**Title**
Three Essays on Non-Monetary Incentives and Employee Compensation



**Author**
Ramirez, Cristian

**Publication Date**
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

# Three Essays on Non-Monetary Incentives and Employee Compensation

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Cristian Ramirez

2017

ABSTRACT OF THE DISSERTATION

Three Essays on Non-Monetary Incentives
and Employee Compensation

by

Cristian Ramirez
Doctor of Philosophy in Management
University of California, Los Angeles, 2017
Professor Ian Israel Larkin, Chair

Scholars in strategy, organizational behavior, and economics have shown increasing interest in the link between non-monetary, extrinsic incentives and employee productivity. However, nearly all of this research examines rewards that have some kind of social recognition mechanism. In the first chapter of my dissertation, I examine the awarding of private, non-monetary badges for hitting performance targets. On Amazon Mechanical Turk, workers receiving this type of badge upon hitting a performance threshold are approximately 9.4 percent more productive than workers in the control group. Interestingly, this increase in productivity was almost the exact same as giving workers hitting the threshold a 20 percent bonus in pay.

The second chapter of my dissertation presents the analysis of an actual incentive scheme that has a unique characteristic: it combines both symbolic and pecuniary incentives under the same platform. By examining the results of this real-life application, I can estimate the extent to which workers respond to an actual application of gamification and how its impact varies over time.

Understanding the determinants of value captured by different stakeholders is a key issue for both practitioners and scholars in strategic management. The final chapter of my dissertation presents a study on variations in worker compensation in the copper mining industry. Our results show that there is a positive effect of copper price on workers' compensation, but this effect is moderated by the characteristics of labor regulation in each country.

The dissertation of Cristian Ramirez is approved.

Marvin Lieberman

Sanford DeVoe

Jana Gallus

Ian Israel Larkin, Committee Chair

University of California, Los Angeles

2017

# Table of Contents

# List of Figures

# List of Tables

# Vita

| | |
|---|---|
| 2008 | Bachelor Degree in Business and Economics, Pontificia Universidad Católica de Chile. |
| 2009 | Master of Science in Management, Pontificia Universidad Católica de Chile (PUC). |
| 2011 | Fulbright Fellowship. |

# CHAPTER 1

# Do workers perform better after a digital badge is awarded? Evidence from Amazon Mechanical Turk

## 1.1 Introduction

Many papers have been written about the importance of incentives. The empirical evidence collected so far supports the idea that monetary incentives increase worker performance (Lazear, 2000; Prendergast, 1999; Shearer, 2004). Nevertheless, firms have not embraced pay-for-performance schemes to the extent predicted by agency theory. The presence of psychological factors such as overconfidence and social comparison costs (Nickerson and Zenger, 2008) makes performance-based compensation a less attractive alternative in reality (Larkin, Pierce, and Gino, 2012).[1]

However, monetary incentives are just one side of the coin. During the last few years, scholars have become increasingly interested in the effects of extrinsic non-monetary incentives on worker performance, with awards receiving the most attention. Although there is no clear definition of what constitutes a corporate award from an academic point of view, an award is colloquially understood as a symbolic reward that comes with some positive feedback and social exposition (Neckermann, Cueni, and Frey, 2014; Gallus and Frey, 2016). This description of what comprises an award includes many different symbolic rewards actually given by companies such as 'employee of the month' and 'best annual performance', and also recognition symbols bestowed by governments and armies (Frey, 2007).

Many of the awards considered in the empirical literature rely on a large amount on the social disclosure of awardees. The public recognition of recipients has been labeled as the

[1] According to a Bloomberg BNA report on Mercer's 2013 Pay for Performance Survey, more than half of the surveyed companies have executed a pay-for-performance scheme, and from those around a 45 percent felt disappointed with their program (Douglas, 2013).

key component of the motivating power of awards (Frey, 2006; Gallus, 2016; Neckermann and Frey, 2013). According to standard economic theory, having a chance to obtain a purely symbolic reward should not make a worker exert more effort unless there are other benefits included; if workers care about social recognition and status, then we can expect them to put more effort to be rewarded with an award that is visible to the members of the organization (Moldovanu, Sela, and Shi, 2007). Public exposure of an award recipient enhances her status and social reputation (Frey, 2007), and helps satisfy her inner desire to feel better than others (Frey, 2006). Therefore, by publicly rewarding recipients, awards explicitly fulfill an individual need for social recognition. In general, the early research on awards and worker performance seems to suggest a positive effect of symbolic rewards on behavior (Gallus, 2016; Gallus and Frey, 2016; Kosfeld and Neckermann, 2011; Restivo and Rijt, 2012). However, since almost all studies include non-monetary rewards with some degree of social disclosure, disentangling the effects of social comparisons on non-social mechanisms becomes an impractical task.

To my knowledge, only one paper has tried to shed some light on this question. Ashraf, Bandiera, and Lee (2014) study the impact of the introduction of awards to health workers in Zambia who are enrolled in a one-year training program. Workers need to attend classes and their performance is measured in term of test scores. The experimental design in Ashraf, Bandiera, and Lee (2014) allows separating the effect of an award (a personalized letter from the program director to the top performers) from the inherent social visibility generated by the public knowledge of who the top scorers are. As in the case of most studies in the literature, Ashraf, Bandiera, and Lee (2014) consider a symbolic reward that is introduced to subjects before a task is performed (Kosfeld and Neckermann, 2011; Lacetera and Macis, 2010; Neckermann and Frey, 2013). The way some of these 'ex-ante' awards are presented to workers resembles the structure of a tournament, where a very reduced number of recipients or winners are publicly recognized with a reward (Frey and Neckermann, 2008; Kosfeld and Neckermann, 2011). These ex-ante awards seem to have an effect on the tails of the distribution of workers, which means that a great portion of subjects are either unmotivated or simply think they do not have the sufficient skills to be bestowed with an award (Kosfeld and Neckermann, 2011). In line with the general findings, Ashraf, Bandiera, and Lee's results show a positive effect of awards on performance.

Although the research to date shows a positive effect of extrinsic social rewards, we are learning that they can carry real costs for firms, similarly to monetary incentives. Charness, Masclet, and Villeval (2014) study whether the desire to achieve a higher status, understanding it as a better position in the performance ranking with no additional pecuniary benefits, could motivate unethical behavior among experimental subjects. Their results show that agents are willing to sacrifice part of their income just to get a better rank position. Similarly, Huberman, Loch, and ONculer (2004) express their concern about how people are willing to devote more resources than what would be optimal to win a contest, whenever winning awards status. Malmendier and Geoffrey (2009) and Borjas and Doran (2013) offer evidence of a negative impact of awards on CEOs and recipients of the Fields medal, respectively.

Gubler, Larkin, and Pierce (2016) show how an actual implementation of an attendance award program can carry considerable costs to an organization, arguing that under a non-monetary incentive system, workers can still engage in detrimental actions to the firm in order to become eligible for the awards. Also importantly, Gubler, Larkin, and Pierce (2016)'s paper shows how the establishment of an extrinsic award system can have a negative effect on intrinsic motivation when rewards are expected or promised, a topic previously considered in the theoretical literature about motivation (Bénabou and Tirole, 2003; Frey and Jegen, 2001; Ryan and Deci, 2000). Gubler, Larkin, and Pierce (2016) estimate the cost on productivity to be around 1.4 percent as a consequence of the introduction of the award program.

There has been little research on whether firms can get some of the benefits of non-monetary extrinsic rewards without the costs if the social dimension is excluded. The present research tries to shed some light on this topic. Particularly, the aim of this paper is to measure the causal effect of digital badges on worker performance. As awards, badges allow people to distinguish themselves from their peers, satisfying the human need for status and social recognition. Nevertheless, badges could also have an effect on an individual's performance by affecting her self-esteem, intrinsic motivation, and/or feeling of competence (Benabou and Tirole, 2002; Bénabou and Tirole, 2003; Stajkovic and Luthans, 1998). Therefore, badges might affect workers' behavior even when recipients are not publicly recognized. Even more, with workers being more familiar with computers, video games, and the Internet, badges do not need to be physically awarded: a digital badge can fulfill the role.

We test the effectiveness of digital badges using a sample of workers from Amazon Mechanical Turk (AMT) who perform a task for 25 minutes in exchange for a fixed payment. Six experimental conditions are considered. Workers in the *badges* condition (four different treatment arms) receive digital badges, while workers in the *money* condition get an extra monetary payment (both the digital badges and the extra payment are unexpected, i.e., workers are not told about their existence). The digital badges considered in this experiment include both absolute performance and relative performance. The effect of relative feedback provision on worker behavior has, in general, found support in the literature (Blanes i Vidal and Nossol, 2011; Charness, Masclet, and Villeval, 2014; Kuhnen and Tymula, 2012), although at least one negative experience has been observed (Ashraf, Bandiera, and Lee, 2014). Workers in the *badges* and *money* conditions receive a written message along with the specific stimulus (the badge or the monetary payment) when it is awarded. In order to isolate the effect of these stimuli from the inherent positive feedback generated by the written message, a last experimental group is considered. Workers in the *feedback* condition observe just base written message that is shown to workers in the other experimental conditions and nothing else. By including this last experimental group, we can quantify the added effect of each stimulus on worker performance.[2]

To actually capture the differences in behavior that occur after the provision of monetary and non-monetary incentives, all stimuli are awarded once workers reach the same point in the task. This way, by keeping all the conditions across experimental groups the same until this point, we can observe not only the differences in performance across groups once they receive the stimuli (between-group comparison), but also control for their own performance before and after the stimuli are given (within-group analysis).

Our results indicate that, in general, the presence feedback messages, monetary rewards, and digital badges increased the performance of male subjects, while the effect on females is for the most part negative. While males increase their output (in terms of quantity) without sacrificing quality, females that received only feedback messages present a lower output in terms of quantity and also quality. The negative quality effect in females is present for all treatment conditions except for females that got additional monetary payments.

The paper is organized as follows. Section 1.2 presents the relevant theoretical backgrounds for this study and the propositions to be tested. Section 1.3 describes the experiment and its implementation while Section 2.5 shows the results of each of the phases of the experiment. Section 2.6 discusses the main findings, the limitations of the present study, and future lines of research. Section 2.7 concludes.

[2]As we will discuss later, three conditions that award badges also mention information about relative ranking of the subject (e.g., top 20 or top 40 percent). That information is not considered in the *feedback* treatment.

## 1.2 Theoretical backgrounds

Even though the term award is generic enough as to include rewards such as badges, we are reserving the latter for symbolic rewards that are not tied to monetary compensation and are given to anyone who fulfills the requirement (i.e., the number of potential recipients is not constrained ex ante). Armies, the Boy Scouts, online forums, and other groups and organizations use badges or medals to denote or highlight a special situation or achievement (Antin and Churchill, 2011; Frey, 2007). Video games are another good example in which badges are used extensively to engage players in certain behaviors.[3] Nowadays, around 42 percent of Americans play video games more than three hours a week, with the average player being thirty-five years old. It would not be hard to argue that a large number of workers in America are already familiar with digital badges and achievements.

### 1.2.1 Gamification

Badges have been mostly studied in the literature on gamification, a term that could be roughly defined as the use of game-like elements such as badges, medals, point systems, and leaderboards outside games (Deterding et al., 2011).[4]

The evidence on the effect of *gamification* applications on subject behavior has been mostly positive, although results vary depending on the context of the implementation and the final users of it (Hamari, Koivisto, and Sarsa, 2014). Studies on the effectiveness of digital badges and achievements have been carried out in topics such as Internet forums (Grant and Betts, 2013), social media websites (Easley and Ghosh, 2013), photo-sharing services (Montola et al., 2009), online newspapers (Jones, Altadonna, and Lindsey, 2012), marketing (Huotari, 2012), and educational platforms (Abramovich, Schunn, and Higashi, 2013; Domínguez et al., 2013; Gibson et al., 2013; Haaranen et al., 2014).

[3] In video games, badges or digital achievements usually have three main components: signifier, completion logic, and rewards. For a good review of these concepts, please see Hamari (2011).

[4] It is important to mention that the objective of gamification is not to transform a given platform into a game, but rather include game mechanics that could help motivate or engage users (Hakulinen, Auvinen, and Korhonen, 2013) However, this approach to gamification is not standard. Mollick and Rothbard (2014) consider gamification as one among many potential managers initiatives with the goal of improving the experience of work. Mollick and Rothbard evaluate the impact of a game imposed by the management in a work context, which deviates with the idea of gamification being the use of game elements without implementing an entire game. .

With respect to badges, Antin and Churchill (2011) discuss the role of badges in social media, identifying five possible roles that badges play in this context. According to the authors, badges and achievements can provide: (i) goals; (ii) instructions about what is possible to do; (iii) reputation signals; (iv) symbols of status; and (v) group identification. Besides these functions, Antin and Churchill recognize that badges can have a non-social component that can motivate some people. Also, Antin and Churchill talk about badges as symbols of status since they also work as a reminder of past performances and how badges can give personal affirmation. Denny (2013) studies the effect of badges that are only visible as personal information. In his setting, students could engage in two activities: the creation of questions for other students to answer, or answering questions created by other students. Subjects were divided into two groups (treatment and control), and the only difference between groups is the access that the treatment condition had to the badge system. Denny's results show that there is no difference in the number of questions authored by each group, but students in the treatment condition answered more questions on average. Landers and Callan (2011) present the results of introducing a gamified platform in an undergraduate class of a North American university. Enrolled students in this class had the chance to respond to some short tests on the platform, and they could gain a rank each time they passed a test (the cutoffs to get a rank increased from thirty percent to one hundred percent of correct answers). Students were able to get a new rank every four days (they could not lose ranks). There were badges associated with each rank, an badges were displayed in the student's profile page and each time the student posted in the forum dedicated to that particular test. The results of this intervention show that almost 400 students enrolled in the platform, out of around 600 who received an invitation to join, and that on average 4.8 ranks per student were awarded. Since there was no 'reward' in terms of credit for the students that took these short quizzes, the fact that so many students enrolled and spent time on the platform answering questions is interpreted by Landers and Callan as evidence of the motivational effect of gamification. Grant and Betts (2013) show how badges can be a motivator by looking at user behavior on the site Stack Overflow. Grant and Betts present evidence that supports the idea that users increase their participation on the site just during the period before a badge is awarded, lowering their participation levels after receiving the badge. Anderson et al. (2013) explore the role and power of badges as incentives, since even though badges are simple instruments, they seem to provoke complex responses from individuals that are not completely understood. Anderson et al.'s model involves the presence of a designer that is in charge of a site in which users make different decisions. The utility function of each user is increasing in the actions she likes and in the number of badges she obtains. In this multi-period model discounting is also incorporated in the traditional fashion of a probability $\delta > 0$ that the user will abandon the system after taking an action. Some of the implications of this model are related to the optimal location of badges in the decision-space. For example, according to their model, it is better to have two badges than a single (larger) badge at any point. Also, with respect to badge placement, badges should be placed almost evenly apart. While not exactly using badges, Gallus (2016) evaluates the impact of digital awards given to new editors in the German Wikipedia. Gallus finds that purely symbolic awards have a positive effect on editor retention, a key dimension to Wikipedia. By focusing her experiment on editors who have just created their accounts, Gallus argue that the results of awards are mostly driven by changes in subjects' private identity.

