Personal accountability and cooperation in teams

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 and Institute for Advanced Studies, Josefstädter Straße 39, 1080 Vienna, Austria, axel.sonntag@univie.ac.at

[b] School of Economics, University of Queensland, St Lucia Qld 4072, Australia, d.zizzo@uq.edu.au

Abstract

In a real effort lab and online team production experiment, we analyze exerted effort under different conditions of individual accountability. In a repeated setting, we vary the degree to which production can be directly traced back to a collaborator's individual or randomly drawn effort level, respectively. We find that individuals produce much less and the decline of effort over time is significantly steeper under high as compared to low and endogenously chosen personal accountability. While endogenous accountability provides an option for monitoring others, it does not force subjects to learn about their under-performing peers, thus limiting the typical decline of contributions over time. We conclude that accountability one step removed may be an interesting institutional setting for repeated collaborations in contexts where low accountability for political, social or legal reasons is not a viable option.

Keywords: team production, imperfect observability, information acquisition, lab and 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 of 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. Such peer accountability occurs in many real-life contexts, and particularly so in the workplace. Therefore, we will investigate the specific role of different levels of individual accountability and will present a team production experiment that tests how team output varies over time under different accountability environments.[1]

We find that accountability one step removed is surprisingly effective in eliciting cooperation, even though people do not actually choose to take advantage of the option to get the information. Conversely, exogenously set high accountability performs very poorly in sustaining cooperation, as this unravels with time.

In order to have a closer correspondence between our experimental setting and natural world teams, which genuinely work over time, we combine an initial lab session with an online part taking place over 3 weeks, as opposed to 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., 2014).

---

[1] In contrast to a vertical notion of accountability, e.g. towards a line manager (e.g. Sonntag and Zizzo, 2015), in this paper, we consider a horizontal notion of accountability, i.e., peer effects among co-workers (e.g. Mas and Moretti, 2009). The latter particularly relates to a huge literature on conditional cooperation in team production and public goods settings (e.g. Fischbacher and Gächter, 2010; Steiger and Zultan, 2014).

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 study decision making over time more accurately.[2] Furthermore, our team compositions are exogenously set, which is a natural choice for the work environment we are trying to model, and also allows for better interpretability in terms of experimental design.

Our experiment compares two information-rich decision making environments.[3] There is clearly reason to believe that workers will behave differently if they feel that what they do is being observed. For example, 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 monitoring one's fellow co-workers, and being monitored by them, matters. When comparing a piece rate with a relative incentive scheme (with negative externalities on the co-workers), they find that the relative incentive scheme reduced production output, but only when workers can observe each other, the bottom line being that, in order to act on some sort of social preferences, one needs to know how well-off the others are. Thus, peer accountability seems to matter.

In contrast to previous experiments, in our setting, subjects always learn the output of each individual group member, and they know that (in expectation) in 50% of all cases, the observed output is caused by a subject's input. That is, even in settings of "low" accountability, subjects get a sense of the level of individual contributions, since about half of the observed outputs are fully informative of the participants' effort (and not subject to noise). In that sense even our "low" accountability condition conveys more information than situations where only the

---

[2] Whereas we consider our hybrid lab and online setup as a useful methodological innovation for the research question at hand, it is clear that allowing for natural distractors also means loosening experimental control. An alternative to our setup would be to run an experiment on Amazon's Mechanical Turk. Although this would most likely also reduce activity bias, AMT experiments are not ideal for investigating repeated interactions of the same participants over longer periods of time (3 weeks in this study).

[3] 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).

aggregate production level is known (the typical "no info" treatment in the existing literature). Since in our condition of "high" accountability, the true source of all subjects' contributions to the joint project is known, that is, one learns who is responsible for his/her output in that period and who is not, what we actually vary is the individual-specific precision of the informative signal on the subjects' effort. This best reflects our notion of accountability used in this paper. And it is relevant, because there is experimental evidence (e.g. Almas et al. 2010; Cappelen et al. 2013) that the extent to which one is responsible (in terms of personal effort) for his/her outcome matters.

In a very different domain, yet related in terms of studying the importance of personal accountability, Abbink and Herrmann (2011), found that agents are more likely to engage in destructive behavior if they can hide their move behind nature. 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 efficiency gains from team production. 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) efficiency gains.[4]

We are not aware of previous public good or team production experiments, which look at endogenous accountability and its effect on cooperation outside of the laboratory (in our case online) and over a longer period of time (in our case three weeks). Corgnet et al. (2015) and Rustagi et al. (2010) come closest.

Using an innovative lab setup to study team production under either individual or team incentive schemes, Corgnet et al. (2015) compare treatments with different monitoring technologies. More specifically, they vary whether peers could monitor the other group members' production behavior and whether the observed individuals know that they are being

[4] 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 efficiency gains from contribution.

monitored, or not. They find that observable peer monitoring led to significantly higher production, on par with the levels achieved in the individual incentive scheme. However, when subjects had the chance to monitor others, without them knowing to be monitored (which comes closest to our endogenous accountability treatment, see section 2.3), the average production was as low as when there was no monitoring at all.

In the context of forest resource management in Ethiopia, Rustagi et al. (2010) 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 effect, or the effect of having a choice, in the lack of controls where there is no choice and where there is, in our terms, 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 (see section 2.2). However, accountability may create a sense of peer pressure that, in turn, increases production efforts (Falk and Ichino, 2006; Mas and Moretti, 2009; Mohnen et al., 2008). 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). 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 high accountability. However, 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 (Johnson and Goldstein, 2003; Madrian and Shea, 2001) or strategic information aversion as postulated by Huck et al. (2017) in an individual choice real effort experiment - workers may not be seeking information that may hurt them if they had it.[5] This would then imply a situation in which workers under accountability one step removed do not work differently than if there was low accountability,

[5] We discuss possible reasons further in section 4.1.

and so do not suffer from the same level of cooperation unraveling that is possible under high accountability. There is an important difference though: knowing that others have an option to get information might be sufficient to induce perceived accountability and therefore greater cooperation – what we label a *Damocles effect*.[6]

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.

2. Experimental Design and Hypotheses

2.1. Experimental setting

Subjects and procedures

The experiment was run in 2014 at the University of East Anglia, UK. Subjects were invited from the CBESS subject pool using ORSEE (Greiner, 2004).[7] In the lab, 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.[8] After completing four tasks, subjects could leave and received a summary sheet of paper, containing their personal ID and a schedule for nine working days.[9] All lab sessions took place in the week before the online part started simultaneously for all treatments. Subjects were randomized into groups of four after all lab sessions were finished. Using this matching protocol had two advantages: (i) gaining flexibility in terms of the number of participants per lab session and (ii) reducing the chance that later group-members get to know each other in person. Neither the

---

[6] Damocles is put in a situation of constant fear in the classical sword of Damocles story; in an analogous way, co-workers in the endogenous accountability treatment may constantly fear that their potential free-riding behavior is exposed to others.

[7] 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.

[8] The full set of instructions is provided in Appendix A3.

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

actual lab session nor the personal ID used to play the game was informative about the matching group.[10] As from the week following the lab sessions, 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 each, but we opted for real effort tasks so that completing the task would take more than a quick decision on a single screen.[11] 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.[12]

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 letter coding 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

[10] Of course, such an online setting provides less control than a standard lab experiment, e.g. we cannot entirely rule out communication among participants. However, the matching protocol implemented should limit issues related to ‘uncontrolled communication’.

