subjects could return to the online environment as often as they wished to complete tasks or to change their decision with respect to learning the true sources of the recorded numbers, before a working day ended at 8pm that day.

¹² That said, our notion of accountability is different from previous research in the sense that it makes transparent the source of any observable output level, but not necessarily the effort put in to achieve the output (in case the source was a random draw).

the source of the recorded number is likely to matter and given the ε nature of the information cost, subjects are expected to seek information about the true effort exerted by the workers.

Hypothesis 2b (H2b): If subjects in EA acquire information, they will be more responsive to this information than subjects in the situation of exogenously imposed information (HA treatment).

Subjects who intend to base their decisions to complete tasks (and how many of them) on the performance of others on the previous working day, will seek information about the true source of the others' recorded numbers of tasks. Consequently, subjects who seek the information are expected to be more responsive to any discovered information (i.e. positively or negatively reciprocate) than the population average in the HA treatment where high accountability was imposed exogenously. They may also be more responsive out of cognitive dissonance (Festinger, 1957) and self-image concerns (e.g. Dunning, 2007): *because* they sought the information, they *need* to make use of it in order to make sense of having sought it in the first place.

Hypothesis 3a (H3a): If subjects have the option to receive information about the other subjects' true source of contribution (EA treatment), they will refrain from acquiring this information.

In line with the discussion in the introduction, H3a expects subjects not to seek acquiring information even though they can do so at an ε cost. In that sense H3a is the antithesis to H2a. There could be a number of reasons for this. First, in line with the psychological literature on self-perception and self-esteem (Bem, 1972), subjects will not seek this information because, in a Kantian fashion, they might categorically object to spying on others as they also would not want to be spied upon. Second, they might be optimistic about the other subjects' contributions (Neugebauer et al., 2009) and, consistently with their beliefs, deem it unnecessary to monitor others and pay ε for doing so. Third, subjects could have pessimistic beliefs about their co-subjects' contributions and, in a self-deceiving manner (Taylor and Brown, 1988), prefer to see the world through rose-colored glasses and refrain from learning the 'painful and unpleasant truth' about the potentially low numbers recorded for the other group members. Fourth, as in Huck et al. (2015) subjects may have a preference for avoiding information because exactly knowing the truth could induce additional stress they could avoid by choosing not to have the

information (in their case about a low or ten times higher piece rate for a real effort task).¹³ Fifth, subjects may simply be inclined to stick to the default (e.g. Choi et al., 2003; Madrian and Shea, 2001).

Hypothesis 3b (H3b): If subjects in EA do not acquire information, they will complete more tasks as in the LA treatment.

If subjects in EA do not acquire information, they will at the very least put in as much effort as they would under LA. That said, in EA, unlike LA, even if a subject does not acquire information about the source of the co-workers' recorded numbers, he or she knows that others might do so. By inducing perceived accountability, the resulting 'Damocles effect' may induce subjects to exert more effort than subjects whose co-workers never have the chance to learn about the true sources of the group's recorded numbers (LA treatment).

3. Results

3.1. Testing hypotheses

Table 2: Average effort, by treatment

Treatment	Mean effort (all working days)	Mean effort (working days 1-3)	Mean effort (working days 7-9)
HA	5.01	7.33	3.01
LA	6.54	7.84	5.69
EA	8.26	9.61	7.14

Note: Effort refers to the number of correctly completed tasks per working day.

Result 1a: Against H1a, we do not find a difference in exerted effort between HA and LA.

Table 2 contains descriptive statistics. Against H1a, we do not find evidence of higher effort when subjects received information about the source of their co-workers' recorded numbers (HA treatment) as compared to when they received did not (LA treatment) (two-sided Wilcoxon: $p=0.280$).¹⁴

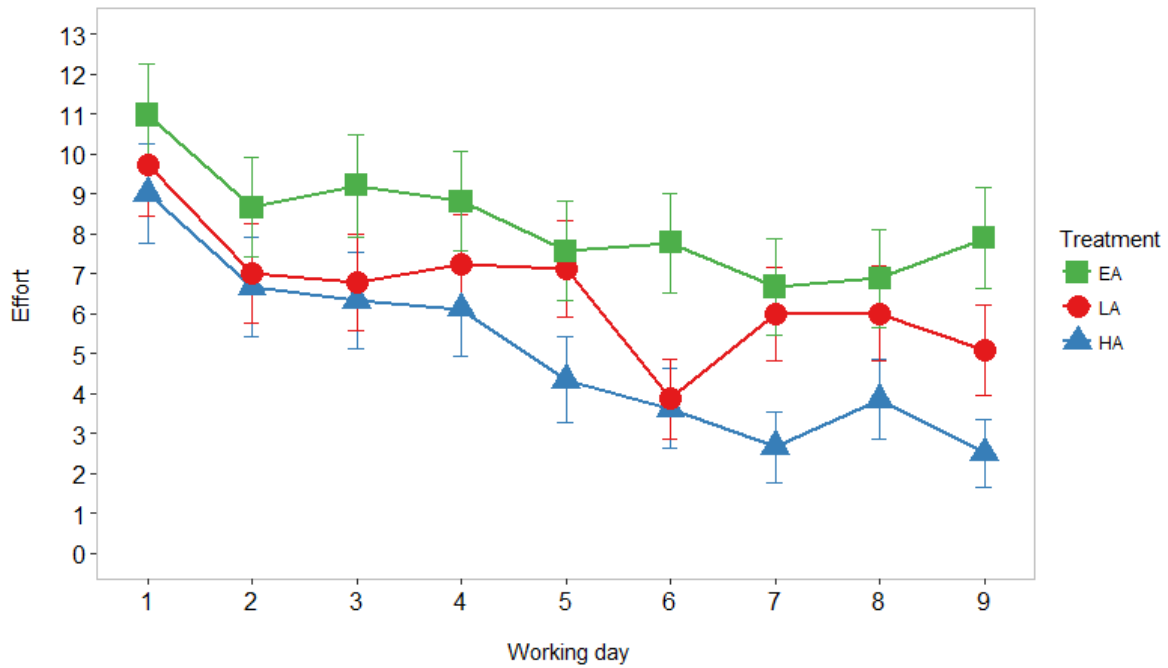
¹³ Huck et al.'s (2015) subjects stated that they did not want to be demotivated by learning about their low wage or they did not want to feel too much pressure by learning about their high wage.