### 1.2.2 Autonomy, feelings of competence and self-esteem

Although activities can be extrinsically motivated, the degree to which they feel autonomous can make a difference in terms of their performance while carrying out the task. There is evidence for elementary children that the more an activity is externally regulated, the less interest or effort children put into the task (Ryan and Connell, 1989). In general, higher degrees of autonomy in extrinsically motivated tasks are related to positive outcomes such as enjoyment and interest (Ryan and Deci, 2000), performace (Deci and Ryan, 2000), and even psychological health (Deci, Olafsen, and Ryan, 2017). Although workers might not be intrinsically motivated in all the dimensions of their jobs, they might be genuinely motivated by parts of if, and when they do they also show greater performance and wellbeing (Deci, Olafsen, and Ryan, 2017).

If private badges are not promised to workers beforehand, but they are awarded to employees as they work, there is a good chance that badges would improve recipients' self-esteem. The rationale for this relies on a learning effect Bénabou and Tirole (2003); the badge signals the worker that the task was hard and that the employer cares about her performance. In this case, workers should not infer that their behavior was controlled when badges are awarded since they were given unexpectedly, and this is a key condition to avoid potentially detrimental effects on intrinsic motivation (Deci and Ryan, 2000). Theory and empirical evidence support the idea of a positive relationship between self-esteem and worker productivity (Frey and Jegen, 2001; Kuhnen and Tymula, 2012).

Additionally, private badges give workers a way to understand their performance (Bénabou and Tirole, 2003), which also improves their feelings of competence (Ryan and Deci, 2000). Loewenstein and Issacharoff (1994) show that people value things more when they have been received as a consequence of skill rather than chance. Loewenstein and Issacharoff's point is based on the idea that causes and consequences are deeply related in the mind, so thinking of one recalls the other and vice versa. Therefore, if a private badge is awarded to a worker adducing her performance, this should evoke thinking of 'success,' which is something enjoyable that should affect her valuation of the private badge. Confidence in one's abilities and capacities plays an important role in an individual's motivation and work performance (Benabou and Tirole, 2002; Grant and Gino, 2010; Stajkovic and Luthans, 1998).

### 1.2.3 Reciprocity and conditional altruism

Workers might consider private badges as gifts from the employer and reciprocate in response. The empirical evidence with respect to the existence of gift-exchange theory is mixed; while lab experiments support the existence of actual reciprocity, field experiments only present moderate evidence (Dur, 2009). In a series of experiments, Kube, Maréchal, and Puppe (2012) show that workers respond to a non-monetary gift by increasing their productivity, but not to money given in the form of cash. Interestingly, Kube, Maréchal, and Puppe results indicate that if the money comes folded like an origami figure, workers do increase their productivity by 30 percent. Kube, Maréchal, and Puppe argue that workers value the time the employer spent choosing the gift and not simply the monetary value of it. Therefore, to the extent that workers think of private badges as a device the employer put thought into their creation and implementation, workers might have a positive reaction to the inclusion of private symbolic rewards by improving their productivity.

In a similar venue, Dur (2009) discuss the idea of conditional altruism, which can be understood as how the degree workers care about their manager depends on how convinced workers are of their manager's genuine altruism. Conditionally altruistic workers would reciprocate only if they believe their employer honestly cares about them.[5] Although the private dimension of the badges implemented in the present study does not allow us to test directly for this idea (more on this later), it is still a potential explanation for an employee's change in behavior in the presence of badges. Additionally, given the mixed evidence with

[5]In online labor markets such as AMT, that will be discussed in detail in Section 1.3, the exertion of more efforts seems to be the only possible way workers could reciprocate employers Bradler et al. (2014).

respect to gift-exchange theory, Dur calls for a more comprehensive vision of how the theory might work in reality; from the past focus on monetary gifts, we should now study other tools at managers disposal to create relationships with their employees.

### 1.2.4 Emotions

Lastly, it could be argued that besides increasing feelings of competence, boosting self-esteem, or generating a desire for reciprocity, badges might also have an effect on employee's affect and emotions towards work. A subject's affect, which is defined as the combination of that subject's affective experiences at work and her beliefs about the job (Latham, 2012), plays an important role in that subject's performance and behavior (Pekrun and Frese, 1992). Weiss and Cropanzano (1996) examine how events that occur at work can have immediate influence on performance via affective reactions, and how these levels of affect are not fixed and can vary over time. It has been shown that people develop positive affect to a task after they have performed well in it (Latham, 2012). Grandey, Tam, and Brauburger (2002) discuss how the emotion of pride with one's job can be related to positive events that have happened at work, and how employees with a higher sense of achievement —as consequence of their own efforts— might also present a higher motivation (Hackman and Oldham, 1976). In Grandey, Tam, and Brauburger (2002), the most important source of pride for workers came from supervisors, which reinforces the idea that employer-generated feedback can have an important effect on employee's affective state. Therefore, if private digital badges can help elicit emotions of such as pride in employees, then badges can potentially be used as an effective way to communicate important achievements and improve motivation or performance.

### 1.2.5 Propositions

Badges are, essentially, an inexpensive way to award and recognize individuals. However, to the best of our knowledge, there is no empirical test of their effectiveness in a workplace. As Denny (2013) shows, subjects seem to respond to badges even when the social recognition dimension is removed. This is interesting in at least two ways. First, the literature on awards has focused on the use of non-monetary awards as a mean to get social recognition or status, while the private dimension of rewards has been neglected for the most part. Second, if digital badges have proven useful in video games and educational platforms to engage users (Abramovich, Schunn, and Higashi, 2013; Denny, 2013), then digital badges might be useful in workplaces as a form to provide real-time feedback and (private) recognition to employees in a way that deviates from the traditional approach of publicly announcing the recipient of an award. Additionally, as discussed previously, in most studies about awards the symbolic rewards are presented to subjects before the task starts Ashraf, Bandiera, and Lee (2014), Kosfeld and Neckermann (2011), and Lacetera and Macis (2010). This implies that most of the evidence gathered so far does not include discretionary awards that are given unexpectedly (Bénabou and Tirole, 2003). Therefore, a better understanding of the effects of introducing non-public and unexpected badges on worker performance should be interesting for scholars and practitioners as well.

Thus, from the theoretical discussion above, we obtain the following propositions.

**Proposition 1** *A worker's overall performance will improve after receiving an unexpected, private badge as a result of her past performance.*

Work performance can be understood in different ways depending on the dimension of work we are interested. Proposition 1 states that a worker's overall performance will improve after receiving a private badge, which means that the dimensions of work performance that show an improvement after a badge has been bestowed need not be the same dimensions that caused the badge to be awarded in the first place. Let us clarify this idea with a simple example. Let's consider a job that is comprised of a task that involves repeating a process multiple times using different inputs, e.g. typing information from written documents to a database. In this case, a worker's performance could be evaluated under at least two dimensions: quantity of work (number of documents she typed over a specified length of time) and quality of work (number of mistakes she made). Although the idea of providing a worker with a badge that highlights the number of articles she typed in a day would increase the number of articles she types in the future seems reasonable, we should not neglect the possibility that other dimensions of performance in this case —quality— might improve as well. Neckermann, Cueni, and Frey (2014) present empirical evidence of the 'spillover' effects of awards by looking at changes in worker performance after receiving an award for voluntary work activities. On average, recipients of the award show a short-term increase in their performance. Interestingly, the effect on performance is related to areas that are hard to perceive for coworkers and supervisors. Neckermann, Cueni, and Frey argue that the causes behind the observed effect might be intrinsic or affective instead of related to peer effects or an image-motivation mechanism.

When we consider private digital badges, i.e. private badges that are awarded in an online or electronic context, we should not expect all types of workers to react in the same way to them.[6] Deci (1972a) test the effect of feedback for males and females that were either paid according to a piece rate or not paid at all. For males, verbal reinforcement always increased intrinsic motivation (i.e., time they spent working on the task during the free-choice period), while for females that did not get paid the verbal reinforcement implied a decrease of more than 50 percent of the time they dedicated to the task while in the free-choice period (the drop was not statistically significant at the 90 percent of confidence).[7] Niederle and Vesterlund (2007) show that differences exist in the propensity of men and women to select competitive situations. According to their results, these disparities can be explained by the presence of overconfidence (Croson and Gneezy, 2009) and a preference for competition in men. Niederle and Vesterlund also mention that men and women react differently to the revelation of relative feedback, with women more prone to assimilate bad feedback than men and see feedback as a reflection of their self-worth. Additionally, differences in the way women and men interact with technology have been reported, with women more interested in social interactions than men, with the latter driven by the desire for winning and achieving goals (Koivisto and Hamari, 2014). The evidence suggests that males and females react differently to the same stimulus and, by the effect on intrinsic motivation, the final impact on performance might not be the same for both groups.

**Proposition 2** *The effect of unexpected, private badges on performance will be moderated by subject's gender and the type of feedback provided.*

[6] Young workers are probably more familiar with digital badges and achievements, given their greater exposure to video games and other social platforms. Older users of computer technology have reported lower levels of self-efficacy and higher computer anxiety than their younger counterparts (Koivisto and Hamari, 2014). Therefore, younger workers will probably react to digital badges in a more positive way than older workers who could be confused by the use of digital badges on a platform. However, since we run our experiment on AMT —an online platform— and only consider subjects that have more than 500 approved tasks (see Section 1.4.1), we do not expect a moderating effect of age on the impact of digital badges to be present in our sample.

[7] Similar results were obtained in Deci, Cascio, and Krusell (1973) and Deci, Cascio, and Krusell (1975).

## 1.3 Experimental Setup

### 1.3.1 The platform

Experimental subjects were recruited on AMT, which is an online platform where requesters post short jobs or HITs.[8] AMT workers, individuals that signed up for an account on AMT, work on as many HITs as they want to (as long as they qualify for working on those HITs). Requesters post a description of the HIT and an example of it for workers to check before they agree to complete it. The requester must approve the work before the worker is paid. If the requester does not approve it, the worker's rating approval —which measures the percentage of work submitted by the worker that has been approved— gets hurt. Since many HITs are only available for AMT workers with an approval rating above a threshold, AMT workers care about their rating.

The minimum payment a requester can offer for a HIT is \$0 and most HITs offer payments between \$0.01 and \$0.1 (Ipeirotis, 2010). Horton and Chilton (2010) determine that in their sample of AMT workers the median reservation wage is of \$1.38 per hour. Ipeirotis (2010) estimates the average hourly wage of AMT workers around \$4.8 per hour, although he recognizes this estimation is based on an oversimplified version of how workers choose and complete HITs. Among the most common HITs are image classification, data entry, and completion of surveys.

Scholars in diverse fields have evaluated the use of AMT as a cheap and reliable source of recruits for experiments (Mason and Suri, 2012). AMT samples have been characterized as more diverse than traditional non-probabilistic samples, but in general not as representative as some Internet panels or samples obtained through a probabilistic method (Buhrmester, Kwang, and Gosling, 2011; Berinsky, Huber, and Lenz, 2012). Researchers have replicated some well-known experiments using samples from AMT and obtained very similar results to the ones presented in published papers (Berinsky, Huber, and Lenz, 2012; Horton and Chilton, 2010; Paolacci, Chandler, and Ipeirotis, 2010). All these elements reinforce the idea that inferences obtained from AMT samples might be as valid as others acquired through more traditional samples. However, AMT still has its own drawbacks. For example, since AMT is an Internet-based platform, researchers have no control over the environment in which subjects perform the tasks (Buhrmester, Kwang, and Gosling, 2011; Kittur, Chi,

[8] HIT stands for human intelligence task.

and Suh, 2008). Also, as the use of AMT to run experiments becomes widespread across AMT workers, concerns about the external validity of the results, as a consequence of having subjects exposed to similar experiments many times, will become more important (Berinsky, Huber, and Lenz, 2012). Another issue with experiments carried out using Internet subjects is that individuals tend to be less attentive than in a traditional laboratory experiment (Paolacci, Chandler, and Ipeirotis, 2010).

### 1.3.2 Stages of the experiment

The experiment considered the following stages common to all workers:

1. A HIT is posted on AMT including the description of the task and an example of it. The description of the task on AMT also states that the requester is a research company building a database from information published in newspaper articles.

2. Potential workers see the post and the description of the task (they can also click on a link to check an example of how to complete the information for a real article). If a subject accepts the HIT, she is then redirected to a second screen in which a login and password appear. To actually complete the HIT, subjects need to visit an external website and use the login and password provided to have access to the task.