[11] 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 (see Cappelen et al., 2010).

[12] 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.

the answers by simply clicking on a dropdown menu below each letter to select the correct number for each letter separately (see Figure 1).[13]

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

In this paper, we are interested in understanding the role that personal accountability plays when deciding on how much effort to exert on a joint project. We attempt to mimic common features of real workplace environments where collaborators can typically observe the contribution of the team members in terms of output (e.g., everyone can read the literature review section of a joint paper, produced by co-author A). However, what typically cannot be observed is how much effort a collaborator needed to exert to produce the (observable) output (e.g., co-author A could have been lucky in finding a recent review paper, making the job of summarizing the relevant literature much easier, or not).

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.[14] Each subject's

[13] 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.

[14] The reason for choosing this specific setup for the noise process (instead of a more standard procedure like an additive noise term) was that (i) subjects should not be able to identify to what extent the output was driven by noise and (ii) the incentive to exert effort should not be distorted. An additive noise term does not satisfy these conditions. Nicklisch et al. (2016) use a similar noise structure.

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, i.e., all subjects of a group received the same experimental earnings. 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 payoff 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 $c_i = £0.125$.[15] Note that we would not like to emphasize this particular threshold or use it to 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 standard public goods games, 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 subjects learned the number of recorded tasks of each subject of the same group.

---

[15] Whereas the prediction for the selfish pay-off maximizer is not affected by her co-workers' effort levels, the appendix A.1 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.

Hence, the only dimension of manipulation across treatments was information about the source of the other subjects' recorded number of tasks on the previous working days.

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 the working days 2-9, all subjects not only learned the previous working day's recorded number of all subjects in their group, but also the true source of each number.[16]

Low accountability (LA)

In the LA treatment, on the working days 2-9, subjects only learn the previous working day'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-workers' 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.[17] Subjects could indicate their wish to learn the true sources of their co-

[16] Note that learning the true source of the recorded number does not reveal a worker's true number of completed tasks if a random number was recorded for that worker on the previous working day.

[17] Some theoretical work suggests that efficient outcomes in repeated dilemma games are achievable when players receive the option to privately monitor their peers (Miyagawa et al., 2008; Yamamoto, 2007). In their spirit, and although in our setting subjects can only obtain noisy signals, they 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 experimental research (a notable exception is Corgnet et al., 2015). We chose

workers' 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 unticking 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.[18]

2.4. Hypotheses

Based on findings from the previous literature, we derive behavioral hypotheses. An explanation as to how we did that is provided directly below each hypothesis.

Hypothesis 1 (H1): The number of completed tasks will be higher under high accountability (HA) than under low accountability (LA). The number of completed tasks under endogenous accountability (EA) will be between LA and HA.

In our setting, accountability could potentially support greater contributions. As previously noted, our notion of accountability is different from previous research on the role of information in public good games in the sense that it makes transparent the source of any observable individual output level, but not necessarily the effort put in to achieve the output (in case the source was a random draw). Importantly, because of our information rich setting in which subjects always receive individual level information (but the information quality is different between our treatments), the studies referred to below are not directly comparable (e.g., since they compare settings where information on either only aggregate or individual contributions is available). However, there is a similarity in terms of varying the precision of information. Thus, we derive hypotheses in terms of how *more precise* information (in our case, higher information quality) would affect contributions to team production.

With the above qualification, our work is related to the literature on information disclosure in social dilemma games. Frey (1993) provides a theoretical account under which circumstances

---

a price of 1 penny (as opposed to costless information) to obtain a clear-cut prediction, in case subjects attached zero value to this information.

[18] 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.

such *monitoring* of information on individual contributions might be a good thing for cooperation. Whereas Cox and Stoddard (2015), van der Heijden and Moxnes (1999) and Weimann (1994) find no effect of individual information on contributions, Carpenter (2004) finds that information on individual contributions leads to faster unravelling of cooperation. In particular, he concludes that when subjects learn about the low contributions of others, they try to conform to the observed group's norms of contributing less. Eventually, such behavior leads to a downward spiral. On the other hand, Bigoni and Suetens (2012), Sell and Wilson (1991) and Nikiforakis (2010) find that providing information on individual contributions increases cooperation in comparison to similar scenarios with aggregate or no information.[19] For reasons of statistical power, the single most important source for our behavioral hypothesis that HA will lead to higher contributions than LA is a recent meta-analysis by Fiala and Suetens (2017). They review 71 studies on voluntary contribution mechanisms (and 18 studies on collusion in oligopoly settings) and find that individual information on contributions "tends to lead to an increase in contributions" (Fiala and Suetens, 2017, p. 755).

This prediction is also in line with Aoyagi and Fréchette (2009), who find that the less noisy signals about the others' cooperation are, the more willing people are to reciprocally cooperate. In our information rich environment, knowledge of personal responsibility could enforce more effective punishment (by reducing one's own contribution to team production), which in turn could sustain higher cooperation. Consequently, while appreciating that the evidence is mixed and that our information setting is quite different from those used in previous studies, as an educated guess based on the existing literature, we expect information about whether a recorded number is the result of the other subjects' real effort or a random draw, to increase contributions.

---

[19] Nikiforakis (2010) showed that while providing information on individual contributions could increase cooperation (as compared to aggregate level information), information about individual payoffs leads to the opposite effect. That explains why Weimann (1994) did not find any difference between individual and aggregate information conditions since in his individual information conditions, he provided both information on individual contributions and payoffs. Thus, two counteracting effects might have cancelled out each other. Furthermore, Bigoni and Suetens (2012) show that when individual information on both contributions and earnings is present, a negative aggregate effect obtains. Cox and Stoddard (2015) investigate several dimensions including partner/stranger matching, give/take framing and individual/aggregate information about contributions. In their partners-give-individual/aggregate comparison (which is closest to our setting), they find that individual information slightly increases cooperation, but not significantly so. Van der Heijden and Moxnes (1999) analyze contributions in a public bad framing and find no difference between individual and aggregate information conditions.

The endogenous accountability condition EA allows subjects to select their informational state, i.e., they either choose to be in an HA or an LA like information setting. Therefore, EA can be thought of as a convex combination of HA and LA environments and we expect that EA results in average cooperation levels between HA and LA.

Hypothesis 2 (H2): Cooperation will unravel faster in the HA than in the LA treatment, with an intermediate decline in EA.

The literature on the role of information on contributions in public good games suggests that, if information on individual contributions is available, higher contributions can be sustained (see H1). However, taking the role of intentions into account, we should only observe such behavior as long as the information conveyed is one of cooperation (Rand et al., 2015). In our information rich environment, where workers can observe the true source of their co-workers' output, even identifying a single free-riding attempt 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 lead to negative reciprocal behavior, as observed in many experimental investigations of social dilemma games without introducing an explicit punishment technology (Fehr and Gächter, 2000). Since in EA subjects can choose the information richness of their situation, i.e., whether they prefer being in an HA or LA like information environment, we expect unravelling in EA to lie between HA and LA.

Hypothesis 3 (H3): If subjects have the option to receive information about the other subjects' true source of contribution, they will acquire this information.