¹⁴ If not stated otherwise, in this article all non-parametric tests are two-sided tests that were performed on group level averages to take care of any non-independence of within-group observations.

Result 1b: In support of H1b, cooperation unravels more in the HA treatment than in the LA treatment as the experiment progresses.

As shown in Figure 2, Result 1a is driven by almost identical effort levels in the first three working days (Wilcoxon: both $p=0.787$) that gradually decline at different slopes, resulting in a significant difference in levels in the last three working days (Wilcoxon $p=0.047$). This is confirmed by the regression analysis in Table 3. Whereas the dummy *Low accountability*, which captures the mean difference in effort levels between LA and HA, is not significant, *Low accountability x working day*, which captures the differences in slopes between LA and HA, is significant at the 5% level across regression models.

Figure 2: Mean effort, by treatment and working day



Notes: Error bars denote standard errors of the mean. Means and standard errors are calculated on the bases of individual observations per working day and treatment. The picture does qualitatively not change when using means and standard errors at the group level.

Table 3: Tobit regressions on real effort

	(1)	(2)	(3)	(4)	(5)
Low accountability	1.665 (1.321)	1.721 (1.348)	-0.248 (1.350)	-1.043 (1.218)	-1.451 (1.162)
Endogenous accountability	3.420*** (1.231)	3.489*** (1.253)	1.216 (1.477)	2.134 (1.440)	2.065 (1.374)
Working day		-0.540*** (0.0821)	-0.853*** (0.155)	-0.891*** (0.158)	-0.934*** (0.154)
Low accountability x working day			0.417** (0.205)	0.424** (0.215)	0.432** (0.219)
Endog. accountability x working day			0.477** (0.201)	0.494** (0.209)	0.509** (0.214)
Gender: male				-2.544*** (0.839)	-2.333** (0.976)
Age				0.509*** (0.0890)	0.466*** (0.101)
Additional controls	No	No	No	No	Yes
Observations	1620	1620	1620	1620	1620
Log. Likelihood	-1363.7	-1345.5	-1342.8	-1283.9	-1217.8
F-test	2.901	5.552	3.705	2.948	2.923

Notes: Reference treatment in all columns: high accountability; all columns contain marginal effects of Tobit models with errors clustered at group level in parentheses. See the appendix for specifications with errors clustered at the subject level. The results of mixed effect Logit, linear and Probit models (each with multi-level error clustering: subjects nested in groups) do not differ qualitatively from the results presented above. Additional controls include: psychological scales (social desirability, Machiavellianism), nationalities and self-assessments as to whether subjects perceive themselves as being an organized and/or a busy person, each measured on a 7-point Likert scale (see Appendix A4 for details on all measures); levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Result 2: Against H2a and in support of H3a, in EA, information about the true source of the recorded number was almost never acquired.

In the EA treatment, only four subjects out of sixty acquired information on the first working day. Three of those continued to seek information on the second working day, with no one else doing so, and no one learnt information as from working day three onwards. Thus information was only sought in 7 out of 540 times (i.e. 1.3% of all cases).¹⁵ Although it is theoretically possible that the cost of 1 penny was perceived prohibitively high, we believe this implausible and interpret this result as evidence that workers were not interested in learning the true source

¹⁵ In a public good setting, Kurzban and Descioli (2008) found that even low information costs decreased the likelihood of information search. Their average acquisition rate was 46% in the costly treatment. However, their setting is considerably different, and in relative terms the cost for information was much higher than in our experiment.

for the recorded number for one on more of the reasons discussed in the motivation for H3a in section 2.4.

As H2b is conditional on people seeking information, and there were too few observations of people seeking information, we cannot reasonable apply statistical tests to identify any behavioral differences between people who did and those who did not seek information.

Result 3: Subjects in EA completed more tasks than those in HA and, consistently with H3b, at least as many tasks as than those in condition LA.

Figure 2 shows that the average effort in EA is always higher than those in HA and LA, in terms of point values.¹⁶ Models 1 and 2 in Table 3 support this claim of a higher average effort in EA than in HA (dummy Endogenous info, z-tests: both $p < 0.01$), but effort in EA is not significantly higher than in LA (test for equality of coefficients: EA dummy vs LA dummy z-tests: both $p \geq 0.147$). Models 3-7 in Table 3 demonstrate that this result is mainly driven by the reduced unravelling in cooperation that is seen in both EA and LA relative to HA.