3. Once workers have access to the external website, the instructions of the task are repeated. Figure 1.1 presents a screenshot of the instructions page on the website (the same set of instructions is included in the job posting on AMT). After checking the instructions, all workers are required to enter a nickname before the task begins.

4. As the page with the first article loads, a 25-minute counter starts running. All subjects, irrespective of their condition (see Section 1.3.4), observe the same sequence of articles. We call these initial 25 minutes of work round 1.

5. At the end of the twenty-five minutes, workers are redirected to a webpage that prompts them to click on a button to start the survey.

6. After answering the survey, workers are redirected to another webpage in which a payment code appears. Workers are told to copy this code and paste it on AMT so to process their payment (this procedure is standard for tasks that are performed outside AMT).

7. At the same time, all workers receive an additional job offer that invites them to work for 15 more minutes in exchange of a bonus payment of \$1.50.[9] If a subject accepts, she is redirected to a new webpage and will start working right away (we call this period of additional work round 2). After these additional fifteen minutes of work, or in case they did not accept the offer, the experiment is over.

#### 1.3.3 The task

Subjects are asked to extract the following information from a series of pictures of short articles from the printed version of a national newspaper: title, authors (separated by commas in case there are more than one), and the first two different private companies mentioned in the article. This means that workers have to type information in four different boxes before they can submit their answers (i.e., click on the 'submit' button). Incomplete submissions, which are ones that have one or more empty camps, are not accepted by the website (subjects are prompted to complete all camps before the submission is processed). Once a subject submits an article (i.e., the subject has sent the four answers required for an article), the website prevents the subject from going back to modify her answers or resubmit an article.

All the articles are about business in general and they mention at least one private company. Articles with subtitles, which could have confused workers, were not included. Rules on how to type the information were presented in detail and even an email address was provided so the workers could contact the author of this paper in case they felt something was unclear.[10]

The task required 25 minutes of work and the completion of a five-minute survey in exchange of a fixed payment of \$3.00.[11] The task was designed so there was variance in terms of both completion time and quality across subjects. Our main goal was to create a task in which both ability and effort play a central role in the determination of performance.

[9] We changed this for the last wave of data wave collection (see Section 2.4). Since we are interested in the effect of our set of stimuli on performance, and given that subjects receive the first treatment after the $10^{\text{th}}$ article —see Section 1.3.4— we only offered the additional job offer to subjects that completed at least 11 articles.

[10] Results from the survey that subjects answered at the end of the first 25 minutes of work show that more than 92 percent of the subjects who completed the task think that instructions of the task are either 'extremely' or 'very' clear. We discuss the results from the survey in Section 1.5.4.

[11] Fixed payments for tasks that involve effort have been used in the literature before (one example is Kosfeld and Neckermann (2011)).

Although other studies have used summations (Azmat and Iriberri, 2012; Eriksson, Poulsen, and Villeval, 2009; Niederle and Vesterlund, 2007), or have made subjects count the number of letters in a paragraph (Rosaz and Villeval, 2012), and even defined the task as reading excerpts from a computer manual (Woolley and Fishbach, 2015), we tried to come up with a task that seemed meaningful and that subjects did not think of it as an experiment but instead as something a real firm would ask them to do.[12] For workers that put more effort in the task, we should observe a higher number of articles completed or a better quality (i.e., fewer errors) in the information submitted for each article.

Three qualification requirements were included. First, AMT workers needed to have a HIT approval rate of at least 95 percent, which is a typical requisite of experiments that hire subjects on AMT (Burbano, 2016), in order to accept the HIT. Second, in order to have a pool of experienced workers, we also required subjects to have submitted at least 500 approved HITs. And third, to keep cultural differences at a minimum, only workers that chose the United States as their location when creating their account were able to accept the job posting on AMT.[13]

### 1.3.4 Thresholds and treatments

The experiment considers one control and six experimental conditions. Workers are randomly assigned to one of these groups at the beginning of the tasks (this information is not revealed to them. Subjects do not know they are taking part in an experiment). Workers in the six experimental groups receive different stimuli after a certain number of articles have been completed. Specifically, the thresholds are 10, 15, 20, 25, 30, and 35 completed articles.[14] Workers in each experimental group can potentially observe the same maximum number of stimuli, although each stimulus is different.

The experimental conditions considered in this study are below:

[12] As mentioned above, the job posting on AMT stated that the requester was building a database from information published in newspaper articles.

[13] Four rounds of data collection occurred between September, 2015 and April, 2017 (more on this in Section 2.4). For all the rounds but the first, we included an additional restriction of no participation in previous rounds to avoid having the same subject working multiple times on the task.

[14] We follow the recommendations in Anderson et al. (2013) with respect to both the number of badges and the 'space' between them. Anderson et al. suggest having multiple badges instead of just one of bigger value when the implementer wants to reward the same dimension (in my case, the number of articles submitted).

- **Feedback**: subjects in all treatment conditions observe a positive and identical message along with the specific stimulus associated with each group. In order to disentangle the particular effect of each stimulus from the inherent feedback given to workers via the written messages, we include one group that we call *feedback* that receives the same positives messages as subjects in the other experimental conditions without including a monetary payment or a digital badge.[15] The messages that accompany each of the other stimuli are presented in Table 1.1.

- **Money**: subjects in the *money* condition get a monetary bonus, additional to the fixed payment offered for the task, of $0.20 after each of the thresholds mentioned above (they are alerted of the extra $0.2 after each of the thresholds is reached). This means that subjects in the money condition could earn up to $5.70, or a 27 percent extra over workers in other conditions, assuming they accept the additional job offer.

- **Badges**: Workers in the badges conditions receive digital badges as recognition for their achievements. As in other conditions, a subject in one of the badges conditions can receive a maximum of six stimuli (digital badges in this case). Each digital badge is simple in its design and highlights the reason why it was awarded.[16] After obtaining a digital badge, a small version of it becomes available as an icon at the top-right corner of the website, next to any other badges that have been awarded. A worker can click on any of the small badges to see it in full size. After receiving a badge, the subject needs to click on the 'close' button in order to continue working on the task (this applies to subjects in the *feedback* and *money* conditions as well). Four different types of digital badges are considered in this study.

  - **BA**: subjects in the *BA* condition receive a digital badge that highlights the number of articles submitted so far after each of the thresholds. *BA* badges only mention the threshold that the worker has just surpassed, but do not make reference to any other attribute (such as number of correct responses or even speed of submission). Figure 1.2 shows how a *BA* badge is actually presented

[15]Although a relative test of the effects of additional monetary payments and digital badges on worker performance could be computed just by comparing the behavior of subjects in both groups, the confounding effect of the feedback provided would not let us capture the added impact of monetary payments and badges on worker productivity without including the feedback condition.

[16]The digital badges were constructed following the model presented on this website: `http://blog.inkydeals.com/illustrator-tutorial-premium-vector-badges/`.

(awarded) to a subject after the completion of the 15th article. The same form of presentation was used for all other stimuli in our study (although for *feedback* and *money* conditions no picture or digital icon was presented above or below the text). Figure 1.3 presents a screenshot of the *BA* badges used in this study.

- **BR-20**: subjects in the *BR* condition receive a badge that emphasizes the relative standing of the subject. In this particular case, a *BR-20* badge means that it was communicated to the subject that she was in the "top 20 percent of users who have worked on this HIT." As with the other stimuli considered in this study, these *BR-20* badges were given to all subjects in this condition irrespective of performance. This type of feedback is similar to the one used by Deci (1972a) and Deci (1972b).[17] To keep variation to a minimum, the same colors and iconography of *BA* badges were used to construct the *BR* badges.

- **BR-2040**: subjects in the *BR-2040* condition first received a 'top 20 percent' badge (first threshold), and then a 'top 40 percent' badge thereafter, irrespective of performance.

- **BR-40**: subjects in the *BR-40* condition received a 'top 40 percent' badge after each of the thresholds, irrespective of performance.[18]

A summary of the main characteristics of the conditions considered in this study appears in Table 1.2.

All digital badges in this study are awarded according to the number of articles completed (i.e., according to the thresholds) and not taking into consideration the number of correct responses (as discussed above). Although providing private badges conditional on the quality of the job submitted during the task was an option (e.g., by the accumulated number of

[17] The feedback given to subjects each time they completed a puzzle in Deci (1972a) and Deci (1972b) was something in the lines of "That's very good, it's the fastest anyone has solved this one."(Deci, 1972b, p.224). Clearly, it is impossible that each subject had solved the puzzles faster than the person preceding her.

[18] The inclusion of the BR conditions is motivated by the empirical papers about peer effects and relative feedback. There has been evidence of the positive effect of peer effects under both flat-rate (Mas and Enrico, 2009) and piece-rate environments (Blanes i Vidal and Nossol, 2011). Eriksson, Poulsen, and Villeval (2009) do not find effect of relative feedback on performance under a pay-for-performance scheme in terms of number of subjects' submissions during their task (quantity), but they do find a decrease in the quality of the submissions (correct answers) if constant relative feedback is provided, a result that Eriksson, Poulsen, and Villeval attribute to stress and anxiety (in Eriksson, Poulsen, and Villeval's paper the feedback provided to each subject is relative to a specific individual to whom she is paired and not relative to the subject's position in the overall performance distribution).

articles correctly completed), we opted not to implement them for two main reasons. First, by making badges harder to get, fewer workers would be bestowed with them. This would make the process of capturing the true effect of badges a more difficult task since we would only observe the effect on performance for already high-ability individuals, which is a vivid concern in the awards literature regarding the causal effect of awards. Second, awarding badges based on the correct number of articles submitted would imply that the requester already knew the correct answers he is asking subjects about. This could have alerted workers that they are part of an experiment, generating a potential bias in their answers.

Private digital badges, as they are considered in this study, should not incorporate some of the drawbacks of awards that are publicly advertised. First, not receiving a badge should not demotivate workers because they are not informed of the existence of the badge system or the conditions under the badges are awarded. Therefore, no social comparison costs should arise in this situation (Gallus and Frey, 2016). Second, since digital badges do not get their value from being scarce, but instead from the feedback and personal recognition incorporated in them, digital badges should not suffer a reduction in value as digital badges are given to other workers in this setting (Gallus and Frey, 2016).

## 1.4 Data

Data were collected in five different waves between September, 2015 and April, 2017. The details of the specific dates of data collections and number of subjects per wave are presented in Table 1.3.

### 1.4.1 Filters

Not all subjects that participated in the experiments were considered in the analyses. We applied four filters to our raw dataset. First, all subjects that completed fewer than 10 articles during the first 25 minutes were excluded from the final sample, since they did not meet the minimum productivity threshold required to study the impact of the set of stimuli described above on worker performance. Second, subjects that completed more than 10 articles but that did not answered the survey were excluded as well.[19] Third, for analyses that consider

[19]This occurred for four subjects. AMT workers have the chance to return the HIT, so that another worker can complete it, without being penalized.

data from the additional work period, we only consider subjects that submitted at least six articles during that time. Fourth, although the job post on AMT included a JavaScript code to avoid an AMT worker to accept the HIT more then once, this code does not preclude the same user, using a different account, to accept the HIT again.[20] We kept track of the IP addresses from where subjects connected to the website and excluded all instances where the connection came from the same IP address (this affected to 27 subjects across the 7 different conditions).

### 1.4.2 Descriptive statistics

Descriptive statistics of the data used in the analyses are presented in Table 1.4. The number of subjects in the control and treatment groups varies because (i) not every wave of data collection 'hired' the same number of subjects; and (ii) subjects are randomly assigned to each condition, which in a small sample can lead to some differences in group size. As we can see, the average number of articles and total points obtained are similar between groups during round 1, although subjects in the *BR-20* have the highest average number of articles submitted and correct responses.[21] However, the proportion of subjects that accepted to work during round two seems to differ among conditions. While a 85 percent of workers took, on average, the additional job offer, only a 77 percent did it in the *BR-20* condition. Interestingly, around 92 percent of subjects in the *money* condition accepted the offer. We analyze these decisions in Section 1.5.5.

In our setting, worker performance can be analyzed in different ways. For example, one could consider the total number of articles submitted as a measure of performance, or rather look at the accuracy at which subjects completed each of the articles, as in Eriksson, Poulsen, and Villeval (2009).[22] These and other measures of performance can be computed for each subject by looking at her behavior during the completion of the first 10 articles.[23] In order

[20]The code was provided by `http://uniqueturker.myleott.com`.

[21]We use the terms correct responses and points interchangeably throughout the text. Since workers need to submit four camps for each article, we define the maximum number of points per article as four (one per correct camp).

[22]Although subjects are randomly assigned to the control and treatment groups, our setting allows us to test whether the speed or ability of subjects is balanced across conditions.

[23]Until the $11^{\text{th}}$ article all conditions —including control— do not receive any stimuli. The order of the articles is the same for all conditions throughout the entire experiment.

to identify different skill levels, we calculate the following variables using the data for the first 10 articles submitted.

- **speed10**: this variable shows how fast, on average, the first 10 articles were submitted by each subject. Specifically, we define it as the average speed (per minute) at which subjects send their responses for the first 10 articles. We compute it dividing the number of articles (10) by the number of minutes that took each subject to submit them ($\text{minutes}_{10}$).

$$\text{speed10} = \frac{10}{\text{minutes}_{10}}. \tag{1.1}$$

- **accuracy10**: this variable shows how precise were the responses sent by each subject while working on the first 10 articles. As explained above, for each article there were four camps that had to be filled before the answers could be submitted. If the answer for one of those camps was correct, we coded it with a one (i.e., one 'point') or zero otherwise. Therefore, four is the maximum number of points available per article.