It typically matters to people whether output is the result of a subject's costly effort or luck (Cappelen et al., 2013b). Furthermore, substantial experimental evidence suggests that subjects behave as conditional cooperators, i.e., they reciprocate to their co-players behavior (Falk and Fischbacher, 2006; Fehr and Gächter, 2000; Fischbacher and Gächter, 2010). In the context of public goods games, what is commonly observed is that subjects reciprocate the effort exerted by other group members (Andreoni and Petrie, 2004). Although the incentive structure of our team production game is not identical to a standard public goods game (as we control for efficiency gains), the behavioral principle of conditional cooperation is applicable in a very similar way. Thus, subjects are expected to take the intentions and real efforts of other subjects into account when making decisions about their own contribution levels.

In a recent article, Tasch and Houser (2018) term the demand for information about others' actions and outcomes in social dilemma games "social curiosity". In particular, they distinguish instrumental from pure curiosity. While learning about others' actions is *instrumental* for making one's own action depend on the action of others in the first place, in addition, there might also exist a genuine non-instrumental wish to learn about what is going on in general (i.e., *pure* curiosity). In an experimental public goods setting, Tasch and Houser (2018) compare situations of varying cost of obtaining social information and find that most of the subjects (86%) learn about the outcomes of others when they have the (costly) option to do so. Similarly, in another public goods experiment with endogenous information look-up, Kurzban and Descioli (2008) find that on average subjects took the opportunity to look-up information in 46% of all cases, strongly suggesting that information on others' contributions matters to subjects in public goods settings.[20] Subjects were willing to pay even substantial economic cost in order to receive information about the other subjects' contributions. Given that, in comparison to the studies mentioned above, the relative cost subjects had to pay for receiving information in our setting were much lower, we expect that subjects will seek information about the source of the recorded numbers of their peers when having the opportunity to do so.[21]

Although, based on the above, we expect that subjects will opt for receiving the additional information, we acknowledge that there may be behavioral mechanisms by which people might refrain from learning the additional information. We will discuss these in section 4.1.

[20] Note that in Tasch and Houser (2018) the cost of having the information were 0, 2%, 10%, and 20% of the subjects' endowment and in Kurzban and Descioli (2008) this cost was equal to 4% of the respective endowment. Since in our design subjects receive no endowments, but need to work for earning any payoff. Nevertheless, we can put the cost of information in perspective. In comparison, our information cost of 1 penny approximately equals 0.1% of subjects' average earnings.

[21] There is also a growing literature on subjects' demand for qualitatively different sorts of information in social dilemma games (e.g., see Eckel and Petrie, 2011 and Samak and Sheremeta, 2013 for varying the identifiability of subjects via portrait pictures).

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.

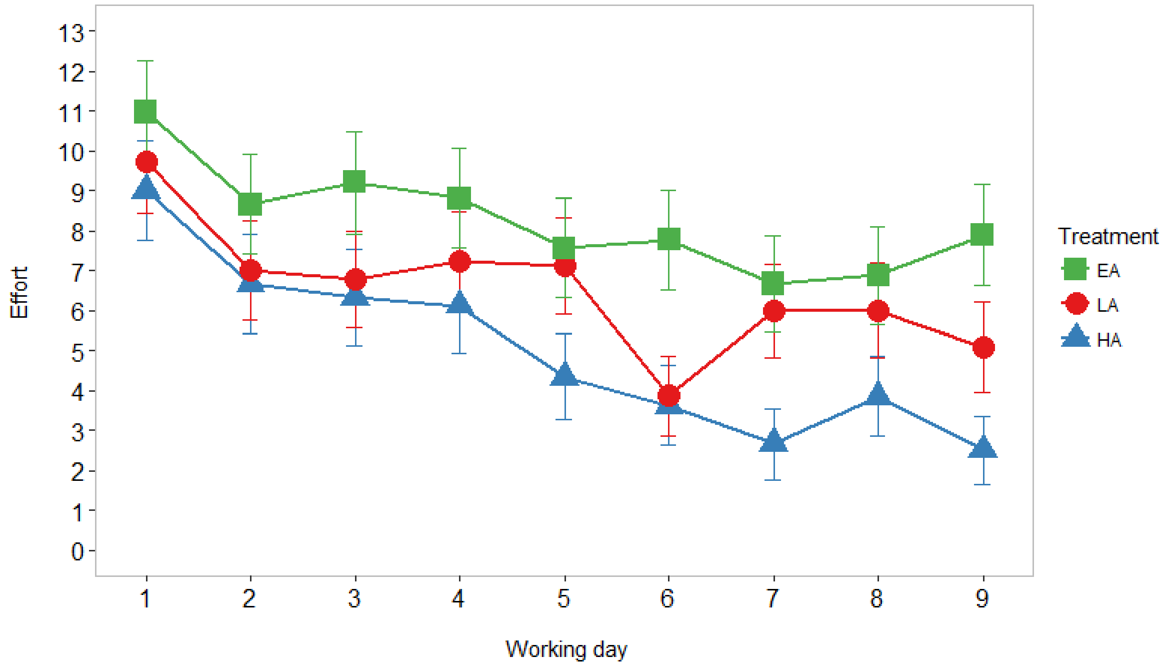
Result 1: Exerted effort under EA is significantly higher than under LA, and effort under LA is significantly higher than under HA.

Table 2 contains descriptive statistics.[22] Overall, we do not find evidence of higher effort when subjects received information about the source of their co-workers' recorded numbers (HA) as compared to when they did not (LA, one-sided Wilcoxon: p = 0.860).[23] A nonparametric comparison of the overall output levels of EA with those in the HA and LA treatments indicates that EA fares about equally well compared to LA (one-sided Wilcoxon: p = 0.145), but substantially exceeds HA (one-sided Wilcoxon: p = 0.008). Models 1 and 2 of the regression analysis presented in Table 3 seem to support these findings. However, extending these regression models by controlling for heterogeneity of treatments over time (interaction effects) and individual characteristics of subjects (Table 3: model 4) indicates significant differences in levels, also between LA and EA (F-test: p = 0.020). Hence, against H1, and in line with the impression that is conveyed by Figure 2, we find statistically significant support for $x^{EA} > x^{LA} > x^{HA}$.

[22] Histograms of real effort provision by treatment and by treatment and working day can be found in Figures A.1 and A.2, respectively (appendix A.2.2).

[23] 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. We support these non-parametric tests by regression analysis in Table 3 (and Tables A.2 – A.6 in the Appendix).

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.

Result 2: In partial support of H2, as the experiment progresses, cooperation unravels faster under HA than under LA and EA.

As shown in Figure 2, at the outset, the three treatments do not differ in terms of exerted effort (Wilcoxon for both the first and the first three working days: all $p > 0.1$). However, there is variation in terms of how fast cooperation unravels over time. Whereas the effort declines at a rate of about 0.9 tasks per working day in HA (coefficient of variable "working day" in models 3 and 4 in Table 3), the decline in LA and EA is only about half that size (adding the coefficients of interaction terms of working day with LA and EA, respectively), and significantly different from HA (z-test, both: $p < 0.05$). The decline of cooperation is not significantly different between LA and EA both with and without controlling for individual characteristics of subjects (models 3 and 4 in Table 3, z-tests: both $p > 0.7$).