3.2. Supplementary Evidence

We elicited beliefs of all workers about their average co-worker's effort in the pre-payment questionnaire.¹⁷ Table 4 shows that, in general, subjects had a good understanding of the work effort of their peers, in the sense that their beliefs about the others' effort are strongly correlated with actual effort levels of their co-workers (Spearman: $\rho=0.453$, $p<0.001$). The beliefs for the average of all nine working days are not significantly different from the beliefs about the others' performance on working day nine (Wilcoxon rank-sum tests: all $p>0.6$).¹⁸ Although we cannot observe significant differences between LA and HA (both for the average of all working days and for working day 9 with both $p>0.7$), EA subjects interestingly hold significantly less pessimistic beliefs than subjects in both LA and HA (Wilcoxon, for the average of all working days: LA vs. EA: $p=0.052$ and HA vs. EA: $p=0.017$ and for working day 9: LA vs. EA: $p=0.039$ and HA vs. EA: $p=0.015$).

¹⁶ This can be replicated excluding the 4 subjects who sought information once or twice.

¹⁷ Although, from a statistical point of view, it would be advantageous to elicit beliefs about others' contributions every working day, we deliberately refrained from doing so, in an attempt not to distort subjects' effort behavior.

¹⁸ We asked for beliefs about the others' effort on average and on working day 9 separately, to control for a potential end-game effects in beliefs on working day 9.

Table 4: Beliefs and perception of peers

Treatment	Belief avg. task others, all w.d., (as percentage of actual average effort)	Belief avg. task others, w.d. 9 only	Let down by others
HA	4.84 (96.6%)	4.70	4.35
LA	5.22 (79.8%)	5.05	3.69
EA	6.94 (84.0%)	7.46	3.85

Notes: Beliefs about the average number of tasks completed by the other group members could be stated from 0 to 20. w.d.: working day. The percentages in parentheses are calculated as the average beliefs divided by the actual average effort level per treatment (see Table 2). The impression whether one felt “let down by other group members” was measured from 1 (not at all agree) to 7 (totally agree).

The average perception of “having been let down by ones’ peers” is highest in HA, though with no statistically significant differences to the other treatments (Wilcoxon: all $p > 0.1$).

4. Discussion and Conclusions

Our experiment varied the degree to which output can be directly traced back to a participant’s individual effort level or a random draw. By running our experiment over three weeks, we were able to minimize any experimenter demand coming from people wanting to ‘do something’ in the lab, particularly, once they made the effort of going there. Accountability is normally seen as a good thing (Baker, 2000; Seabright, 1996), but in our experiment high accountability led to the lowest mean team production effort as cooperation most quickly unraveled with time. Our results are consistent with Neugebauer et al. (2009), who found the strongest decline in their information treatment. They are also, in spirit if not in terms of specific setting or adopted solution, close to Steiger and Zultan (2014). In their full information and no information treatments, subjects respectively do or do not learn about their co-participants’ public good contributions. In an intermediate chain information treatment each subject only observes the action of his/her immediate predecessor. They find that the intermediate chain treatment does remarkably well and on average results in higher contributions than their other treatments. They conclude that “partial information can be used to balance the positive and negative effects of transparency” (p. 1).

Under accountability one step removed, workers did not seek the information even though this cost only a symbolic amount. This reflects genuine real world environments where the information is there, and you know it is there, but you need to make an effort to get it (for example, by writing an email or searching the internet). Accountability one step removed was sufficient to do just as well as under no information, and just possibly better. There are settings where no accountability may not be politically, socially or legally viable, and so even a result of equal effectiveness of accountability one step removed is an interesting finding.

While we have controlled for efficiency maximization in our experiment, based on comparing team production and public good contribution research, there is no reason to believe that our key results would not extend to a public good contribution setting. Nevertheless, this is an obvious avenue of extension of this research.

We have provided a number of reasons why workers may not seek information. There is some suggestive evidence that EA workers have significantly higher beliefs about their co-workers' effort, which suggests self-deception as subjects prefer to see the world through rose-colored glasses and refrain from learning the 'painful and unpleasant truth' about the potentially low output recorded for the other group members. There may be other effects at work such as strategic information avoidance, Kantian avoidance of spying or simply sticking to the default. It would be useful to discriminate among different reasons of information avoidance in future research.

Acknowledgements

The financial support of the ESRC (NIBS Grant ES/K002201/1) is gratefully acknowledged. We are especially grateful for helpful comments from Elena Cettolin, Shaun Hargreaves Heap, Stefan Palan, James Tremewan, Alexander Wagner, Ro'i Zultan, and from participants to presentations at the London Behavioural and Experimental Group, the Vienna Center for Experimental Economics, the Choice Lab in Bergen, the 3rd International Meeting on Experimental and Behavioral Social Sciences (IMEBESS), Rome and the 86th Annual Meeting of the Southern Economic Association, Washington, DC. Funding: This work was supported by the UK Economic and Social Research Council [NIBS Grant ES/K002201/1]. The usual disclaimer applies.