  We define accuracy10 as the total number of points obtained from the first 10 articles divided by 10 (the total number of articles submitted at that point).

$$\text{accuracy10} = \frac{\sum_{j=1}^{10} \text{points}_j}{10}. \tag{1.2}$$

- **speed10p**: similarly to speed10, speed10p is also a measure of speed, but in this case speed10p computes how quickly points were earned (i.e., how fast, on average, a correct response is submitted) while working on the first 10 articles. Specifically, we define speed10p as the total number of points obtained during the first 10 articles divided by the time it took each subject to submit them.

$$\text{speed10p} = \frac{\sum_{j=1}^{10} \text{points}_j}{\text{minutes}_{10}}. \tag{1.3}$$

As can be seen in Table 1.4, although groups do not seem to differ in terms of accuracy10, their average values of speed10 and speed10p do not appear to be the same across treatment arms, especially for subjects in the *BR-20* condition.[24] Since different values of speed10 and

[24]The $p$-values of an F-test with null hypothesis that the treatment arms do not predict the values of speed10, accuracy10, and speed10p are 0.031, 0.249, and 0.024, respectively.

speed10p can have an effect in the total number of articles submitted or points obtained, we control for all these measures of skill (including accuracy) in all our models (except where it is stated otherwise).

### 1.4.3 About the randomization

Tables 1.5, 1.6, and 1.7 presents the average values of the survey questions across treatment arms.[25] For each question in the survey, Tables 1.5, 1.6, and 1.7 present its average value (i.e., the percentage of subjects in each treatment arm that chose each answer) and standard error for the control and treatment groups. If the randomization process was successful, we should observe that conditions are balanced with respect to these variables. The last column in Tables 1.5, 1.6, and 1.7 shows the $p$-value of an F-test with a null hypothesis that groups do not predict the answer to the survey questions. In general, we can appreciate that conditions do not differ for most of the characteristics, although the proportion of females and certain ranges of age, and exposure to video games are statistically different across groups when a level of confidence of 95 percent is chosen. Although our results do not change significatively when analysis are not conditioned on these characteristics, we include controls for the categories that appear to be unbalanced across treatments (considering a 95 percent of confidence) in all the reported regressions models.

## 1.5 Results

### 1.5.1 Round 1: articles and points

We start our analyses focusing on round 1. First, we study the average number of articles submitted per group, controlling for the speed at which the first 10 articles were submitted. Second, we examine how the average number of points obtained after the first 10 articles changes by treatment condition controlling for speed10p $\times$ remaining time, where remaining time represents the minutes left (from the original 25) after submitting the $10^{\text{th}}$ article. This term reflects what the total points of a subject would be at the end of round 1 had she kept her initial speed10p constant after the first 10 articles. In other words, we are controlling for our best guess about a subject's total number of points according to her data during the first

[25] The survey questions and their answers are listed in Appendix 1.8.

10 articles. One might question why we cannot control just for speed10p instead of using the composed term mentioned previously. Contrary to what occurs with speed10, speed10p —by itself— does not allow for a meaningful comparison between subjects. Let's clarify this with an example. Consider subject A that submits her first 10 articles in 10 minutes. Also assume that subject A obtains 10 points in total. Subject A's speed10p value would be 1 point/minute. Now let's examine subject B, who submits her first 10 articles in five minutes but that only gets five points in total. Subject B would have the exact same value of speed10p as subject A: 1 point/minute. However, these two speed10p values, although being numerically equivalent, reflect very different behaviors. While subject A would have 15 more minutes to get points at the 1 point/minute rate, subject B would have 20 additional minutes to obtain points. Therefore, both subjects, although having the same speed10p, would get different total scores keeping everything else constant. Multiplying the subjects' speed10p values by the remaining time after the $10^{\text{th}}$ article yields the correct comparison. Another way to think about this is to recognize that the actual number of points acquired by a subject is a function of her remaining time, speed of submission, and accuracy (see Equation 2.1). We will study the latter two components (speed and accuracy) separately in Section 1.5.2.

$$\text{points}_i = \text{remaining time}_i\,(\text{minutes}) \times \text{speed}_i \left(\frac{\text{articles}}{\text{minutes}}\right) \times \text{accuracy}_i \left(\frac{\text{points}}{\text{articles}}\right). \tag{1.4}$$

As mentioned in Section 1.4.3, all regressions models include dichotomous variables $X_i$ that control for the unbalanced categories identified previously in Tables 1.5, 1.6, and 1.7. Additionally, since some of our models incorporate interaction terms, we include mean-centered versions of all continuous variables to facilitate the interpretation of the estimates of the main effects. All regression results involve robust standard errors. Our main regression models are below.

$$\text{total articles}_i = \alpha + \sum_{c=1}^{6} \beta_c \text{condition}_c + \lambda \text{speed10}_i + X_i'\delta + \epsilon_i. \tag{1.5}$$

$$\text{total points after } 10^{\text{th}} \text{ article}_i = \alpha + \sum_{c=1}^{6} \beta_c \text{condition}_c + \lambda\left(\text{speed10p}_i \times \text{remaining time}_i\right) + X_i'\delta\epsilon_i. \tag{1.6}$$

Table 1.8 presents the results of the regression of total number of articles submitted by subject for different samples (males and females) and with the inclusion of interaction terms between speed10 and group conditions (the same set of control variables is considered in each regression). Under column (1) we have the baseline results for the whole sample. As we can see, no estimate is statistically significant at the 90 percent of confidence with exception of speed10, which is significant in both statistical and substantive terms. A positive one-unit change in speed10 is related to an increment in around 25 articles submitted, which makes sense given the way speed10 is defined (articles/minute) and that the total time alloted for working on the task was 25 minutes. Column (2) presents the same model but now including an interaction term for speed10 and each of the treatment conditions to test for the presence of heterogenous effects. However, no estimate is statistically significant using the traditional thresholds with exception of speed10. Therefore, based on the results of these two regression models it seems that our interventions do not have any statistically significant effect on the total number of articles submitted by subjects.

Column (3) presents our baseline model now considering the subsample of males that participated in the experiment. Our results show that male subjects in the *feedback*, *money*, and *BA* conditions submit, on average, around 1.5 extra articles than males in the control group (estimates of *feedback* and *BA* conditions are statistically significant at the 90 percent of confidence. The estimate for the *money* condition is significant at the 95 percent of confidence). Male subjects in the *BR-40* condition submit almost 2 more articles on average than male workers in the control group, a results that is significant at the 95 percent of confidence. No statistically significant effect is found for males in the *BR-20* and *BR-2040* conditions. When we test for the presence of heterogenous effects in the subsample of males (column (4)), we find that the main effects of the *feedback*, *money*, and *BA* are now around between 1.7 and 2.2 extra articles for males with a speed10 equal to the sample mean (0.9 articles/minute). The estimate for males in the *BR-40* condition now indicates that subjects in this condition submit, on average, 2.4 extra articles than subjects in the control group. The interaction estimates show significant effects for most of the treatment arms. To help with the interpretation, let's consider speed10 a measure of productivity or even motivation before the introduction of any stimuli. This way, a one unit increase in a subject's productivity is related to an additional 9.4 articles submitted if the subject is in the *BA* condition. We find similar results for every other interaction effect with exception of *BR-2040*, for which there is no main or interaction effect statistically different from zero. These results suggest that the effect of our selected set of stimuli on performance depends on the productivity/motivation levels of subjects. We will discuss this more extensively in Section 1.5.3.

For females, the results show a very different picture. Under column (5) in Table 1.8, the only significant estimate besides speed10 is the negative coefficient of the feedback condition. This suggests that females that received only a positive message, but no digital badges or money, had a worse performance than subjects in the control group (-1.8 articles submitted, an estimate that is significant at the 90 percent of confidence). This result remains relatively unchanged when the interaction terms between speed10 and the dichotomous variables that denote the treatment arms are included. This reinforces the idea that the negative impact of feedback on female subjects' performance was similar for everyone in that treatment arm and did not depend on the subjects' initial productivity or motivation.

Table 1.9 presents the results using points obtained after the $10^{\text{th}}$ article. In this case, since we are controlling for speed10p$_i$ $\times$ remaining time$_i$, we do not include speed10 in the regression as a control given that speed10 is one of the components of speed10p (speed10p = speed10 $\times$ accuracy10).[26] As occurred with the total number of articles submitted, there is no significant result for the sample as a whole, not even when the interaction effects are included. For males, subjects in the *BA* condition get, on average, 6.2 more points than subjects in the control group. Subjects in the *BA-40* condition obtain 5.8 extra points when compared to the control group (column (3)). Both estimates are statistically significant at the 95 percent of confidence. Although the estimates for the *feedback* and *money* conditions are positive, they are non-significant in statistical terms even when a threshold of 90 percent of confidence is chosen ($p$-values of 0.18 and 0.54, respectively). When interaction effects are considered (column (4)), the effects for the *BA* and *BR-40* conditions retain their statistical significance and their point estimates are even higher. For example, subjects in the badge effect that have average values of speed10p $\times$ remaining time obtain almost 7.3 additional correct entries than subjects in the control group. While the main effect for the *feedback* condition almost reaches statistically significance at the 90 percent of confidence ($p$-value of 0.101), the estimate for males in the *money* condition remains non-statistically significant. The interaction effects also show interesting results. Male workers in the *BA*, *BR-20*, and *BR-40* conditions present estimates for the interaction terms that are highly significant in statistical and also in substantive terms. A one unit increase in speed10p $\times$ remaining time (i.e., for each extra point expected to be obtained after submitting the $10^{\text{th}}$ article), subjects in the *BA*, *BR-20*, and *BR-40* are expected to get around 0.46, 0.35, and 0.31 additional correct entries, respectively. As it was the case when we analyzed articles submitted, our results support the idea that highly productive workers benefit more from the presence of

[26] Variables speed10 and speed10p have a correlation coefficient higher than 0.9.

these types of stimuli than low-performance workers.

For female subjects, the only significant result in columns (5) and (6) in Table 1.9 makes reference to the lower number of points obtained by subjects in the *feedback* condition. Since female subjects in the *feedback* condition also submit a lower number of articles than subjects in the control group (see results under columns (5) and (6) in Table 1.8), it was natural to expect a lower number of correct entries as well if we assume accuracy did not increase substantially.

### 1.5.2 Round 1: speed and accuracy

There are two main components to the total number of correct entries obtained by a subject during the task: speed of submission and accuracy. We can explain a difference between any of the treatment conditions and the control group as (i) a change in the average speed at which the articles were submitted; (ii) an increase or decrease in the subjects' accuracy; or (iii) a combination of both factors. As explained in Section 1.2.5, we expect our set of stimuli to affect subjects' behavior in different ways, with the total effect on number of correct entries being dependent on the direction and strength of the effects over speed and accuracy. With this in mind, we estimate the effect of our selected set of stimuli on speed and accuracy, measuring both variables using the information of articles submitted after the $10^{\text{th}}$. Specifically, we define our dependent variables of interest for each subject $i$ as follows.

$$\text{speed}_i = \frac{\text{articles}_i - 10}{\text{remaining time}_i}. \tag{1.7}$$

$$\text{accuracy}_i = \frac{\sum_{j=11}^{\text{articles}_i} \text{points}_j}{(\text{articles}_i - 10)}. \tag{1.8}$$

Tables 1.10 and 1.11 present the results considering the new dependent variables above.

For the regression models that consider the speed as the dependent variable, the results in Table 1.10 are very similar to the ones in Table 1.8. However, for male subjects some of the statistical significance of the estimates of the effect of *feedback*, *money*, and *BA* conditions on speed of submission is lost (column (3)). This change in statistical significance is somewhat larger for the *feedback* and *BA* conditions ($p$-values around 0.15) than for the *monet* treatment ($p$-values of 0.11). In any case, the size of the effects is in line with what is reported in Table 1.8. For males in the badge condition, the estimate under column (3) indicates that subjects in this treatment increase their average speed in 0.091 articles/minutes after submitting the 10$^{th}$ article. On average, subjects in this condition submitted the 10$^{th}$ article around the 12:35 (minutes:seconds) mark, which means they have 12:25 left to work on the task after observing the first digital badge. Multiplying the increase in speed (0.091 articles/minutes) by the time left (roughly 12.42 minutes), this gives us around 1.3 extra articles. Although this estimate is lower than the one we found for males in the badges treatment in Table 1.8, it is inside the 95 percent confidence interval for the effect of the badges on the total number of articles submitted. The estimate for the impact on subjects in the *BR-40* condition remains positive highly significant in statistical terms. Results under column (4) in Table 1.10 are also similar to those in Table 1.8. The main effects for *feedback*, *money*, *BA* and *BR-40* are positive and statistically significant at the 90 percent of confidence (the coefficient of *BA* and *BR-40* are significant at the 95 and 99 percent of confidence, respectively). The interaction effects for these treatment arms are also statistically significant and, as it was reported in Table 1.8, the interaction effect of speed10 and *BR-20* is also significant in statistical terms. All together, our results imply that male subjects with an average value of speed10 increase their speed of submission after starting receiving feedback messages, monetary payments, and certain digital badges (*BA* and *BA-40*). Additionally, the effect of this set of stimuli (plus *BR-20*) on submission speed is increasing in speed10. High-performance males seem to benefit the most from the interventions.

For females, the speed of submission is negatively affected by the presence of feedback messages, a result that is statistically significant at the 90 and 95 percent of confidence for the models under columns (5) and (6), respectively. No other main of interaction effect is statistically significant (or close to).

With respect to the effects on accuracy, the results in Table 1.11 show a different side of the story. Although there are some negative effects of the treatment conditions on accuracy for the whole sample, these are driven mostly by female subjects. For males, there is no estimate statistically significant at the 90 percent of confidence under column (3), although the estimate for the *money* condition becomes statistically significant at the 90 percent of confidence under column (4) when interaction effects are included. This suggests that males in this treatment arm lowered their accuracy in around 0.24 points per article after the 10$^{th}$ submission when their accuracy10 was the same as the sample mean (2.99 correct entries per article). On the other hand, females subjects show lower levels of accuracy in almost every treatment arm with exception of *money*. Females, in general, lowered their accuracy in between 0.17 and 0.2 points per article after they started observing digital messages and badges. These negative effects for females do not seem to differ with pre-intervention performance levels (column (6)).