Table 3: Tobit regressions on real effort with group level error clustering

	(1)	(2)	(3)	(4)
Low accountability	1.665	1.721	-0.248	-1.451
	(1.321)	(1.348)	(1.350)	(1.162)
Endogenous accountability [a)]	3.420***	3.489***	1.216	2.065
	(1.231)	(1.253)	(1.477)	(1.374)
Working day		-0.540***	-0.853***	-0.934***
		(0.0821)	(0.155)	(0.154)
Low accountability x working day			0.417**	0.432**
			(0.205)	(0.219)
Endog. accountability x working day			0.477**	0.509**
			(0.201)	(0.214)
Additional controls	No	No	No	Yes
Observations	1620	1620	1620	1620
Left/right censored obs.	1055/512	1055/512	1055/512	1055/512
Log. Likelihood	-1363.7	-1345.5	-1342.8	-1217.8
F-test	2.901	5.552	3.705	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. The results of a Tobit model with errors clustered at the subject level (Table A.2) as well as a panel Tobit (panel variable: subjects, Table A.3), and of mixed effect Logit (Table A.4), Probit (Table A.5) and linear models (Table A.6), each with multi-level error clustering: subjects nested in groups, do not differ qualitatively from the results presented above and can be found in Appendix A.2.1. Additional controls include: gender, age, 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 A.4 for details on all measures); levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a)] Whereas the coefficients of the treatments dummies for LA and EA are not significantly different from one another in models 1-3 (all $p \geq 0.141$, F-tests), by controlling for individual heterogeneity in subjects, model 4 yields statistically significant differences between LA and EA: F-test, $p = 0.019$.

Result 3: Against H3, in the treatment 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 learned 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).[24] Although it is possible that the cost of 1 penny was perceived as economically substantial, we believe this is

[24] 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. In an experimental prisoners' dilemma setting Jarke et al. (2017) obtain similar results: they compare zero monitoring costs with costs of 10% and 16% of the maximum defection payoff, respectively, and find a strong drop from 80% to 10% and 4%, respectively. In contrast, Tasch and Houser (2018) find very frequent information look-up across information cost conditions of partially substantial magnitude (see section 2.4).

implausible, as 1 penny is e.g. much smaller than the cut-off point of 12.5 pennies that rational selfish pay-off maximizers should use to decide whether to work fully or not at all (see section 2.2). We interpret this result as evidence that workers were not interested in learning the true source for the recorded number for some of the reasons we will come back to in the discussion.

3.2. Supplementary Evidence

We elicited beliefs of all workers about their average co-worker's effort in the pre-payment questionnaire.[25] 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: $\rho = 0.338$, $p < 0.001$).[26] 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 $p > 0.6$).[27] 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 $p > 0.7$), subjects in EA 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: $p = 0.015$). Provided that subjects in EA almost never looked up the optional information (only in 1.3% of all cases) makes the comparison to LA particularly interesting. Subjects in EA might have anticipated the potential working of the 'Damocles effect' in their team members as well. Not only one anticipates that the others might observe one's real cooperation level, but so do the others. In turn, this might have resulted in less pessimistic beliefs, and eventually, higher exertion of effort.[28]

[25] Although, from a statistical point of view, it would be advantageous to elicit beliefs about others' contributions every working day, we refrained from doing so, in an attempt not to distort subjects' effort decisions by highlighting the role of the others' behavior.

[26] Since in our information rich setting, 50% of all recorded (and therefore displayed) numbers were actual effort levels in each of the three treatments, it is not surprising that subjects had a good understanding about their peers' levels of exerted effort. The distance between beliefs and actual effort levels is not significantly different across treatments (two-sided Wilcoxon: all $p > 0.236$).

[27] 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.

[28] Besides the notion that others could potentially learn of one's own contribution behavior, it would also be interesting to control for the subjects' beliefs about how many of the team members will actually choose to the optional information. Whereas in this experiment we did not elicit such beliefs, we do so in a companion experiment.

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). In total, 83% of all participants filled the final questionnaire. By not filling this questionnaire and not collecting their earnings, on average £3.85 were left on the table. Table 4 contains average values from 49, 46 and 54 individuals who filled the final questionnaire in the treatments LA, HA and EA, respectively.

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

Our experiment varied the degree to which cooperation can be directly traced back to a participant's individual effort level or a random draw. By running our experiment over three weeks online, 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 (Zizzo, 2010).

In line with the stylized finding of the vast literature of public goods games, and lacking an explicit punishment technology (e.g. Fischbacher and Gächter, 2010), we also find diminishing contributions over time, across all treatment conditions. However, the decline in contributions is by far the steepest under high accountability, where information about the others' true source of output is 'thrown into the faces' of all participants whether they wanted to have it or not. That is, although accountability, and in a broader sense transparency, is normally seen as a good thing (Baker, 2000; Seabright, 1996) or might even be required institutionally, 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, and, if in a different setting, with Frey (1993), who finds that excess supervision may backfire, in our context by fueling distrust among workers. 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). Interestingly, Corgnet et al. (2015) only find a positive effect of monitoring if subjects are explicitly informed about when they are monitored by their peers. In contrast, in our experiment, lacking this element of signalling, the notion that one's co-workers could potentially find out whether one shirked or not seemed to be enough to slow down the speed of unravelling cooperation.

Under accountability one step removed, team members 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 a small effort to get it (e.g. searching the internet). Accountability one step removed was sufficient to do at least as well as under low accountability. Nevertheless, 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 would be an interesting finding (we even find that EA fares significantly better than LA when controlling for individual characteristics).

The question remains why providing a notion of optional accountability performed so well. Observability can in principle support contribution in both the HA and the EA treatments. If there is such an effect, though, it is not evident by comparing initial contributions in the HA and LA treatment, before any unraveling takes place. If it exists, it is at any rate dominated by the strong unraveling taking place in the HA treatment as time goes by. That said, we cannot rule out that a 'Damocles effect' of potentially being observed shirking by others might be the reason for the marginally better performance of the EA treatment relative to the LA treatment.[29] However, this difference is only significant after controlling for additional subject specific characteristics.

4.1. Why do subjects not seek information in EA?

Only a minority of subjects actually learned the true source of their co-members' recorded numbers. In line with the psychological literature on self-perception and self-esteem (Bem,

[29] There was the same 50% change of a random number overwriting actual effort levels in the LA treatment, but no-one ever knew the true source of the recorded number, such that no 'Damocles effect' could exist in this treatment.

1972), subjects may 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. Subjects may also have pessimistic beliefs about their co-subjects' exerted effort 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. This would be in line with Huck et al. (2017) who found that 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).[30] It is also possible that subjects may simply be inclined to stick to the default of, in our case, not getting the information (e.g. Choi et al., 2003; Madrian and Shea, 2001). This potential mechanism is investigated in a companion paper.

A further mechanism for information avoidance might be that subjects foresee that having information could lead to cooperation unravelling, and therefore they withdraw from having it. That said, in contrast to the literature on information avoidance (see Golman et al., 2017 for a review) where each subject could put itself into a better position individually by not seeking information, even anticipating that an LA information scheme would fare better than an HA information scheme would create a social dilemma situation in terms of voluntary information look-up in EA. The 'benefit' of information avoidance is there only if no one (or only very few subjects) would look up the info and thereby could be disappointed by the low contributions of others, which in turn, could limit an overall negative effect on effort. Specifically, our subjects would face a social dilemma situation between having info (benefiting from instrumental curiosity, as per Tasch and Houser, 2018, discussed in section 2) and not having info (avoid cooperation unravelling). It seems plausible that, given the evidence of subjects' instrumental curiosity and given standard experimental evidence on free riding, subjects should have sought the info more frequently than actually observed in our data, if this were the key mechanism explaining information avoidance.