References

- Abbink, K., Herrmann, B., 2011. The moral costs of nastiness. *Econ. Inq.* 49, 631–633. doi:10.1111/j.1465-7295.2010.00309.x
- Andreoni, J., 1988. Why free ride? *J. Public Econ.* 37, 291–304. doi:10.1016/0047-2727(88)90043-6
- Bag, P.K., Pepito, N., 2012. Peer transparency in teams: Does it help or hinder incentives? *Int. Econ. Rev. (Philadelphia)*. 53, 1257–1286. doi:10.1111/j.1468-2354.2012.00720.x
- Baker, G., 2000. The use of performance measures in incentive contracting. *Am. Econ. Rev. Pap. Proc.* 90, 415–420.
- Bandiera, O., Barankay, I., Rasul, I., 2005. Social preferences and the response to incentives: Evidence from personnel data. *Q. J. Econ.* 120, 917–962.

doi:10.1162/003355305774268192

- Bem, D.J., 1972. Self-Perception Theory. *Adv. Exp. Soc. Psychol.* 6, 1–62. doi:10.1016/S0065-2601(08)60024-6
- Blanco, M., Engelmann, D., Normann, H.-T., 2011. A within-subject analysis of other-regarding preferences. *Games Econ. Behav.* 72, 321–338. doi:10.1016/j.geb.2010.09.008
- Burlando, R.M., Guala, F., 2005. Heterogeneous agents in public goods experiments. *Exp. Econ.* 8, 35–54. doi:10.1007/s10683-005-0436-4
- Cappelen, A.W., Moene, K.O., Sørensen, E.Ø., Tungodden, B., 2013. Needs Versus Entitlements-an International Fairness Experiment. *J. Eur. Econ. Assoc.* 11, 574–598. doi:10.1111/jeea.12000
- Carpenter, J., Bowles, S., Gintis, H., Hwang, S.H., 2009. Strong reciprocity and team production: Theory and evidence. *J. Econ. Behav. Organ.* 71, 221–232. doi:10.1016/j.jebo.2009.03.011
- Cason, T.N., Khan, F.U., 1999. A laboratory study of voluntary public goods provision with imperfect monitoring and communication. *J. Dev. Econ.* 58, 533–552. doi:10.1016/S0304-3878(98)00124-2
- Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Q. J. Econ.* 117, 817–869.
- Choi, J.J., Laibson, D., Madrian, B.C., Metrick, A., 2003. Optimal Defaults. *Am. Econ. Rev.* 93, 180–185. doi:10.1126/science.151.3712.867-a
- Corngnet, B., Hernán-González, R., Schniter, E., 2015. Why real leisure really matters: incentive effects on real effort in the laboratory. *Exp. Econ.* 18, 1–22. doi:10.1007/s10683-014-9401-4
- Crump, M.J.C., McDonnell, J. V, Gureckis, T.M., 2013. Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS One* 8, e57410. doi:10.1371/journal.pone.0057410
- Dijk, F. Van, Sonnemans, J., Winden, F. Van, 2001. Incentive systems in a real effort experiment. *Eur. Econ. Rev.* 45, 187–214.
- Dunning, D., 2007. Self-image motives and consumer behavior: How sacrosanct self-beliefs sway preferences in the marketplace. *J. Consum. Psychol.* 17, 237–249. doi:10.1016/S1057-7408(07)70033-5
- Eckartz, K., 2014. Task enjoyment and opportunity costs in the lab: The effect of financial incentives on performance in real effort tasks. *Jena Econ. Res. Pap.* 2014–5.
- Erkal, N., Gangadharan, L., Nikiforakis, N., 2011. Relative Earnings and Giving in a Real-Effort Experiment. *Am. Econ. Rev.* 101, 3330–3348.
- Falk, A., Kosfeld, M., 2006. The hidden costs of control. *Am. Econ. Rev.* 96, 1611–1630.
- Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 90, 980–994.
- Fehr, E., Rockenbach, B., 2003. Detrimental effects of sanctions on human altruism. *Nature* 422, 137–40. doi:10.1038/nature01474
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Quarterly J.*