Overall, our results for speed and accuracy tell a similar story to the one we found when we analyzed number of articles submitted and points obtained.[27] Male subjects seem to increase their average speed of submission after the intervention started for most of the treatment arms, and these effects seem to be greater for highly-productive workers. Only a negative effect on accuracy for males in the *money* condition is observed for males. Females in the feedback group show a lower submission speed, and women in all treatment groups (with exception of subjects in the *condition* group) present a lower average accuracy after the $10^{\text{th}}$ article is submitted. There is no significant interaction of pre-intervention accuracy and treatment condition for males or females in our sample.

### 1.5.3 Panel data analysis

The evidence presented in the two previous sections indicate that the higher the initial productivity levels of a male subject, the higher the expected effect of certain monetary and non-monetary incentives on his performance (measured it as number of articles submitted). If we assume that our set of stimuli can have positive effects on worker performance the longer a subject is exposed to them (i.e., the greater the number of messages/badges/monetary incentives she receives), then workers who initially are high performers will be the ones that will receive the higher number of stimuli and, in consequence, will be more likely to show any reaction to their presence.

There is one more subtle thing to examine. Let's consider number of articles submitted or speed of submission as our measures of performance (i.e., quantity). Since our performance indicators only consider submissions that were actually completed, small improvements in submission time after receiving a stimulus are not captured in the analyses presented above if that reduction in time is not big enough to guarantee the submission of an additional article that would not have been submitted otherwise. Therefore, small changes in submission time —especially for those subjects with relatively low speeds of submission— might not be reflected in the results when looking at data aggregated at the individual level.

To study the effects of our stimuli on both speed of submission and accuracy, we order our data as if they corresponded to a panel (longitudinal) setting, with article as the time

[27]Remember that changes in total points obtained reflects variation in both speed and accuracy.

dimension. The logic is the following: although we do not observe subjects at the same points in time, we do observe their behavior whenever they submit an article. Since articles appear in the same order for everybody, we can use the information recorded for each article (i.e., time required to submit and points obtained) to make inference on the effect of our set of stimuli on speed and accuracy controlling for all time-invariant characteristics of subjects (such as ability, age, educational level, video game familiarity, etc.) that we could only partially control for in the analyses presented in Sections 1.5.1 and 1.5.2 by using the responses from the survey.[28] This way, we can study the impact of each type and number of stimulus given to subjects in their subsequent submissions.

For each treatment arm we define dichotomous variables according to the thresholds discussed in Section 1.3.4. Each dichotomous variable includes the articles in the range defined by each threshold starting with the 11$^{\text{th}}$ article. This way, the first dichotomous variable for each treatment spans articles 11 to 15 (including extremes), the second comprises articles 16 to 20 (including extremes), and so on. Since the last stimulus is given after completion of the 35$^{\text{th}}$ article, the corresponding dichotomous variable includes from the article number 36 until the last one submitted by a subject. Constructing the dichotomous variables in this form allows us to test the effect of our selected stimuli in a difference-in-difference fashion, where the effects are always understood in comparison to the pre-intervention period (first 10 articles submitted).

We consider two different dependent variables: time required to submit an article (natural logarithm of seconds) and number of correct entries per article. Since most of the subjects complete fewer than 40 articles, we exclude any article submission from the 41$^{\text{st}}$ article onwards so to focus on parameters that are estimated with the 'within' variation of at least a couple of subjects. Besides including fixed effects for each subject in the sample, we also control for any particularities of articles (such as their difficulty or length) that might affect either the speed of submission or the number of correct entries (or both) using article fixed effects. Tables 1.12 and 1.13 present the results for the whole sample and male and female subjects.

For the natural logarithm of seconds, our results show that for the sample as a whole only two estimates (out of 36) are statistically significant at the 90 and 99 percent of significance,

[28]As we saw above, the randomization was for the most part effective, but not completely successful. Additionally, we can only claim the randomization procedure was effective for the variables we can actually observe, which might leave other confounding factor without proper control in our analyses.

which could be mostly consequence of random variation and the presence of some outlier observations rather than the representation of a true relationship in the data. For male subjects (column (2)), the results indicate that individuals in the *feedback* condition decreased their average submission time in about 22 percent (relative to the average submission time of their first 10 articles) after receiving the fourth message and in around 30 percent after receiving the sixth, estimates that are statistically significant at the 90 and 99 percent of confidence, respectively. Male subjects in the *money* condition show a similar behavior regarding submission time as individuals in the *feedback* group: they show a statistically significant decrease in submission time after the fifth and sixth monetary payment. Male workers in the *BA* condition reduced their average submission time in almost 28 percent after receiving the fourth badge, and in around 19 percent after observing the fifth badge (both estimates are significant at the 95 percent of confidence). Subjects in the *BA* condition even started to show a lower average submission time (-10 percent) after getting the third badge, although this result is only statistically significant at the 89 percent of confidence. On the other hand, estimates for males in the *BR-20* condition denote that the reaction of individuals to the digital badges was more on the opposite direction of what occurred with most of the other treatments. According to the results in Table 1.15, males in the *BR-20* condition took, on average, more time per article after receiving the first three badges than during the submission of the first 10 articles. Subjects in the *BR-2040* do not present any variation in their speed of submission after obtaining any of the badges. Finally, males in the *BR-40* show lower submission times after the first and sixth badges. All the results discussed so far are in line with what we found in Tables 1.8 and 1.10.

For females (column (3)), regression results regarding submission time look very different. All significant estimates (at the 90 percent of confidence and above) are positive, which means that most of our stimuli were ineffective and negatively affected female subjects' performance in terms of speed of submission.

For the number of points obtained per article, the results in columns (4) to (6) present a similar picture to the one observed in Tables 1.8 and 1.10. Males (column (5)) in the *feedback* condition experience a drop in the average number of correct entries after the fifth and sixth messages. In the *feedback* condition, females (column (6)) show a decrease in the average number of correct entries after as early as the second message is received ($15^{\text{th}}$ article). This negative relationship is found in almost all estimates for females that received positive feedback, with exception of the effect of the last feedback message, which is estimated as positive. For *BA* badges, males show a decrease in their average accuracy only after the sixth badge, while for females there is no significant effect in statistical terms. Interestingly, male subjects that received badges that show information about relative performance present a decline in their accuracy after the fifth badge and in some cases the reduction in correct entries is even more pronounced that the one experimented by subjects in the *BA* condition. For females, in general the estimates present a negative relationship for some of the badge conditions and accuracy, although no one is as strong as the one observed for the *feedback* condition. However, although our estimates suggest that subjects lower their accuracy after a certain number of stimuli have been received (probably as consequence of 'gaming' and the interest of acquiring additional stimuli —messages and badges— at expenses of the quality of the work being submitted), the results for subjects in the *money* condition show a more extreme situation. Male subjects in this condition present a decrease in their accuracy after the third monetary reward has been given. This reduction in the number of correct entries is increasing in the number of monetary rewards received, reaching -1.3 correct entries per article (relative to the overall accuracy observed during the first 10 articles) after a male subject in the money condition has started working on his 36$^{st}$ article. The decrease for females is less pronounced and it is noticeable mostly after the fourth and fifth payment has been received.

It is important to mention that the analyses presented so far are complementary to the ones in Sections 1.5.1 and 1.5.2. As explained at the beginning of the current section, relationships observed in Tables 1.8 to 1.11 should be reflected in the results of the analyses of the data in panel form. However, some of the results of our panel data analysis do not, necessarily, need to be observed in our main analyses. Take as an example the higher submission time of females after receiving the sixth badge (column (3)). Although this estimate is highly significant in statistical and substantive terms, it is an effect that does not appear in the aggregate analysis mainly for two reasons. First, the number of subjects that actually reach the 40$^{th}$ article is relative low when compared to the number of subjects that have the chance to complete the first 20 articles, which dilutes the average effect found for high performers in the estimates. Second, estimates in our panel data analysis are more prone to be affected by extreme observations than in the analyses where we use the aggregate data. For example, it is quite possible for a subject to dramatically reduce their average speed if she observes that, after submitting many articles, there is no enough time to submit more than one (a similar argument can be applied to the effect of some stimuli on accuracy in the case where a subject assumes there is not enough time to carefully complete a submission and opts for sending the answers without making sure the information in them is accurate).

Therefore, even though we could observe a significant estimate for the effect of a stimulus on some variable of interest when we use our panel data specification, that would not imply that a similar effect will be detected in the aggregate data we use in our main analyses.

As we discussed in Section 1.5.1, our results suggest that high-performance males show the most benefit from our set of stimuli. At the beginning of this section, we argued that individuals with higher productivity levels will receive more stimuli than low-productivity subjects, which increases their chances of showing a reaction to the presence of any of the interventions. To test the impact of our set of incentives on low and high-productivity subjects we divide workers into quartiles according to their pre-intervention values of speed and accuracy and check for the presence of any effect after the introduction of the treatments.[29] We focus our attention on the first 20 articles submitted, so in the analyses presented below we discard any information on articles submitted after the $20^{\text{th}}$ to make sure we can compare the effects on different types of workers. As before, our models include both subject and article fixed effects.

To facilitate the understanding of the results, Tables 1.14 to 1.17 present the average marginal effects of each treatment by quartile for submission time (natural logarithm of seconds) and points obtained per article, respectively.

As we can observe in Tables 1.14 and 1.15 , in general the effect of the first two stimuli given in each treatment arm have a positive impact on speed of submission, especially for the subjects in the first quartile after they received the second stimuli. Interestingly, the positive effect of most of the treatments is similar for both male and female subjects. For subjects in the top quartiles, the effect of the treatments on submission time between articles 11 and 20 are usually non-significant and close to zero, although for some the effect is negative and statistically significant. When we analyze the effect on correct entries, the results in Tables 1.16 and 1.17 indicate that workers in the higher quartiles tend to do worse after receiving a stimulus while working in articles 11 to 20. Although not all estimates for the top two quartiles are statistically significant at the 90 percent level of confidence, in general the estimates show a negative relationship between a high-performance subject's number of correct entries per article and the presence of a stimulus. For low-performance subjects, most of the results show a positive effect after the first stimulus is observed. However, the

[29] Instead of having two different set of quartiles (one for each dimension), we also estimate the models ordering the workers just using their speed10p values. The results obtained are similar and are available upon request.

effects of the second stimulus tend to be negative.

Based on the results exposed in Tables 1.14 to 1.17, we can observe that the effect of our treatments are not limited to high-performance individuals (although these are more likely to receive, and be affected by, our stimuli). Low-performance subjects, specially those with lower speed of submission, tend to improve their submission time. Although the effects on number of correct entries are less clear, in general we can appreciate that our set of stimuli affects both high- and low-performance subjects, sometimes even in opposing directions.

### 1.5.4 Survey

After completing the first 25 minutes of work, subjects answered 16 multiple-choice questions as the final step to complete the HIT.[30] Although no numeric scale was offered as a potential response to non-demographic questions, choices were always worded in a similar way in order to keep consistency. For example, when asked about their opinion on the clarity of instructions, how interesting they found the HIT, or how challenging the task was, subjects always faced options in the following order:

- Extremely clear/interesting/challenging.
- Very clear/interesting/challenging.
- Moderately clear/interesting/challenging.
- Slightly clear/interesting/challenging.
- Not at all clear/interesting/challenging.

Since answers to this question express a logical order (e.g., 'extremely clear' implies the instructions are clearer than in the case where 'very clear' is chosen), we analyze the responses using an ordered logistic model. We operationalize the responses in the following way: we give the lowest score (1 in our case) to the answers that show the highest disagreement (e.g., 'not at all challenging' or 'strongly disagree') and the highest score (5) to the answers that express the highest degree of conformity (e.g., 'extremely interesting' or 'strongly agree').

[30] The questions and their potential answers are listed in Appendix 1.8.

Table 1.18 presents the results of our analyses for the first three questions of the survey, which are reproduced below. All models control for the actual number of articles completed in the first round and the unbalanced categories identified previously. [31]

- **Question 1**: How clear were the instructions of this HIT?
- **Question 2**: How interesting was the HIT?
- **Question 3**: How challenging was the HIT?

As we can appreciate, the treatment arm does not help predict the response of a subject regarding the clarity of the instructions; no estimate is statistically significant at the 90 percent of confidence. The estimate for males in the *BA* condition can be considered large (in absolute value) when compared to the estimates for the other conditions, but the result is statistically significant only at the 86 percent of confidence. When asked about how interesting the task was, subjects in the *BA* and *BR-20* conditions considered the HIT to be less interesting than what individuals in the control group expressed (both estimates significant at the 90 percent of confidence). When looking at the sub-samples, the only statistically significant result found is for males in the *BA* condition. Finally, as it occurred with the first question, the treatment arms do not help predict the response given by and individual about how challenging the HIT was.

The survey also included the following two questions about the recognition and compensation provided by the employer who posted the HIT on AMT. In both of these questions, the answers ranged from 'strongly disagree' to 'strongly agree' (five options in total).

- **Question 5**: The requester of this HIT recognizes good job performance.
- **Question 6**: I am satisfied with the overall compensation I received for working on this HIT.

The results of the ordered logistic models for the questions above are presented in Table 1.19. According to the estimates in columns (1) to (3), males in the *money* condition are more likely to choose an answer with a higher degree of conformity when asked about the recognition

[31] We also considered models with interaction terms for treatment arm and number of articles completed, but results, in general, showed little difference between this specification and the one without interaction effects.

given by the requester of the HIT. No other treatment arm seems to have a significant effect on the answer selected by an individual (relative to the control group). Interestingly, there is no effect for females in this treatment arm. When asked about satisfaction with the overall compensation received, we find that subjects in the *money* condition are more likely to choose answers that express full agreement with the question, while the contrary occurs to subjects in the *BA* (these results are statistically significant at the 95 and 90 percent of confidence). The estimate for the *BR-20* condition is similar in magnitude to the estimate for the *BA* condition, but it is only significant at the 88 percent of confidence. Although the estimates for the *BA* and *BR-20* conditions are similar for both males and females, males seem to be the driver behind the effect found for the full sample.