That said, we cannot entirely rule out that subjects deliberately gave up the chance of having the info, in the belief that, if everyone did so, the overall outcome would be beneficial to each

[30] Subjects in Huck et al. (2017) 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. Also see Golman et al. (2017) for a comprehensive review on information avoidance.

member of the team. Since we aimed for minimizing experimenter demand, we did not record production beliefs on each working day. However, this could be an interesting avenue for future research.

4.2. Why do subjects exert more effort under EA than under LA and HA?

Not getting information in EA, from a purely informational point of view, puts subjects in a similar environment as in treatment 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. Since getting the information does not increase one's own accountability, but the accountability of one's co-workers, there exists a latent 'Damocles effect' of potentially being observed, even when not learning the information oneself. In this environment of informational uncertainty about whether one's co-workers do or do not get the information may induce subjects in EA to exert more effort, than subjects whose co-workers never had the chance to learn about the true sources of their co-workers' recorded numbers (LA treatment). We do not make any hypothesis on whether this 'Damocles effect' is weaker or stronger in the EA treatment relative to the HA treatment. On the one hand, it may be weaker because in the HA treatment the likelihood of learning the real number of completed tasks is 50% and in EA this probability can at most be 50% but never higher. That is, the expected chance of getting caught working very little is higher in HA than in EA. On the other hand, subjects in the EA treatment may believe that those who choose to acquire the information will be more sensitive to it than subjects in HA who always have the information. Thus, the danger of 'punishment' in terms of negatively reciprocating by lowering their contributions may be greater in EA than in HA.[31]

Furthermore, the very fact that subjects had the opportunity to get information about the other subjects' true source of their recorded numbers might have highlighted that other subjects could make use of this information strategically, e.g. by judging intentions, and consequently reciprocating to them. Recent experimental evidence shows that intentions, in addition to outcomes, are an important factor determining subjects' behavior (Bartling and Fischbacher, 2012; Rand et al., 2015). Taking intentions into account and since information is made salient

[31] Subjects might be more sensitive to information they voluntarily acquired and paid for, because of sunk cost effects (Thaler, 1980) or because they attempted to avoid cognitive dissonance (Festinger, 1957) and act in a self-consistent manner (Rustichini and Villeval, 2014).

in EA, but not in LA, subjects in EA should be more careful in not getting caught free riding, which could explain higher effort levels than in LA in which no one can be caught at all.

An alternative explanation as to why subjects exerted higher levels of effort in EA is in line with research in participatory management (e.g. Lawler and Hackman, 1969). It would be that people appreciate to have more freedom of choice (in our case whether or not to get the optional information) per se. Such a mechanism could explain why subjects exerted the highest levels of effort in the endogenous accountability treatment, irrespective of whether they actually had the optional information or not. In line with such an interpretation, there is also growing evidence from research on endogenous institutions. People seem to value the fact of having a say, independent of - and in addition to - the allocative quality of the outcome that is achieved by allowing for freedom of choice. For example, letting subjects vote on the type and intensity of reward or sanctioning schemes to encourage high contributions to a public good results in higher contributions than imposing the endogenously selected policy exogenously (Markussen et al., 2014; Sutter et al., 2010; Tyran and Feld, 2006). Moreover, receiving the optional information as a result of individual choice might be perceived as procedurally fair (Tyler, 1988) and might increase the legitimacy of the information scheme (Castillo, 2011), i.e., having or not having the info. Dal Bó et al. (2010) used an elaborated experimental design to control for selection effects and find causal evidence of an “endogeneity premium”. In their setting, endogenously chosen policies resulted in more efficient outcomes than imposing the same policies exogenously did. Nevertheless, more research is clearly needed to substantiate support for a potential endogeneity mechanism in team production environments – or preferably – test it directly in an experiment.

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. Nonetheless, this is an obvious avenue of extension of this research.

Acknowledgements

The financial support of the ESRC (NIBS Grant ES/K002201/1) is gratefully acknowledged. We would like to thank the Associate Editor and two anonymous reviewers for their valuable comments and suggestions. 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, the annual conference of the German Economic Association, Vienna, Austria and the 86th Annual Meeting of the Southern Economic Association, Washington, DC. 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

Almas, I., Cappelen, A.W., Sorensen, E.O., Tungodden, B., 2010. Fairness and the Development of Inequality Acceptance. Science (80-.). 328, 1176–1178.

Andreoni, J., 1988. Why free ride? J. Public Econ. 37, 291–304. doi:10.1016/0047-2727(88)90043-6

Andreoni, J., Petrie, R., 2004. Public goods experiments without confidentiality: A glimpse into fund-raising. J. Public Econ. 88, 1605–1623. doi:10.1016/S0047-2727(03)00040-9

Aoyagi, M., Fréchette, G., 2009. Collusion as public monitoring becomes noisy: Experimental evidence. J. Econ. Theory 144, 1135–1165. doi:10.1016/j.jet.2008.10.005

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:10.1162/0033553054268192

Bartling, B., Fischbacher, U., 2012. Shifting the Blame: On Delegation and Responsibility. Rev. Econ. Stud. 79, 67–87. doi:10.1093/restud/rdr023

Bem, D.J., 1972. Self-Perception Theory. Adv. Exp. Soc. Psychol. 6, 1–62. doi:10.1016/S0065-2601(08)60024-6

Bigoni, M., Suetens, S., 2012. Feedback and dynamics in public good experiments. J. Econ. Behav. Organ. 82, 86–95. doi:10.1016/j.jebo.2011.12.013

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., Konow, J., Sørensen, E.Ø., Tungodden, B., 2013a. Just Luck: An Experimental Study of Risk Taking and Fairness. Am. Econ. Rev. 103, 1398–1413. doi:10.1257/aer.103.4.1398

Cappelen, A.W., Moene, K.O., Sørensen, E.Ø., Tungodden, B., 2013b. Needs Versus Entitlements-an International Fairness Experiment. J. Eur. Econ. Assoc. 11, 574–598. doi:10.1111/jeea.12000

Cappelen, A.W., Sørensen, E., Tungodden, B., 2010. Responsibility for what? Fairness and individual responsibility. Eur. Econ. Rev. 54, 429–441. doi:10.1016/j.euroecorev.2009.08.005

Carpenter, J., 2004. When in Rome: conformity and the provision of public goods. J. Socio. Econ. 33, 395–408. doi:10.1016/j.socec.2004.04.009

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

Castillo, J.C., 2011. The legitimacy of economic inequality: An empirical approach to the case of Chile. Universal-Publishers.

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

Corgnet, B., Hernán-González, R., Rassenti, S., 2015. Peer Pressure and Moral Hazard in Teams: Experimental Evidence. Rev. Behav. Econ. 2, 379–403. doi:10.1561/105.00000040

Corgnet, B., Hernán-González, R., Schniter, E., 2014. 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

Cox, C., Stoddard, B., 2015. Framing and Feedback in Social Dilemmas with Partners and Strangers. Games 6, 394–412. doi:10.3390/g6040394

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

Dal Bó, P., Foster, A., Putterman, L., 2010. Institutions and Behavior: Experimental Evidence on the Effects of Democracy. Am. Econ. Rev. 100, 2205–2229. doi:10.1257/aer.100.5.2205

Dijk, F. Van, Sonnemans, J., Winden, F. Van, 2001. Incentive systems in a real effort experiment. Eur. Econ. Rev. 45, 187–214.

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. 2014-005.