- Econ. 114, 817–868.
- Festinger, L., 1957. A theory of cognitive dissonance. Row, Peterson, Evanston, Il.
- Fischbacher, U., Gächter, S., 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *Am. Econ. Rev.* 100, 541–556.
- Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* 71, 397–404. doi:10.1016/S0165-1765(01)00394-9
- Greiner, B., 2004. The online recruitment system ORSEE 2.0 - a guide for the organization of experiments in economics. Univ. Col. Work. Pap. Ser. 10.
- Huck, S., Szech, N., Wenner, L.M., 2015. Designed Beliefs and Performance: A Real-Effort Experiment. Work. Pap.
- Johnson, E.J., Goldstein, D.G., 2003. Do Defaults Save Lives? *Science* (80-.), New Series 302, 1338–1339.
- Jones, M., Mckee, M., 2004. Feedback Information and Contributions to Not-for-Profit Enterprises: Experimental Investigations and Implications for Large-Scale Fund-Raising. *Public Financ. Rev.* 32, 512–527. doi:10.1177/1091142104267064
- Karakostas, A., Sonntag, A., Zizzo, D.J., n.d. Efficiency and Fairness in Revenue Sharing Contracts. *Scand. J. Econ.*
- Kritikos, A., Bolle, F., 2004. Punishment as a public good. When should monopolists care about a consumer boycott? *J. Econ. Psychol.* 25, 355–372. doi:10.1016/S0167-4870(02)00197-6
- Kurzban, R., Descioli, P., 2008. Reciprocity in groups: Information-seeking in a public goods game. *Eur. J. Soc. Psychol.* 38, 139–158.
- Madrian, B.C., Shea, D.F., 2001. The power of suggestion: Inertia in 401(k) participation and savings behavior. *Q. J. Econ.* CXVI, 1149–1187.
- Mohnen, A., Pokorny, K., Sliwka, D., 2008. Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence. *J. Labor Econ.* 26, 693–720. doi:10.1086/591116
- Nalbantian, H.R., Schotter, A., 1997. Productivity under group incentives: an experimental study. *Am. Econ. Rev.* 87, 314–341.
- Neugebauer, T., Perote, J., Schmidt, U., Loos, M., 2009. Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *J. Econ. Psychol.* 30, 52–60. doi:10.1016/j.joep.2008.04.005
- Nikiforakis, N., 2010. Feedback, punishment and cooperation in public good experiments. *Games Econ. Behav.* 68, 689–702. doi:10.1016/j.geb.2009.09.004
- Rustagi, D., Engel, S., Kosfeld, M., 2010. Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management. *Science* (80-.). 330, 961–965.
- Seabright, P., 1996. Accountability and Decentralization in Government: An Incomplete Contracts Model. *Eur. Econ. Rev.* 40, 61–89.
- Sell, J., Wilson, R.K., 1991. Levels of information and contributions to public goods. *Soc. Forces* 70, 107–124.

- Sitzia, S., Zheng, J., Zizzo, D.J., 2015. Inattentive consumers in markets for services. *Theory Decis.* 79, 307–332. doi:10.1007/s11238-014-9466-8
- Steiger, E.-M.M., Zultan, R., 2014. See no evil: Information chains and reciprocity. *J. Public Econ.* 109, 1–12. doi:10.1016/j.jpubeco.2013.10.006
- Taylor, S., Brown, J.D., 1988. Illusion and well-being: A social psychological perspective on mental health. *Psychol. Bull.* 103, 193–210.
- Vranceanu, R., Ouardighi, F. El, Dubart, D., 2014. Team production with punishment option: insights from a real-effort experiment. *Manag. Decis. Econ.* doi:10.1002/mde

Appendix

A.1 Social preferences

A.1.1 Inequity aversion a la Fehr and Schmidt (1999)

The standard linear inequality averse utility function in Fehr and Schmidt (1999) defines an individual i 's utility $U_i(\pi) = \pi_i - \alpha_i \frac{1}{n-1} \sum_{-i \neq i} \max(\pi_{-i} - \pi_i, 0) - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max(\pi_i - \pi_{-i}, 0)$ where π_i and π_{-i} denote the monetary payoffs for individuals i and $-i$, respectively. The parameters α and β describe the degree to which the individual i dislikes being worse off and better off than the other individuals of the same group $-i$, respectively.

Given such preferences, individual i does not only care about its own costs of completing tasks, but also about how much other peers earn. As can be seen from Table A1, under Fehr and Schmidt (1999) preferences, individuals will complete 20 tasks (the maximum) if the constant unit costs $c < 0.0625$ and they will not complete a single task if the constant unit costs per task $c > 0.1923$. However, if an individual faces unit costs such that $0.0625 < c \leq 0.1923$ he/she will try to match his/her expectations as to how many tasks the other members of his/her group will complete.¹⁹ For this intermediate region of unit costs, higher expectations about the effort of others will increase the effort provided.

Table A.1: Utility maximizing predictions as a function of marginal costs

Narrowly selfish		Fehr & Schmidt (1999)		Charness & Rabin (2002)			
				'behaved'		'did not behave'	
$c < 0.125$	20	$c < 0.0625$	20	$c < 0.128$	20	$c \leq 0.1148$	20
$c = 0.125$	indiff.	$c = 0.0625$	indiff.				
		$0.0625 < c \leq 0.1923$	match exp.	$0.128 \leq c < 0.2171$	match exp.	$0.1148 < c \leq 0.1819$	match exp.
$c > 0.125$	0	$c > 0.1923$	0	$c \geq 0.2171$	0	$c > 0.1819$	0

¹⁹ Note that both the Fehr and Schmidt (1999) as well as the Charness and Rabin (2002) utility functions are not differentiable at $\pi_i = \pi_{-i}$. Nevertheless, it is possible to calculate the threshold levels between the behavioral responses of 0, match expectation and 20.

Notes: c denotes the marginal cost of completing one task. *indiff.* and *match exp.* denote being indifferent between producing any possible levels of effort and to exactly match one's expectation regarding the exerted effort of other group members. The table is based on previously used reasonable parameter values (see Blanco et al. 2011; Karakostas et al. 2016): Fehr and Schmidt (1999) model $\alpha = 1, \beta = 0.35$; Charness and Rabin (2002) model $\rho = 0.424, \sigma = 0.023, \theta = -0.111$.