According to the results, only male subjects that received extra monetary payments are more likely to agree to the idea that good performance was recognized and also express more satisfaction with the overall compensation received. On the other hand, both males and female subjects in the *BA* and *BR-20* groups seem to consider badges as a cheap substitution of money. Badges that reward absolute performance (*BA*) and those that acknowledge a high performance (*BR-20*) might have created the expectation of something more tangible as a reward rather than just a digital icon on the screen.

### 1.5.5 Additional job offer

Badges, as awards in general, can also be understood as signals from the giver to the recipient, especially in the case of discretionary awards. By bestowing an award on a worker, an employer is signaling her interest in creating loyalty bond (Frey and Gallus, 2014). However, this loyalty bond is in part mediated by the positive feedback that creates the idea of a psychological contract with a long-term employment vision between the employer and the employee (Suazo, Martínez, and Sandoval, 2009). Therefore, by the establishment of this loyalty bond, workers should be more prone to work for the firm in the future. Also, workers' vision of the task might have changed as well towards a more positive one after being awarded a badge. They could be more inclined to accept another task as a consequence of a more positive attitude.

We estimate a logistic regression model to study the effect of our set of stimuli on a subject's likelihood of accepting the additional round offer. Our binary dependent variable takes the value 1 if the subject accepted the additional round and 0 otherwise. We keep the same set of control variables that we have used so far in this study and we also add the total number of articles submitted during the first round as an additional covariate to control for any potential effects of past performance on the chances of accepting the offer. We present the results of our estimations in Table 1.20.

For the sample as a whole, subjects in the *BR-20* condition show a lower probability of accepting the second round of work. When we divide the sample by gender, we notice that there is no significant difference between treatment arms and the control group in the case of males. For females, being part of the *BR-20* condition makes subjects less likely to accept the additional round (a result that is statistically significant at the 95 percent of confidence). The coefficient for the feedback condition is negative, but it is not statistically significant at the 90 percent of confidence ($p$-value: 0.21). Tablẽreftab: accepted probabilities presents the adjusted probabilities of accepting the second round by condition, considering all other covariates at their sample means.[32] The overall probability of accepting the second round is close to 90 percent for most of the conditions, with the exception of females subjects in the*BR-20* condition that show an adjusted probability of accepting the offer of roughly 75 percent.

### 1.5.6 Round 2: points and articles

Figure 1.4. Presents a histogram of the number of articles submitted during the second round by the subjects who accepted the additional job offer. Figure 1.5 shows a scatterplot between the number of articles submitted in the first and second rounds. As we can see, the number of articles submitted during round 2 is roughly the 60 percent of what subjects submitted in round 1.

Tables 1.22 and 1.23 show the results of our regression models for the articles and points obtained during the second round of the task. Besides our regular set of controls for the unbalanced categories discussed in Section 1.4.3, we include the total number of articles submitted and the total number of points obtained in the first round as additional covariates in our first and second models, respectively.[33] We include these two variables also as controls for the potential effect past-performance can have on self-goals (Mathieu and Button, 1992).

[32] In this case, sample means are defined by each subsample.

[33] We do not include speed10 as a control since the total number of articles submitted in the first round already takes into account the pre-intervention speed of submission. Speed10 and total articles submitted have a coefficient of correlation above 0.92.

As mentioned previously, we only consider observations in which at least 6 articles were submitted during the second round.

In terms of the number of articles submitted during the second round (Table 1.22), we can appreciate that for the sample as a whole, subjects in conditions *BR-2040* and *BR-40* submit, on average, around one more extra articles than the control group. The estimate for those who received *BA* badges in the first round is positive, but not statistically significant at the 90 percent of confidence (column (1)). For males, subjects in conditions *BA* and *BR-40* submit, on average, almost 1.7 additional articles than the control group (column (3)). Individuals in the *feedback* and *BR-2040* send almost one more article than the control group, but these estimates are only significant around the 85 percent of confidence. Results for females are similar, although no estimate is significant at the 90 percent of confidence. Females in the *BR-2040* submit almost one extra article than the control group, but this estimate is statistically significant just at the 89 percent of confidence. Estimates for *BR-20* and *BR-40* are similar in magnitude, but fail to reach statistically significance. The negative impact of feedback messages on females seems to persist even after the end of the first round, although results do not allow us to reject the null hypothesis of no effect in this case.

When considering interaction terms in the full sample (column (2)), only one of interaction term is statistically significant at the 95 percent of confidence: for subjects in the *money* condition, the higher the number of articles they submitted in the first round, the lower the number of articles submitted in the second round. This result appears to be significant only for males (column (4)). Figure 1.6 presents the average marginal effect of the *money* condition on the number of articles submitted for different values of productivity (i.e., articles sent) in the first round. As we can observe, there is a clear downward slope as the number of articles submitted in the first round increases. For low-productivity workers, the extra monetary payments received seem to increase effort in the second round relatively to the control group, although the higher the productivity in round one (and, in consequence, the number of monetary bonuses awarded) the lower the observed productivity in the second round. Since this negative interaction effect of *money* and initial productivity occurs only for the condition that obtained monetary payments, this can be considered evidence of a reduction in intrinsic motivation for subjects that received a high number of monetary payments as in Deci (1971) and Deci (1972b). According to self-determination theory, an extrinsic reward —such as an unexpected monetary payment— should not, per se, decrease a subject's internal motivation unless the subject generates an expectation for the monetary payment in the future. The fact that this negative relationship between performance in the first round and performance in the second round only occurs for subjects in the *money* condition and not for the other treatments might be evidence that badges can help improve performance without the crowding-out effect observed for monetary payments in the literature. This evidence is coherent with the idea that badges can help foster a subject's autonomy and feelings of competency, which can improve her intrinsic motivation for a task.

## 1.6 Discussion and limitations

Although in this paper I have analyzed the impact of a specific set of conditions on worker performance, I do not consider the effect of the stimuli on affective states such as happiness and feelings of dominance (Azmat and Iriberri, 2012). The effect of badges on affective levels might be of importance to organizations, since motivation, employee morale, and even retention rates are variables that can influence worker and firm performance (Azmat and Iriberri, 2012; Frey and Stutzer, 2002).

It is important to remember that the estimated effects in this paper are under a situation were workers were hired under a fixed wage. In that respect, the effects analyzed might represent, in some cases, a lower bound of what the true could be under a piece-rate scenario. For example, the monetary payments given to workers might Azmat and Iriberri (2012) find that the impact of relative feedback is larger when subjects are paid according to their performance rather than when they receive a fixed amount.[34]

Similarly, the effectiveness of financial incentives and badges will probably be very different in a scenario in which they are presented to subjects before the task starts. This is specially true for the *money* and digital badges conditions, since there is ample evidence in the economics literature of the motivating power of financial incentives (Prendergast, 1999) and also of awards (Ashraf, Bandiera, and Lee, 2014; Kosfeld and Neckermann, 2011; Neckermann, Cueni, and Frey, 2014). As Antin and Churchill note, a fully implemented badge system can be used to transmit information to workers relative to what is feasible to do on a platform. This way employers could design badge structures such as to communicate what is expected from the users of the platform.

[34] In Azmat and Iriberri (2012), subjects work for four periods, and they are provided with their relative performance after the second period (subjects knew beforehand that they would receive that information at that point). Azmat and Iriberri provide estimates of the interaction between relative feedback and gender, but only for subjects under the piece-rate condition. This way, we do not know if the negative estimate they got for females was different for subjects that received positive or negative relative feedback information.

Although there was no proper badge system in place, i.e. subjects were not alerted of its presence beforehand and its characteristics (such as total number of badges available and the requirements to obtain them) remained hidden, there is still a chance that subjects might have thought of badges (and the other stimulus as well) as 'controlling' initiatives that implicitly told workers to improve their performance and get closer to the requester's 'ideal' target. Clearly, that was not the idea behind the stimuli, and even though I do not have any data to support or reject that conjecture, it is certainly a possibility. The way positive feedback or rewards are transmitted/given to employees can have an effect on how they perceive it, influencing on whether they think of them as controlling or informational (Deci, Connell, and Ryan, 1989). For example, Ryan (1982) finds that subjects that received controlling feedback performed slightly worse than those who got informational feedback. The messages used in this study (and that were included in all treatment conditions) were created with the purpose of highlighting the informational rather than the controlling aspect (Deci and Ryan, 1980; Kast and Connor, 1988; Ryan, 1982). [35] Of the six messages that accompany all digital stimuli, only the first one ('Keep up the good work!') looks like it could be considered controlling by workers. Future research should look into this issue reducing the possibility that feedback embedded in other stimuli or by itself could be thought of as controlling by participants.

The literature on awards and gamification has identified some interesting areas for future research, specially in firms. The introduction of badges on school settings has proved beneficial, but the application of similar mechanisms in a workplace faces some important challenges. First, scholars do not know how the demographic differences among employees in firm might shape their disposition or attitude towards some regularly used elements in gamification such as badges (Landers and Callan, 2011). Second, there are doubts about the optimal design of an award or badge system (Neckermann, Cueni, and Frey, 2014), and while some theoretical research has been done on this topic, there are still few unanswered questions (Anderson et al., 2013). Third, more empirical research is needed in order to understand how badges, and awards more largely, can affect performance (Neckermann et al. 2014). Hence, improving our knowledge about these topics appears as fruitful avenues for scholars to pursue.

[35]However, it is quite possible that females had interpreted the messages in a different way than males did as in Deci (1972a) and Deci, Cascio, and Krusell (1973), which could have affected their disposition to work and their overall performance.

## 1.7 Conclusion

The focus of this paper has been the collection of causal empirical evidence with respect to the effectiveness of discretionary awards (digital badges) that are kept private to the recipient. According to Frey and Gallus (2014), gathering data on award distribution is one difficult direction researchers should move into. I show how researchers can test the effects of digital badges on worker performance by coding a website without complex features.

The results indicate that, in general, digital badges that are awarded as consequence of quantity (in our case, number of articles submitted) increase the output in terms of quantity for males, but not for females. In fact, females subjects show a decrease in their total output when they only receive feedback messages, similarly to what occurred in Deci (1972a). While the positive effects for quantity in males were not accompanied by a decrease in quality (accuracy), for females the effects were quite different. For all treatment conditions except for *money*, there was a drop in average accuracy. Dissimilar effects for male and female subjects have been found before in the literature as consequence of feedback (Azmat and Iriberri, 2012; Deci, Cascio, and Krusell, 1973). I also discuss how already high-performance subjects were more likely to be affected by the treatments considered in this experiment, and in consequence, more likely to drive the results. Although this is true, I show that subjects that received one or two stimuli also reacted to them.

The overall effect of a certain stimulus on a subject's performance should be the net impact it has on both her intrinsic and extrinsic motivation. So as long as the potential reduction in a subject's intrinsic motivation as consequence of a treatment is overcome by a desire of getting more of it (extrinsic motivation), which might lead her to increase her effort, then net effect on performance will be positive. However, since in our case treatments only make explicit the number of articles submitted and not their quality, the positive effect on number of articles submitted might also arise as the outcome of a trade-off between quantity and quality. If subjects derive some utility from our set of stimuli, and since the only condition to obtain them is based on the number of articles submitted, then it might be optimum for a rational worker to not increase their effort but rather increase their speed of submission at the expense of accuracy (conditional on a low probability that her submissions could be considered of bad quality by the requester).[36] We only observe this trade-off for

[36]Requesters can reject submissions on AMT if they consider the quality of the work to be low. A rejection negatively affects a workers' rating on AMT.

male subjects that received monetary payments. While subjects that received badges (*BA* and *BR-40*) increased their speed of submission without impacting their accuracy, workers in the *money* condition decreased their average accuracy. Additionally, for male workers in the *money* condition, the data suggest that the greater the number of monetary payments received during round 1, the smaller the output (in number of articles and correct entries) during round 2. Since the average output of workers in the *money* condition do not differ by much from what workers in other treatments completed in round 1, we can discard fatigue as a potential explanation (as it is suggested by Prendergast (1999)). One possibility is that the presence of monetary payments in round 1 affected the locus of causality (Ryan, 1982) of workers in the *money* condition, especially those who got a greater number of monetary payments. Then, when they start the second round, the absence of monetary incentives traduced in a lower output for those who experimented a crowding-out effect Frey and Jegen (2001).

With respect to the external validity of the results, it is worth mentioning that the causal effect of private badges on worker performance captured in the experiment might represent either a lower or an upper bound of what the true effect would be in a more traditional workplace. Although subjects were not specifically told that this was a one-time job, nothing in the instructions or on the website mentioned future options of employment. To the extent workers did not consider future HITs from us a possibility, this should have affected their beliefs about the requester's interests of developing a long-term relationship and, as a consequence, the effectiveness of badges as motivators. On the other hand, the literature on gamification has shown that the users' enjoyment of *gamified* systems tend to decrease with time (Farzan et al., 2008; Koivisto and Hamari, 2014). This would mean a decreasing impact of private badges on worker performance overt time (unless new features such as other game mechanics or different badges bring freshness are included periodically).

Lately, scholars have started to analyze the strategic use of awards by companies to motivate people and enhance firm performance, a topic mostly absent from the management literature in previous years. Gallus and Frey (2016) discuss how firms can appropriate some of the value created by awards in an organization and the potential sources of value destruction when awards are not implemented adequately. I certainly believe the use of non-monetary extrinsic rewards can be a fruitful subject of research in the future. However, as this paper shows, the private dimension of non-monetary extrinsic rewards plays a relevant role in employee motivation and can be a source of important productivity improvements.