Eckel, C.C., Petrie, R., 2011. Face Value. Am. Econ. Rev. 101, 1497–1513. doi:10.1257/aer.101.4.1497

Erkal, N., Gangadharan, L., Nikiforakis, N., 2011. Relative Earnings and Giving in a Real-Effort Experiment. Am. Econ. Rev. 101, 3330–3348.

Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54, 293–315. doi:10.1016/j.geb.2005.03.001

Falk, A., Ichino, A., 2006. Clean Evidence on Peer Effects. J. Labor Econ. 24, 39–57. doi:10.1086/497818

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. Quaterly J. Econ. 114, 817–868.

Festinger, L., 1957. A theory of cognitive dissonance. Row, Peterson, Evanston, Il.

Fiala, L., Suetens, S., 2017. Transparency and cooperation in repeated dilemma games: a meta study. Exp. Econ. doi:10.1007/s10683-017-9517-4

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

Frey, B.S., 1993. Does monitoring increase work effort? The rivalry with trust and loyalty. Econ. Inq. 31, 663–670. doi:10.1111/j.1465-7295.1993.tb00897.x

Golman, R., Hagmann, D., Loewenstein, G., 2017. Information Avoidance. J. Econ. Lit. 55, 96–135. doi:10.1257/jel.20151245

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., 2017. More effort with less pay: On information avoidance, optimistic beliefs, and performance. WZB Discuss. Pap.

Jarke, J., Gsottbauer, E., Engel, S., Goeschl, T., 2017. Peer monitoring and cooperation in repeated social dilemmas. Unpubl. Work. Pap.

Johnson, E.J., Goldstein, D.G., 2003. Do Defaults Save Lifes? 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.

Lawler, E.E., Hackman, J.R., 1969. Impact of employee participation in the development of pay incentive plans: A field experiment. J. Appl. Psychol. 53, 467–471. doi:10.1037/h0028657

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.

Markussen, T., Putterman, L., Tyran, J.-R., 2014. Self-organization for collective action: An experimental study of voting on sanction regimes. Rev. Econ. Stud. 81, 301–324. doi:10.1093/restud/rdt022

Mas, A., Moretti, E., 2009. Peers at Work. Am. Econ. Rev. 99, 112–145.

Miyagawa, E., Miyahara, Y., Sekiguchi, T., 2008. The folk theorem for repeated games with observation costs. J. Econ. Theory 139, 192–221. doi:10.1016/j.jet.2007.04.001

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

Nicklisch, A., Grechenig, K., Thöni, C., 2016. Information-sensitive Leviathans. J. Public Econ. 144, 1–13. doi:10.1016/j.jpubeco.2016.09.008

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

Rand, D.G., Fudenberg, D., Dreber, A., 2015. It's the thought that counts: The role of intentions in noisy repeated games. J. Econ. Behav. Organ. 116, 481–499. doi:10.1016/j.jebo.2015.05.013

Rustagi, D., Engel, S., Kosfeld, M., 2010. Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management. Science (80-.). 330, 961–965.

Rustichini, A., Villeval, M.C., 2014. Moral hypocrisy, power and social preferences. J. Econ. Behav. Organ. 107, 10–24. doi:10.1016/j.jebo.2014.08.002

Samak, A., Sheremeta, R., 2013. Visibility of Contributors and Cost of Information: An Experiment on Public Goods. MPRA Work. Pap.

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., John, D., Zizzo, D.J., 2014. Inattentive consumers in markets for services. Theory Decis. 79, 307–332. doi:10.1007/s11238-014-9466-8

Sonntag, A., Zizzo, D.J., 2015. Institutional Authority and Collusion. Sounthern Econ. J. 82, 13–37. doi:10.2139/ssrn.2387743

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

Sutter, M., Haigner, S., Kocher, M.G., 2010. Choosing the Carrot or the Stick Endogenous Institutional Choic in Social Dilemma Situations. Rev. Econ. Stud. 77, 1540–1566. doi:10.1111/j.l467-937X.2010.00608.x

Tasch, W., Houser, D., 2018. Social preferences and social curiosity. Georg. Mason Univ. Work. Pap.

Taylor, S., Brown, J.D., 1988. Illusion and well-being: A social psychological perspective on mental health. Psychol. Bull. 103, 193–210.

Thaler, R.H., 1980. Toward a positive theory of consumer choice. J. Econ. Behav. Organ. 1, 39–60.

Tyler, T., 1988. What is Procedural Justice?: Criteria used by Citizens to Assess the Fairness of Legal Procedures. Law Soc. Rev. 22, 103–136.

Tyran, J.-R., Feld, L.P., 2006. Achieving Compliance when Legal Sanctions are Non-deterrent. Scand. J. Econ. 108, 135–156. doi:10.1111/j.1467-9442.2006.00444.x

van der Heijden, E., Moxnes, E., 1999. Information Feedback in Public-Bad Games: A Cross-Country Experiment. Cent. Discuss. Pap. 102.

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

Weimann, J., 1994. Individual behaviour experiment in a free riding. J. Public Econ. 54, 185–200. doi:/10.1016/0047-2727(94)90059-0

Yamamoto, Y., 2007. Efficiency results in N player games with imperfect private monitoring. J. Econ. Theory 135, 382–413. doi:10.1016/j.jet.2006.05.003

Zizzo, D.J., 2010. Experimenter demand effects in economic experiments. Exp. Econ. 13, 75–98.

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.[32] 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≤0.1148	20
c=0.125	indiff.	c=0.0625	indiff.				
		0.0625<c≤0.1923	match exp.	0.128≤c<0.2171	match exp.	0.1148<c≤0.1819	match exp.
c>0.125	0	c>0.1923	0	c≥0.2171	0	c>0.1819	0

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.

32 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*.

A.2 Additional results

A.2.1 Regressions

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 higher 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. Nevertheless, Table 3 and Table A.2 show a qualitatively very similar picture. Tables A.3 to A.6 demonstrate that the results presented in Table 3 are robust to a variety of different regression models.

Table A.2: Tobit regressions on real effort with subject level error clustering

	(1)	(2)	(3)	(4)
Low accountability	1.665	1.721	-0.248	-1.451
	(1.185)	(1.206)	(1.389)	(1.345)
Endogenous accountability	3.420***	3.489***	1.216	2.065
	(1.210)	(1.229)	(1.431)	(1.412)
Working day		-0.540***	-0.853***	-0.934***
		(0.0763)	(0.170)	(0.177)
Low accountability x working day			0.417**	0.432*
			(0.212)	(0.229)
Endog. accountability x working day			0.477**	0.509**
			(0.204)	(0.219)
Additional controls	No	No	No	Yes
Observations	1620	1620	1620	1620
Log. Likelihood	-1363.7	-1345.5	-1342.8	-1217.8
F-test	3.176	5.915	3.573	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: gender, age, 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; levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01

Table A.3: Panel Tobit regressions on real effort (panel variable: subjects)

	(1)	(2)	(3)	(4)
Low accountability	1.669	1.677	-0.0789	-1.225
	(1.163)	(1.177)	(1.398)	(1.389)
Endogenous accountability	3.322***	3.345***	1.283	1.963
	(1.157)	(1.171)	(1.391)	(1.389)
Working day		-0.548***	-0.851***	-0.929***
		(0.0703)	(0.132)	(0.144)
Low accountability x working day			0.398**	0.424**
			(0.172)	(0.189)
Endog. accountability x working day			0.458***	0.489***
			(0.169)	(0.186)
Additional controls	No	No	No	Yes
Observations	1620	1620	1620	1620
Left/right censored obs.	1055/512	1055/512	1055/512	1055/512
Log. Likelihood	-1123.4	-1089.1	-1084.9	-1055.4
F-test	7.239	32.81	33.80	40.97