A.1.2 Tastes for efficiency and reciprocity a la Charness and Rabin (2002)

Although Charness and Rabin's model could be seen as an extension of the Fehr and Schmidt model in the sense that subjects do not only dislike inequity but also care about efficiency and reciprocating behavior, the predictions for the game in question are quite similar. Whereas there are ranges of unit costs that would result in either full or zero effort (see Table A1), there are also ranges in which the best-response lies in matching the expectation about the number of tasks other members of the group will complete. As the Charness and Rabin model allows for reactions to good or bad behavior of other members of a group (reciprocity), they distinguish the case of behaving from misbehaving. If other group members 'behaved', for a range of $0.128 \leq c < 0.2171$ and if other group members 'did not behave' for a range $0.1148 < c \leq 0.1819$, the best response would be to match the expected effort of others. This, again, means that for certain intermediate ranges of unit costs, higher expectations about the production of others will result in higher effort.

A.2 Additional regression results

Table 3 contains regression results based on Tobit specifications with errors clustered at group level. In contrast, Table A2 contains Tobit specifications with errors clustered at subject level. Although a group consists of four subjects, and therefore is the larger level at which observations are non-independent, we argue that also clustering at subject level has its merits. Firstly, it presumably better captures the higher level of heterogeneity between subjects, than between groups and secondly, it also captures within group endogeneity to the extent that subjects' individual behavior is influenced by the group's outcome.

Table A.2: Tobit regressions on real effort

	(1)	(2)	(3)	(4)	(5)
Low accountability	1.665 (1.185)	1.721 (1.206)	-0.248 (1.389)	-1.043 (1.341)	-1.451 (1.345)
Endogenous accountability	3.420*** (1.210)	3.489*** (1.229)	1.216 (1.431)	2.134 (1.424)	2.065 (1.412)
Working day		-0.540*** (0.0763)	-0.853*** (0.170)	-0.891*** (0.177)	-0.934*** (0.177)
Low accountability x working day			0.417** (0.212)	0.424* (0.224)	0.432* (0.229)
Endog. accountability x working day			0.477** (0.204)	0.494** (0.214)	0.509** (0.219)
Gender: male				-2.544** (0.995)	-2.333** (0.979)
Age				0.509*** (0.117)	0.466*** (0.120)
Additional controls	No	No	No	No	Yes
Observations	1620	1620	1620	1620	1620
Log. Likelihood	-1363.7	-1345.5	-1342.8	-1283.9	-1217.8
F-test	3.176	5.915	3.573	2.826	1.791

Notes: Reference treatment in all columns: high accountability; all columns contain marginal effects of Tobit models with errors clustered at subject level in parentheses. Additional controls include: psychological scales (social desirability, Machiavellianism), nationalities and self-assessments as to whether subjects perceive themselves as being an organized and/or a busy person, measured on a 7-point Likert scale; levels of significance: * p<0.10, ** p<0.05, *** p<0.01

A.3 Experimental Instructions

A.3.1 Lab-instructions part 1²⁰

Instructions – part 1

Welcome to this experiment. The session will begin shortly. Before we start, we ask you to turn off your mobile phone and other devices completely. Please refrain from talking to other participants during the experiment. If you have a question at any point during this session, please raise your hand.

This session consists of two parts. In part one you will be asked to complete a questionnaire. After all participants completed the questionnaire you will receive instructions for part two.

At the beginning of the questionnaire in part 1 you will be asked to provide your email address. Please note that you will receive important information via email, so please ensure that you provide a valid email address that you check frequently.

Personal ID

You have been provided with a separate sheet of paper that, among other information, contains your Personal ID. You will be asked to type in your Personal ID before filling the questionnaire of part 1. Please provide your Personal ID when asked to do so. You will learn more about the additional information on the same page as the Personal ID in part 2 of the session.

When you read and understood the instructions of part 1, please indicate that you are ready to start with part 1 by clicking “I understood the instructions of part 1”.

Should you have any questions, please raise your hand.

²⁰ Part 1 instructions were identical across treatments.

A.3.2 Lab-instructions part 2²¹

Instructions – part 2

In part 2 of this session you learn how you can earn money. How much you will earn depends on your decisions and on the decisions of other participants. All decisions will be absolutely anonymous, i.e. your identity will neither be revealed to your co-participants nor to the experimenters at any time during or after the experiment. You will be matched with three other participants to form a group of four. In every group, each participant will be randomly assigned to take the role of participant 1, 2, 3 or 4. The composition of each group and the roles of participants will not change throughout the experiment. Groups are independent, in the sense that what happens in the other groups will not affect the earnings of your group in any way.

Tasks

During this experiment you can complete tasks to earn money. The more tasks you complete, the more money you can earn. The task in the following example is similar to the tasks you will be asked to complete during the experiment. In every task you will see a table of 26 letters and 26 numbers. Note that all table columns will be ordered by letters from A to Z. Each letter is assigned a unique number (in the same column and below the corresponding letter). For example, in the table below the letters A and B are assigned the numbers 17 and 21, respectively.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
17	21	23	20	19	1	15	12	16	13	2	22	8	26	18	11	6	24	10	4	9	5	7	25	14	3

In all the tasks you will be asked to find the numbers corresponding to a sequence of letters. In the example below, you are asked to find the numbers assigned to the letters C, B, M, G, H, H, T, Z, X and R. You can provide your answers by simply clicking on the dropdown menu below each letter to select the correct number for each letter separately.