Table 1.1: Common feedback messages

| Threshold | Message |
|---|---|
| 10 Articles | Congratulations, *nickname*! You completed 10 articles. Keep up the good work! |
| 15 Articles | Congratulations, *nickname*! You completed 15 articles. Way to go! |
| 20 Articles | Congratulations, *nickname*! You completed 20 articles. You're doing a great job! |
| 25 Articles | Congratulations, *nickname*! You completed 25 articles. Excellent progress! |
| 30 Articles | Congratulations, *nickname*! You completed 30 articles. You should be proud! |
| 35 Articles | Congratulations, *nickname*! You completed 35 articles. Amazing performance! |

Table 1.2: Summary of treatment arms

| Condition | Characteristics |
|---|---|
| Feedback | Positive feedback message after each threshold. |
| Money | Feedback + $0.2 after each threshold. |
| BA | Feedback + digital badge (number of articles submitted) after each threshold. |
| BR-20 | Feedback + digital badge (top 20 percent) after each threshold. |
| BR-2040 | Feedback + digital badge (top 20 percent) after first threshold.<br>Digital badge (top 40 percent) after all other thresholds. |
| BR-40 | Feedback + digital badge (top 40 percent) after each threshold. |

Table 1.3: Data collection

| Data | Date | Conditions | N |
|---|---|---|---|
| Wave 1 | September 2-3, 2015 | Control, Feedback, Money, BA | 130 |
| Wave 2 | April 13, 2017 | Control, Feedback, Money, BA | 66 |
| Wave 3 | April 19, 2017 | Control, Feedback, Money, BA | 79 |
| Wave 4 | April 27-28, 2017 | BR-20, BR-2040, BR-40 | 251 |

Table 1.4: Descriptive statistics

| Group | Statistic | Round 1 Articles | Round 1 Points | speed10 | accuracy10 | speed10p | Accepted Offer | Round 2 Articles | Round 2 Points |
|---|---|---|---|---|---|---|---|---|---|
| Control | Mean | 22.08 | 66.15 | 0.86 | 3.01 | 2.6 | 0.89 | 15.77 | 51.81 |
| | SD | 7.38 | 25.86 | 0.3 | 0.55 | 1.08 | 0.32 | 5.44 | 20.7 |
| | N | 71 | 71 | 71 | 71 | 71 | 71 | 62 | 62 |
| Feedback | Mean | 22.36 | 65.55 | 0.88 | 3 | 2.65 | 0.84 | 15.45 | 48.96 |
| | SD | 7.12 | 24.94 | 0.26 | 0.54 | 0.94 | 0.37 | 4.75 | 17.94 |
| | N | 64 | 64 | 64 | 64 | 64 | 64 | 53 | 53 |
| Money | Mean | 23.47 | 70.62 | 0.89 | 3.12 | 2.8 | 0.92 | 16.8 | 53.85 |
| | SD | 8.56 | 27.27 | 0.31 | 0.49 | 1.06 | 0.28 | 4.8 | 18.29 |
| | N | 60 | 60 | 60 | 60 | 60 | 60 | 55 | 55 |
| BA | Mean | 22.99 | 68.49 | 0.88 | 2.99 | 2.65 | 0.84 | 17.23 | 56.33 |
| | SD | 8.12 | 30.11 | 0.29 | 0.52 | 1.08 | 0.37 | 5.72 | 22.91 |
| | N | 80 | 80 | 80 | 80 | 80 | 80 | 66 | 66 |
| BR-20 | Mean | 25.84 | 78.16 | 1.02 | 3.05 | 3.14 | 0.77 | 17.83 | 57.28 |
| | SD | 9.18 | 33.9 | 0.35 | 0.54 | 1.28 | 0.42 | 6.05 | 23.66 |
| | N | 70 | 70 | 70 | 70 | 70 | 70 | 54 | 54 |
| BR-2040 | Mean | 22.46 | 64 | 0.87 | 2.89 | 2.53 | 0.88 | 16.95 | 54.4 |
| | SD | 7.33 | 26.4 | 0.28 | 0.58 | 1.01 | 0.33 | 5.32 | 19.88 |
| | N | 91 | 91 | 91 | 91 | 91 | 91 | 78 | 78 |
| BR-40 | Mean | 24.59 | 72.41 | 0.93 | 2.97 | 2.82 | 0.84 | 18.11 | 57.28 |
| | SD | 9.62 | 34.26 | 0.36 | 0.57 | 1.33 | 0.36 | 6.52 | 23.85 |
| | N | 90 | 90 | 90 | 90 | 90 | 90 | 76 | 76 |
| Total | Mean | 23.41 | 69.24 | 0.9 | 2.99 | 2.73 | 0.85 | 16.93 | 54.45 |
| | SD | 8.31 | 29.56 | 0.31 | 0.55 | 1.13 | 0.35 | 5.62 | 21.31 |
| | N | 526 | 526 | 526 | 526 | 526 | 526 | 444 | 444 |

The number of subjects in round 2 per condition might not match the number of subjects during round 1 times the proportion of those who accepted the second round. We incorporate an extra filter during round 2 of at least six articles submitted during that round (in addition to the filters already considered for observations in round 1 — see Section 1.4.1) for an observation to be included in our sample.

Table 1.5: Balance 1

| | (1) Control | (2) Feedback | (3) Money | (4) B-A | (5) B-R20 | (6) B-R2040 | (7) B-R40 | (8) Overall | (9) p-value |
|---|---|---|---|---|---|---|---|---|---|
| Female | 0.592 | 0.469 | 0.517 | 0.550 | 0.757 | 0.626 | 0.600 | 0.591 | 0.025 |
| | (0.059) | (0.063) | (0.065) | (0.056) | (0.052) | (0.051) | (0.052) | (0.021) | |
| Age | | | | | | | | | |
| 18-20 | 0.014 | 0.016 | 0.000 | 0.037 | 0.014 | 0.011 | 0.011 | 0.015 | 0.707 |
| | (0.014) | (0.016) | (0.000) | (0.021) | (0.014) | (0.011) | (0.011) | (0.005) | |
| 21-29 | 0.296 | 0.281 | 0.367 | 0.362 | 0.257 | 0.319 | 0.211 | 0.297 | 0.319 |
| | (0.055) | (0.057) | (0.063) | (0.054) | (0.053) | (0.049) | (0.043) | (0.020) | |
| 30-39 | 0.324 | 0.453 | 0.300 | 0.338 | 0.343 | 0.308 | 0.356 | 0.344 | 0.597 |
| | (0.056) | (0.063) | (0.060) | (0.053) | (0.057) | (0.049) | (0.051) | (0.021) | |
| 40-49 | 0.155 | 0.125 | 0.150 | 0.163 | 0.271 | 0.176 | 0.311 | 0.198 | 0.025 |
| | (0.043) | (0.042) | (0.046) | (0.042) | (0.054) | (0.040) | (0.049) | (0.017) | |
| 50-99 | 0.211 | 0.125 | 0.183 | 0.100 | 0.114 | 0.187 | 0.111 | 0.146 | 0.294 |
| | (0.049) | (0.042) | (0.050) | (0.034) | (0.038) | (0.041) | (0.033) | (0.015) | |
| Race | | | | | | | | | |
| White | 0.648 | 0.781 | 0.783 | 0.762 | 0.700 | 0.626 | 0.767 | 0.721 | 0.124 |
| | (0.057) | (0.052) | (0.054) | (0.048) | (0.055) | (0.051) | (0.045) | (0.020) | |
| Black | 0.099 | 0.031 | 0.067 | 0.062 | 0.086 | 0.143 | 0.100 | 0.087 | 0.298 |
| | (0.036) | (0.022) | (0.032) | (0.027) | (0.034) | (0.037) | (0.032) | (0.012) | |
| Native American | 0.000 | 0.000 | 0.000 | 0.013 | 0.000 | 0.011 | 0.000 | 0.004 | 0.654 |
| | (0.000) | (0.000) | (0.000) | (0.012) | (0.000) | (0.011) | (0.000) | (0.003) | |
| Asian | 0.113 | 0.109 | 0.100 | 0.062 | 0.043 | 0.099 | 0.067 | 0.084 | 0.660 |
| | (0.038) | (0.039) | (0.039) | (0.027) | (0.024) | (0.031) | (0.026) | (0.012) | |
| Hispanic | 0.099 | 0.062 | 0.033 | 0.037 | 0.114 | 0.077 | 0.022 | 0.063 | 0.158 |
| | (0.036) | (0.030) | (0.023) | (0.021) | (0.038) | (0.028) | (0.016) | (0.011) | |
| Multiple | 0.042 | 0.016 | 0.017 | 0.062 | 0.057 | 0.044 | 0.044 | 0.042 | 0.769 |
| | (0.024) | (0.016) | (0.017) | (0.027) | (0.028) | (0.022) | (0.022) | (0.009) | |
| Income | | | | | | | | | |
| $0-$24,999 | 0.408 | 0.312 | 0.433 | 0.388 | 0.343 | 0.308 | 0.311 | 0.354 | 0.539 |
| | (0.059) | (0.058) | (0.065) | (0.055) | (0.057) | (0.049) | (0.049) | (0.021) | |
| $25,000-$49,999 | 0.338 | 0.359 | 0.350 | 0.338 | 0.329 | 0.385 | 0.344 | 0.350 | 0.993 |
| | (0.057) | (0.060) | (0.062) | (0.053) | (0.057) | (0.051) | (0.050) | (0.021) | |
| $50,000-$74,999 | 0.141 | 0.203 | 0.183 | 0.200 | 0.186 | 0.176 | 0.200 | 0.184 | 0.969 |
| | (0.042) | (0.051) | (0.050) | (0.045) | (0.047) | (0.040) | (0.042) | (0.017) | |
| $75,000-$99,999 | 0.070 | 0.109 | 0.017 | 0.037 | 0.129 | 0.077 | 0.100 | 0.078 | 0.174 |
| | (0.031) | (0.039) | (0.017) | (0.021) | (0.040) | (0.028) | (0.032) | (0.012) | |
| $100,000-$199,999 | 0.042 | 0.016 | 0.017 | 0.037 | 0.014 | 0.055 | 0.044 | 0.034 | 0.720 |
| | (0.024) | (0.016) | (0.017) | (0.021) | (0.014) | (0.024) | (0.022) | (0.008) | |
| Reason for Hits | | | | | | | | | |
| Income purposes | 0.310 | 0.312 | 0.333 | 0.287 | 0.271 | 0.231 | 0.300 | 0.289 | 0.864 |
| | (0.055) | (0.058) | (0.061) | (0.051) | (0.054) | (0.044) | (0.049) | (0.020) | |
| Extra cash | 0.521 | 0.578 | 0.600 | 0.600 | 0.700 | 0.659 | 0.622 | 0.614 | 0.420 |
| | (0.060) | (0.062) | (0.064) | (0.055) | (0.055) | (0.050) | (0.051) | (0.021) | |
| Fruitful | 0.127 | 0.062 | 0.050 | 0.113 | 0.029 | 0.044 | 0.056 | 0.068 | 0.152 |
| | (0.040) | (0.030) | (0.028) | (0.036) | (0.020) | (0.022) | (0.024) | (0.011) | |
| It is fun | 0.028 | 0.016 | 0.017 | 0.000 | 0.000 | 0.033 | 0.011 | 0.015 | 0.529 |
| | (0.020) | (0.016) | (0.017) | (0.000) | (0.000) | (0.019) | (0.011) | (0.005) | |
| To kill time | 0.014 | 0.031 | 0.000 | 0.000 | 0.000 | 0.033 | 0.011 | 0.013 | 0.312 |
| | (0.014) | (0.022) | (0.000) | (0.000) | (0.000) | (0.019) | (0.011) | (0.005) | |
| N | 71 | 64 | 60 | 80 | 70 | 91 | 90 | 526 | |