Notes: Reference treatment in all columns: high accountability; all columns contain marginal effects of Panel Tobit models with standard errors in parentheses. Additional controls include: gender, age, 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; levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Mixed effect Logit regressions on real effort (subjects nested in groups)

	(1)	(2)	(3)	(4)
Low accountability	0.104	0.105	-0.0224	-0.0984
	(0.0762)	(0.0762)	(0.0901)	(0.0845)
Endogenous accountability	0.216***	0.219***	0.0832	0.113
	(0.0787)	(0.0795)	(0.0919)	(0.0858)
Working day		-0.0352***	-0.0552***	-0.0549***
		(0.00623)	(0.0109)	(0.0104)
Low accountability x working day			0.0286**	0.0279**
			(0.0121)	(0.0119)
Endog. accountability x working day			0.0296**	0.0296**
			(0.0122)	(0.0120)
Additional controls	No	No	No	Yes
Observations	1567	1567	1567	1567
Log. Likelihood	-740.85	-710.75	-706.71	-676.23
Chi-squared	8.06	61.29	64.65	110.05

Notes: Reference treatment in all columns: high accountability; all columns contain marginal effects of hierarchical Logit models with standard errors in parentheses (random effects of subjects nested in groups). Additional controls include: gender, age, 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; levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Mixed effect Probit regressions on real effort (subjects nested in groups)

	(1)	(2)	(3)	(4)
Low accountability	0.110	0.110	-0.0128	-0.0910
	(0.0809)	(0.0814)	(0.0967)	(0.0880)
Endogenous accountability	0.225***	0.229***	0.0896	0.124
	(0.0824)	(0.0834)	(0.0982)	(0.0895)
Working day		-0.0373***	-0.0575***	-0.0567***
		(0.00589)	(0.0105)	(0.00995)
Low accountability x working day			0.0276**	0.0264**
			(0.0123)	(0.0122)
Endog. accountability x working day			0.0305**	0.0297**
			(0.0124)	(0.0122)
Additional controls	No	No	No	Yes
Observations	1567	1567	1567	1567
Log. Likelihood	-741.57	-710.51	-706.73	-676.43
Chi-squared	7.7	64.49	68.58	116.41

Notes: Reference treatment in all columns: high accountability; all columns contain marginal effects of hierarchical Probit models with standard errors in parentheses (random effects of subjects nested in groups). Additional controls include: gender, age, 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; levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01

Table A.6: Mixed effect linear regressions on real effort (subjects nested in groups)

	(1)	(2)	(3)	(4)
Low accountability	1.526	1.526	0.0468	-1.175
	(1.214)	(1.214)	(1.449)	(1.296)
Endogenous accountability	3.256***	3.256***	1.543	1.951
	(1.214)	(1.214)	(1.449)	(1.299)
Working day		-0.525***	-0.738***	-0.738***
		(0.0647)	(0.112)	(0.112)
Low accountability x working day			0.296*	0.296*
			(0.158)	(0.158)
Endog. accountability x working day			0.343**	0.343**
			(0.158)	(0.158)
Additional controls	No	No	No	Yes
Observations	1620	1620	1620	1620
Log. Likelihood	-5610.1	-5577.9	-5575.2	-5546.8
F-test	7.199	73.05	78.81	148.7

Notes: Reference treatment in all columns: high accountability; all columns contain coefficients of hierarchical linear models with standard errors in parentheses (random effects of subjects nested in groups). Additional controls include: gender, age, 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; levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01

A.2.2 Figures

Figures A.1 and A.2 display the distribution of work effort exerted by the subjects in aggregate and separated per working day, respectively. Both figures indicate a binary distribution of efforts. One explanation for such an observation could be that it was not working on the task itself that caused too much effort, but rather 'getting back to work'.

On the one hand, the observed effort patterns suggests that the majority of subjects who engaged in the task might have perceived the subjective costs of solving a task as less costly than 12.5 pence – the theoretical threshold for either completing all 20 tasks of a working day or none (see section 2.2). On the other hand, the subjects' subjective costs of 'returning to work' (consciously or unconsciously) could have been much higher. We do not claim any specific mechanism as to why this might be the case, but it could well be that self-control, grit or conscientiousness play an important role with respect to the costs of repeatedly returning to work. Subjects with higher self-control might have found it easier (or less costly) to return to work on a regular basis than those with lower self-control. Since we did not elicit scales related to self-control, grit or conscientiousness, we cannot establish this association in our data. However, doing so would be an interesting avenue for future research, because such personal characteristics could also be important determinants in real team production environments, such as in the workplace.