²¹ Part 2 instructions varied across treatments. Sections in $[[\dots]]$, $\{\{\dots\}\}$ and $\|\dots\|$ were exclusively used in the treatments HA, LA and EA, respectively.

C	B	M	G	H	H	T	Z	X	R
...

↓

C	B	M	G	H	H	T	Z	X	R
23	21	8	15	12	12	4	3	25	24

You can only proceed to the next task if all ten letter – number combinations are entered correctly. You will have the opportunity to practice tasks similar to the one described above later on in this session. None of the trial tasks that you complete during this session will contribute to your earnings or to the earnings of other participants.

Timing

To earn money in this experiment you can complete up to 20 tasks per working day online. The total experiment consists of 3 working days per week for the total duration of 3 weeks (i.e. 9 working days in total). A working day starts at 8:00 hours in the morning (8am) and ends at 20:00 hours in the evening (8pm). You are entirely free to complete as many tasks as you wish during these working hours (up to 20 per working day). It is not possible to complete any tasks outside the working hours. Please find a list, which you can take home, of all working days and times as well as the internet address where you can complete tasks online.

Earnings

On each working day you have a 50% probability that your actual number of correctly completed tasks is recorded and a 50% probability that a whole number is drawn from 0 to 20 (i.e. 0, 1, 2, ..., 18, 19, 20) at random and is recorded instead of your actual number of correctly completed tasks. In the latter case, any number from 0 to 20 has an equal chance of being drawn and recorded as your number of tasks. For example, assume that you completed 8 tasks. Then you have a 50% chance that 8 is recorded as your number of tasks and a 50% chance that a randomly drawn whole number from 0 to 20 is recorded as your number of tasks.

[[All members of your group will be informed whether your recorded number was your actual number of tasks or a random number.]]

{{ No one will be informed whether your recorded number was your actual number of tasks or a random number. }}

|| As a default, no one will be informed whether your recorded number was your actual number of tasks or a random number. However, all members of the group, at a cost of 1 penny in the randomly selected working day (see below for details), can learn whether each group member's recorded number was his/her number of completed tasks or the result of a random draw. You can learn this information by ticking a box on the screen, and in every working day you will be allowed to change your mind by ticking or un-ticking this box until the last task you complete. Getting this information is entirely optional. ||

On each working day the number of recorded tasks of all four participants of your group will be summed up and then split equally across all group members, i.e. one quarter of the total number of recorded tasks of the group will be credited to each group member. For example, assume that for participants 1, 2, 3 and 4 of a group 8, 12, 0 and 20 tasks were recorded, respectively. This group's total number of recorded tasks would be $8+12+0+20=40$. Consequently, $40/4 = 10$ recorded tasks would be credited to each group member on this working day. As from working day 2 onwards, on every working day, you will be informed about the individual recorded number of tasks per participant as well as about the total number of recorded tasks of your group on the previous working day.

Final Payment

After nine working days one working day will be randomly selected for payment per group. Each working day has an equal chance of being selected (1 out of 9).

[[{{ All participants will be paid £1 per credited task on the randomly chosen working day. If, for example, the above working day was chosen at random (i.e. the total number of completed tasks in that group was 40 on that working day), every member of this group would earn $10 \times £1 = £10$.]}]}

|| On the randomly chosen working day all participants will be paid £1 per credited task minus the cost of learning about the real source of all group members' recorded numbers (if

applicable), i.e. 1 penny (£0.01). If, for example, the above working day was chosen at random (i.e. the total number of completed tasks in that group was 40 on that working day), each member of this group who learned about the source of the recorded numbers of the chosen working day would earn $10 \times £1 - £0.01 = £9.99$. Every member of this group who did not learn this information would earn $10 \times £1 = £10$. ||

Every participant will be paid in private and in cash in the week from Monday 16th to Friday 20th of March 2015.

Personal ID and session number

You have been provided with a separate sheet of paper that contains your Personal ID, your session number and a working day schedule. Please keep it safe and show it to no one. You need your Personal ID and your session number to identify yourself before you can complete tasks online on all nine working days. This information will also be sent to you by email. Furthermore, please note that you will be informed about the specific payment time and location by email, so please ensure that the email is not treated as spam by your email provider.

A.3.3 Personal ID and session number

This sheet of paper was taken home by all participants and contained a unique combination of Personal ID and sessions number which was used to identify each subject and as a password combination to log on to complete tasks during working hours.

Personal ID and session number

Personal ID: ACT021406061700

Session number: 25

Schedule of working days

Working day 1	09/06/2014	8am until 8pm
Working day 2	11/06/2014	8am until 8pm
Working day 3	13/06/2014	8am until 8pm
Working day 4	16/06/2014	8am until 8pm
Working day 5	18/06/2014	8am until 8pm
Working day 6	20/06/2014	8am until 8pm
Working day 7	23/06/2014	8am until 8pm
Working day 8	25/06/2014	8am until 8pm
Working day 9	27/06/2014	8am until 8pm
Payment week	30/06-04/07	Exact time and location to be confirmed by email

Website to complete tasks

<http://www.socexp.org>

A.4 Questionnaires

A.4.1 Lab session part 1 questionnaire

This questionnaire was implemented electronically using LimeSurvey (www.limesurvey.org) and had to be filled during the first part of the initial lab session.

Page 1

Please provide an email address that you regularly check!

- Please enter your email address below.
- Please re-enter your email address.

Page 2

Please read each of the below statements and indicate the extent to which you personally agree or disagree with the statement! [For each statement subjects had to select one response from a 7-point scale ranging from *Strongly agree* to *Strongly disagree*]

- Never tell anyone the real reason you did something unless it is useful to do so.
- The best way to handle people is to tell them what they want to hear.
- One should take action only when sure it is morally right.
- Most people are basically good and kind.
- It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.
- Honesty is the best policy in all cases.
- There is no excuse for lying to someone else.
- Generally speaking, people won't work hard unless they're forced to do so.
- All in all, it is better to be humble and honest than important and dishonest.
- When you ask someone to do something for you, it is best to give the real reason for wanting it rather than giving reasons that might carry more weight.
- Most people who get ahead in the world lead clean, moral lives.
- Anyone who completely trusts anyone else is asking for trouble.
- The biggest difference between most criminals and other people is that criminals are stupid enough to get caught.
- Most people are brave.
- It is wise to flatter important people.
- It is possible to be good in all respects.
- The saying that there's a sucker born every minute mistakenly underestimates people.
- It is hard to get ahead without cutting corners here and there.
- People suffering from incurable diseases should have the choice of being put painlessly to death.
- People more easily forget the death of their father or mother than the loss of their property.

Page 3

Below you will find a list of statements. Please read each statement carefully and decide if that statement describes you or not. If it describes you, check the word "yes"; if not, check the word "no". [For each statement subjects had to select one of two options: “yes” or “no”]

- I sometimes litter.
- I always admit my mistakes openly and face the potential negative consequences.
- In traffic I am always polite and considerate of others.
- I always accept others' opinions, even when they don't agree with my own.
- I take out my bad moods on others now and then.
- There has been an occasion when I took advantage of someone else.
- In conversations I always listen attentively and let others finish their sentences.
- I never hesitate to help someone in case of emergency.
- When I have made a promise, I keep it--no ifs, ands or buts.
- I occasionally speak badly of others behind their back.
- I would never live off other people.
- I always stay friendly and courteous with other people, even when I am stressed out.
- During arguments I always stay objective and matter-of-fact.
- There has been at least one occasion when I failed to return an item that I borrowed.
- I always eat a healthy diet.
- Sometimes I only help because I expect something in return.

Page 4

Please answer the following questions where 1 means "not at all" and 7 means "definitely".

- Would you consider yourself to be a very ORGANIZED person?
- Would you consider yourself to be a very BUSY person?

Page 5

[Subjects had to provide an integer number in response to each of the following two questions.]

- How many emails do you on average receive per day?
- How many of those would you consider worth reading?

[Subjects were requested to provide information about their demographics]

- What is your current age (in years)? [Integer number]
- What is your gender? [Female/Male]
- What is your main Field of Study? [Free text]
- What is your nationality? [British/Chinese/other (free text)]
- How would you rate your command of the English language? [Beginner/Moderate/Good/Excellent/Native language]

A.4.2 Pre-payment questionnaire

Subjects were reminded to collect their experimental pay-off by email. In the same message they were also informed that they needed to fill a pre-payment questionnaire before collecting their payment. This questionnaire was also implemented electronically using LimeSurvey (www.limesurvey.org).

- Did you feel let down by the other participants in your group? Please tick a number from 1-7 where 1 means "Not at all" and 7 means "Totally".

- Please guess how many tasks the other participants of your group on average completed per working day? [R^+ numbers]
- How confident are you about this guess? Please tick a number from 1-7 where 1 means "Very unconfident" and 7 "Very confident". [7-point scale]

- Please guess how many tasks the other participants of your group on average completed on the last working day only, i.e. working day 9? [R^+ numbers]
- How confident are you about this guess? Please tick a number from 1-7 where 1 means "Very unconfident" and 7 "Very confident". [7-point scale]

Page 4

- What do you think is the objective of this experiment?
 - I do not know / I have not thought about it
 - I thought about it and I think it was the following: [free text]

Page 5

Thank you for filling this final questionnaire!

Remember that you can collect your payment in the week from Monday 30th of June and Friday 4th of July. Please come to office 3.68 on the 3rd floor of the Arts 2 building between 9:00 (9am) and 13:00 (1pm) on any of these days to collect your payment.

For any problems, please send an email to admin@socexp.org.

A.5 Screenshot of decision screen

Figure A.1: Screenshot of task screen in treatment Endogenous Accountability

Working Day 5
0% 100%

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
20	10	7	23	13	18	4	5	8	17	16	11	3	12	1	15	9	22	2	19	26	24	6	21	25	14

Please enter the numbers that correspond to the following sequence of letters:

U	Z	V	G	Y	E	Y	C	A	X
26	14	24	4	25	13	25

At the beginning of the next working day, you will learn about the recorded number of tasks of all group members. Optionally, at the cost of 1 penny (if this working day will be randomly selected for final payment), you can also learn whether each group member's recorded number was his/her actual number of completed tasks or the result of a random draw.

Click the box below if you want to learn about the source of your co-group members' recorded numbers. (This is entirely optional.)

Learn the source of the recorded number of each group member

Notes: The top part of the task screen in the treatments Low and High Accountability was identical, but the bottom part was only displayed in the Endogenous Accountability condition.