Figure A.1: Distribution of exerted effort

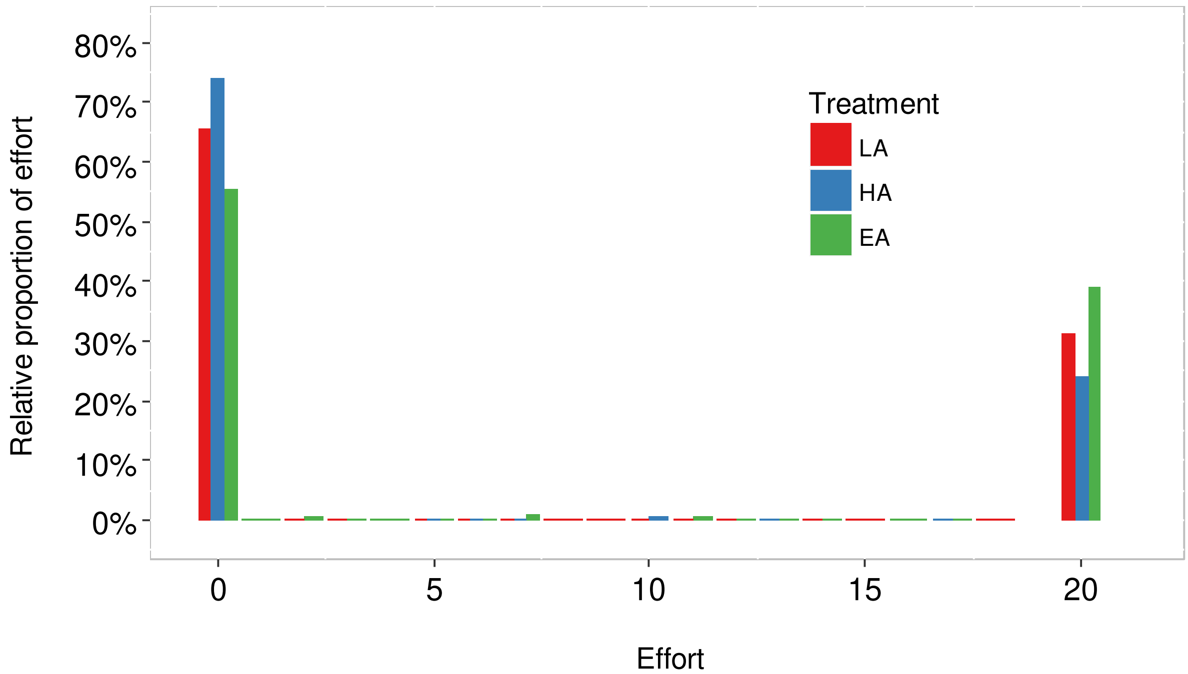
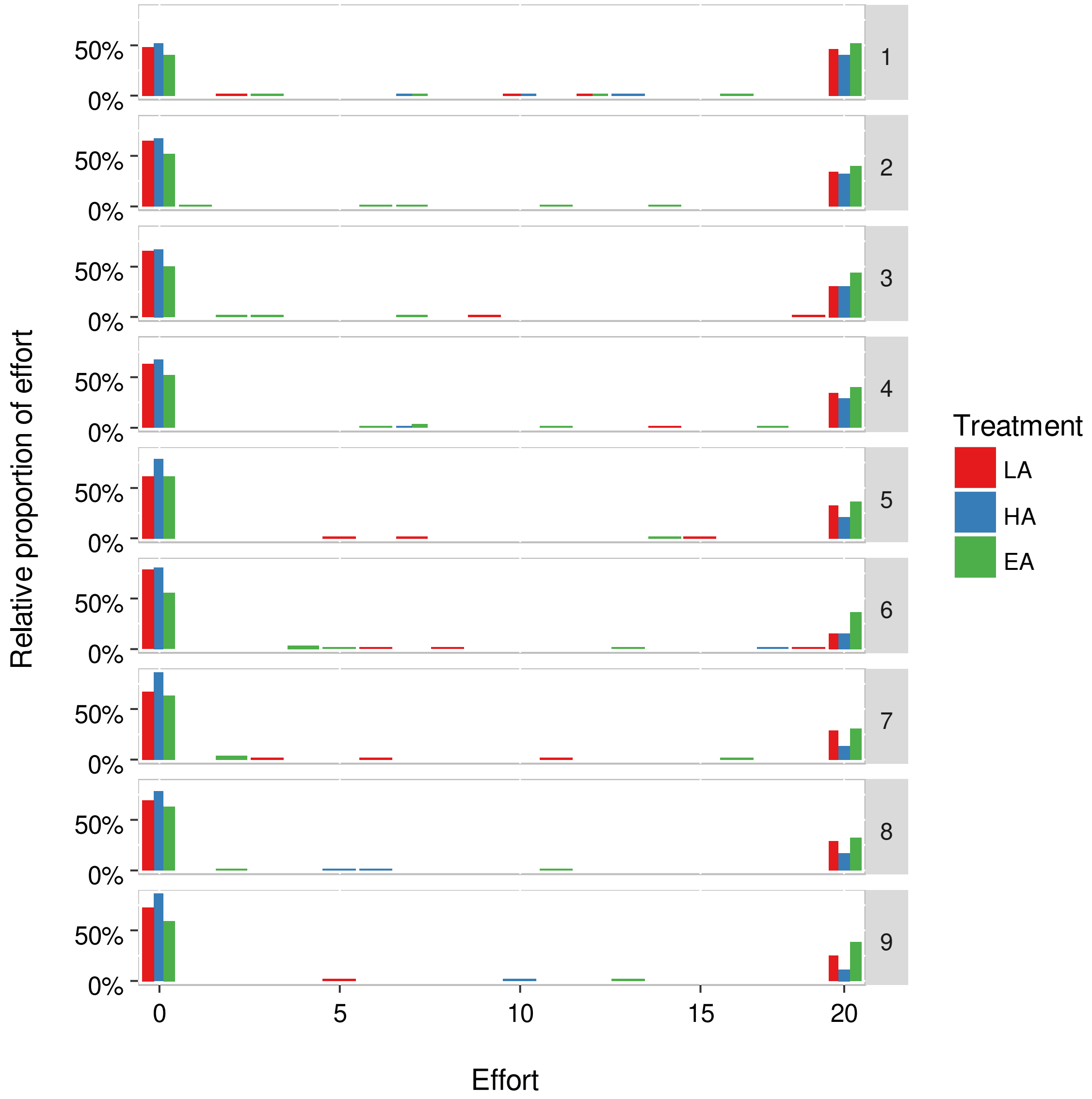


Figure A.2: Distribution of exerted effort, by working day



A.3 Experimental Instructions

A.3.1 Lab-instructions part 1[33]

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

[33] Part 1 instructions were identical across treatments.

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.

A.3.2 Lab-instructions part 2[34]

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.

[34] Part 2 instructions varied across treatments. Sections in [[…]], {{…}} and ‖…‖ were exclusively used in the treatments HA, LA and EA, respectively.

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.

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 x £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 x £1 – £0.01= £9.99. Every member of this group who did not learn this information would earn 10 x £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?

Page 6

[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).

Page 1

- 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".

Page 2

- 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]

Page 3

- 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

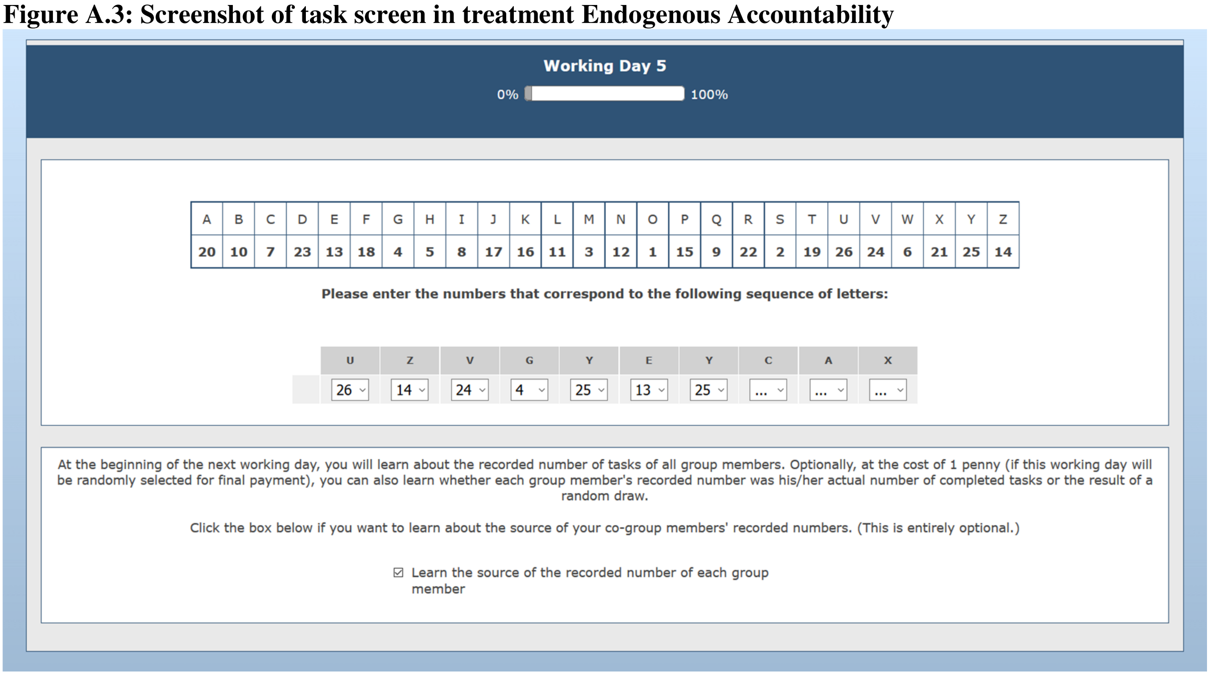
Thank you for filling this final questionnaire!

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.3: Screenshot of task screen in treatment Endogenous Accountability



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.