Table 1.6: Balance 2

| | (1) Control | (2) Feedback | (3) Money | (4) B-A | (5) B-R20 | (6) B-R2040 | (7) B-R40 | (8) Overall | (9) p-value |
|---|---|---|---|---|---|---|---|---|---|
| Hours working on HITs | | | | | | | | | |
| 1-5 | 0.169 | 0.141 | 0.117 | 0.163 | 0.171 | 0.242 | 0.156 | 0.169 | 0.533 |
| | (0.045) | (0.044) | (0.042) | (0.042) | (0.045) | (0.045) | (0.038) | (0.016) | |
| 6-9 | 0.169 | 0.172 | 0.200 | 0.263 | 0.314 | 0.220 | 0.211 | 0.222 | 0.370 |
| | (0.045) | (0.048) | (0.052) | (0.050) | (0.056) | (0.044) | (0.043) | (0.018) | |
| 10-20 | 0.310 | 0.359 | 0.433 | 0.225 | 0.214 | 0.286 | 0.356 | 0.308 | 0.065 |
| | (0.055) | (0.060) | (0.065) | (0.047) | (0.049) | (0.048) | (0.051) | (0.020) | |
| 21-29 | 0.099 | 0.094 | 0.117 | 0.138 | 0.143 | 0.088 | 0.078 | 0.106 | 0.797 |
| | (0.036) | (0.037) | (0.042) | (0.039) | (0.042) | (0.030) | (0.028) | (0.013) | |
| 30-168 | 0.254 | 0.234 | 0.133 | 0.212 | 0.157 | 0.165 | 0.200 | 0.194 | 0.548 |
| | (0.052) | (0.053) | (0.044) | (0.046) | (0.044) | (0.039) | (0.042) | (0.017) | |
| Education | | | | | | | | | |
| Associate | 0.155 | 0.203 | 0.050 | 0.125 | 0.229 | 0.154 | 0.111 | 0.146 | 0.083 |
| | (0.043) | (0.051) | (0.028) | (0.037) | (0.051) | (0.038) | (0.033) | (0.015) | |
| Bachelor | 0.394 | 0.406 | 0.400 | 0.362 | 0.371 | 0.341 | 0.322 | 0.367 | 0.921 |
| | (0.058) | (0.062) | (0.064) | (0.054) | (0.058) | (0.050) | (0.050) | (0.021) | |
| Graduate | 0.056 | 0.094 | 0.083 | 0.087 | 0.086 | 0.132 | 0.133 | 0.099 | 0.654 |
| | (0.028) | (0.037) | (0.036) | (0.032) | (0.034) | (0.036) | (0.036) | (0.013) | |
| High school (HS) | 0.141 | 0.031 | 0.117 | 0.113 | 0.129 | 0.099 | 0.111 | 0.106 | 0.519 |
| | (0.042) | (0.022) | (0.042) | (0.036) | (0.040) | (0.031) | (0.033) | (0.013) | |
| Less than HS | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.022 | 0.000 | 0.004 | 0.143 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.015) | (0.000) | (0.003) | |
| Some college | 0.254 | 0.266 | 0.350 | 0.312 | 0.186 | 0.253 | 0.322 | 0.278 | 0.379 |
| | (0.052) | (0.056) | (0.062) | (0.052) | (0.047) | (0.046) | (0.050) | (0.020) | |
| Employment | | | | | | | | | |
| Full-time | 0.451 | 0.562 | 0.550 | 0.600 | 0.543 | 0.571 | 0.489 | 0.538 | 0.561 |
| | (0.059) | (0.062) | (0.065) | (0.055) | (0.060) | (0.052) | (0.053) | (0.022) | |
| Part-time | 0.211 | 0.156 | 0.200 | 0.150 | 0.171 | 0.198 | 0.244 | 0.192 | 0.755 |
| | (0.049) | (0.046) | (0.052) | (0.040) | (0.045) | (0.042) | (0.046) | (0.017) | |
| Looking | 0.141 | 0.156 | 0.133 | 0.075 | 0.186 | 0.088 | 0.067 | 0.116 | 0.169 |
| | (0.042) | (0.046) | (0.044) | (0.030) | (0.047) | (0.030) | (0.026) | (0.014) | |
| Not looking | 0.099 | 0.078 | 0.100 | 0.113 | 0.100 | 0.088 | 0.133 | 0.103 | 0.949 |
| | (0.036) | (0.034) | (0.039) | (0.036) | (0.036) | (0.030) | (0.036) | (0.013) | |
| Retired | 0.070 | 0.031 | 0.000 | 0.013 | 0.000 | 0.022 | 0.033 | 0.025 | 0.110 |
| | (0.031) | (0.022) | (0.000) | (0.012) | (0.000) | (0.015) | (0.019) | (0.007) | |
| Disabled | 0.028 | 0.016 | 0.017 | 0.050 | 0.000 | 0.033 | 0.033 | 0.027 | 0.621 |
| | (0.020) | (0.016) | (0.017) | (0.025) | (0.000) | (0.019) | (0.019) | (0.007) | |
| N | 71 | 64 | 60 | 80 | 70 | 91 | 90 | 526 | |

Table 1.7: Balance 3

| | (1) Control | (2) Feedback | (3) Money | (4) B-A | (5) B-R20 | (6) B-R2040 | (7) B-R40 | (8) Overall | (9) p-value |
|---|---|---|---|---|---|---|---|---|---|
| Newspaper frequency | | | | | | | | | |
| Every day | 0.310 | 0.359 | 0.417 | 0.338 | 0.271 | 0.374 | 0.311 | 0.338 | 0.648 |
| | (0.055) | (0.060) | (0.064) | (0.053) | (0.054) | (0.051) | (0.049) | (0.021) | |
| Once a week | 0.282 | 0.219 | 0.200 | 0.275 | 0.300 | 0.242 | 0.278 | 0.259 | 0.830 |
| | (0.054) | (0.052) | (0.052) | (0.050) | (0.055) | (0.045) | (0.047) | (0.019) | |
| Times a month | 0.169 | 0.250 | 0.217 | 0.125 | 0.243 | 0.220 | 0.222 | 0.205 | 0.491 |
| | (0.045) | (0.055) | (0.054) | (0.037) | (0.052) | (0.044) | (0.044) | (0.018) | |
| Times a year | 0.127 | 0.109 | 0.083 | 0.200 | 0.143 | 0.088 | 0.100 | 0.122 | 0.306 |
| | (0.040) | (0.039) | (0.036) | (0.045) | (0.042) | (0.030) | (0.032) | (0.014) | |
| Never | 0.113 | 0.062 | 0.083 | 0.062 | 0.043 | 0.077 | 0.089 | 0.076 | 0.800 |
| | (0.038) | (0.030) | (0.036) | (0.027) | (0.024) | (0.028) | (0.030) | (0.012) | |
| Video games frequency | | | | | | | | | |
| Every day | 0.268 | 0.203 | 0.350 | 0.200 | 0.229 | 0.242 | 0.200 | 0.238 | 0.390 |
| | (0.053) | (0.051) | (0.062) | (0.045) | (0.051) | (0.045) | (0.042) | (0.019) | |
| Once a week | 0.268 | 0.344 | 0.317 | 0.312 | 0.186 | 0.264 | 0.278 | 0.279 | 0.499 |
| | (0.053) | (0.060) | (0.061) | (0.052) | (0.047) | (0.046) | (0.047) | (0.020) | |
| Times a month | 0.141 | 0.188 | 0.133 | 0.163 | 0.143 | 0.231 | 0.222 | 0.179 | 0.526 |
| | (0.042) | (0.049) | (0.044) | (0.042) | (0.042) | (0.044) | (0.044) | (0.017) | |
| Times a year | 0.169 | 0.219 | 0.150 | 0.150 | 0.229 | 0.077 | 0.189 | 0.165 | 0.168 |
| | (0.045) | (0.052) | (0.046) | (0.040) | (0.051) | (0.028) | (0.041) | (0.016) | |
| Never | 0.155 | 0.047 | 0.050 | 0.175 | 0.214 | 0.187 | 0.111 | 0.139 | 0.018 |
| | (0.043) | (0.027) | (0.028) | (0.043) | (0.049) | (0.041) | (0.033) | (0.015) | |
| Feedback from other requesters | | | | | | | | | |
| Extremely often | 0.028 | 0.016 | 0.017 | 0.000 | 0.000 | 0.000 | 0.033 | 0.013 | 0.298 |
| | (0.020) | (0.016) | (0.017) | (0.000) | (0.000) | (0.000) | (0.019) | (0.005) | |
| Very often | 0.042 | 0.094 | 0.017 | 0.087 | 0.071 | 0.055 | 0.033 | 0.057 | 0.392 |
| | (0.024) | (0.037) | (0.017) | (0.032) | (0.031) | (0.024) | (0.019) | (0.010) | |
| Moderately often | 0.169 | 0.188 | 0.150 | 0.200 | 0.200 | 0.165 | 0.111 | 0.167 | 0.745 |
| | (0.045) | (0.049) | (0.046) | (0.045) | (0.048) | (0.039) | (0.033) | (0.016) | |
| Slightly often | 0.408 | 0.406 | 0.483 | 0.463 | 0.457 | 0.341 | 0.511 | 0.437 | 0.344 |
| | (0.059) | (0.062) | (0.065) | (0.056) | (0.060) | (0.050) | (0.053) | (0.022) | |
| Not at all often | 0.352 | 0.297 | 0.333 | 0.250 | 0.271 | 0.440 | 0.311 | 0.325 | 0.175 |
| | (0.057) | (0.058) | (0.061) | (0.049) | (0.054) | (0.052) | (0.049) | (0.020) | |
| N | 71 | 64 | 60 | 80 | 70 | 91 | 90 | 526 | |

Table 1.8: Articles submitted.

| (Variable) | (1) (All) | (2) (All) | (3) (Male) | (4) (Male) | (5) (Female) | (6) (Female) |
|---|---|---|---|---|---|---|
| Feedback | −0.340 | −0.247 | 1.399* | 1.709** | −1.798* | −2.172* |
| | (0.590) | (0.673) | (0.817) | (0.760) | (0.917) | (1.156) |
| Money | 0.575 | 0.695 | 1.775** | 2.238*** | −0.181 | −0.225 |
| | (0.564) | (0.619) | (0.897) | (0.840) | (0.745) | (0.804) |
| BA | 0.426 | 0.552 | 1.592* | 2.048** | −0.289 | −0.386 |
| | (0.499) | (0.566) | (0.901) | (0.895) | (0.585) | (0.654) |
| BR-20 | −0.266 | −0.076 | −0.821 | −1.250 | −0.423 | −0.437 |
| | (0.602) | (0.585) | (1.297) | (1.162) | (0.655) | (0.686) |
| BR-2040 | 0.049 | 0.138 | 0.790 | 0.992 | −0.400 | −0.460 |
| | (0.493) | (0.567) | (0.944) | (0.922) | (0.567) | (0.653) |
| BR-40 | 0.619 | 0.707 | 1.974** | 2.366*** | −0.313 | −0.468 |
| | (0.456) | (0.512) | (0.780) | (0.753) | (0.569) | (0.633) |
| Speed10 | 24.526*** | 22.240*** | 25.323*** | 18.515*** | 24.215*** | 25.862*** |
| | (0.594) | (2.398) | (1.123) | (2.977) | (0.701) | (2.643) |
| Feedback × Speed10 | | 1.803 | | 8.301** | | −7.248 |
| | | (3.354) | | (3.285) | | (6.180) |
| Money× Speed10 | | 3.450 | | 10.482*** | | −2.135 |
| | | (2.954) | | (3.365) | | (3.387) |
| BA × Speed10 | | 3.264 | | 9.416*** | | −1.399 |
| | | (2.936) | | (3.414) | | (3.360) |
| BR-20 × Speed10 | | 1.656 | | 9.589** | | −2.945 |
| | | (2.828) | | (3.860) | | (3.256) |
| BR-2040 × Speed10 | | 1.649 | | 4.193 | | −1.727 |
| | | (2.676) | | (4.139) | | (2.888) |
| BR-40 × Speed10 | | 3.329 | | 6.889* | | −0.169 |
| | | (2.552) | | (3.541) | | (2.806) |
| Constant | 22.250*** | 22.250*** | 21.618*** | 21.855*** | 27.335*** | 27.404*** |
| | (1.510) | (1.503) | (1.818) | (2.060) | (1.636) | (1.659) |
| Observations | 526 | 526 | 215 | 215 | 311 | 311 |
| R-squared | 0.861 | 0.863 | 0.858 | 0.873 | 0.874 | 0.878 |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.9: Correct answers (points) after 10$^{th}$ article.

| (Variable) | (1) (All) | (2) (All) | (3) (Male) | (4) (Male) | (5) (Female) | (6) (Female) |
|---|---|---|---|---|---|---|
| Feedback | −1.632 | −1.234 | 3.630 | 4.030 | −5.928** | −6.246** |
| | (1.735) | (1.917) | (2.719) | (2.492) | (2.345) | (2.897) |
| Money | −0.429 | −0.004 | 1.891 | 2.131 | −1.323 | −1.231 |
| | (1.909) | (2.014) | (3.074) | (3.044) | (2.442) | (2.503) |
| BA | 1.283 | 1.796 | 6.186** | 7.295*** | −1.819 | −1.970 |
| | (1.670) | (1.846) | (2.933) | (2.790) | (1.903) | (2.079) |
| BR-20 | −1.119 | −0.867 | −1.820 | −3.576 | −1.916 | −2.097 |
| | (1.965) | (1.874) | (4.334) | (3.637) | (2.054) | (2.081) |
| BR-2040 | −0.645 | −0.309 | 1.595 | 1.764 | −1.900 | −2.125 |
| | (1.618) | (1.884) | (2.930) | (3.082) | (1.852) | (2.120) |
| BR-40 | 0.802 | 1.136 | 5.822** | 6.150** | −2.434 | −2.689 |
| | (1.553) | (1.694) | (2.680) | (2.551) | (1.840) | (1.990) |
| Speed10p × Rem. Time | 0.969*** | 0.855*** | 0.971*** | 0.718*** | 0.970*** | 0.999*** |
| | (0.023) | (0.096) | (0.045) | (0.138) | (0.026) | (0.073) |
| Feedback × (Speed10p × Rem. Time) | | 0.115 | | 0.231 | | −0.053 |
| | | (0.106) | | (0.147) | | (0.140) |
| Money × (Speed10p × Rem. Time) | | 0.072 | | 0.215 | | −0.086 |
| | | (0.116) | | (0.173) | | (0.116) |
| BA × (Speed10p × Rem. Time) | | 0.184 | | 0.460*** | | −0.021 |
| | | (0.121) | | (0.149) | | (0.111) |
| BR-20 × (Speed10p × Rem. Time) | | 0.130 | | 0.352** | | −0.023 |
| | | (0.111) | | (0.165) | | (0.107) |
| BR-2040 × (Speed10p × Rem. Time) | | 0.098 | | 0.222 | | −0.055 |
| | | (0.112) | | (0.180) | | (0.101) |
| BR-40 × (Speed10p × Rem. Time) | | 0.146 | | 0.311** | | −0.007 |
| | | (0.101) | | (0.156) | | (0.081) |
| Constant | 35.923*** | 35.913*** | 35.353*** | 36.107*** | 48.416*** | 48.642*** |
| | (3.537) | (3.637) | (4.291) | (5.212) | (1.859) | (1.988) |
| Observations | 526 | 526 | 215 | 215 | 311 | 311 |
| R-squared | 0.862 | 0.864 | 0.846 | 0.861 | 0.883 | 0.884 |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.10: Average speed after $10^{\text{th}}$ article.

| (Variable) | (1) (All) | (2) (All) | (3) (Male) | (4) (Male) | (5) (Female) | (6) (Female) |
|---|---|---|---|---|---|---|
| Feedback | −0.028 | −0.024 | 0.080 | 0.097* | −0.119* | −0.143** |
| | (0.041) | (0.044) | (0.056) | (0.053) | (0.062) | (0.072) |
| Money | 0.026 | 0.033 | 0.096 | 0.122** | −0.018 | −0.018 |
| | (0.040) | (0.041) | (0.060) | (0.056) | (0.055) | (0.058) |
| BA | 0.013 | 0.021 | 0.091 | 0.115* | −0.034 | −0.038 |
| | (0.036) | (0.038) | (0.063) | (0.062) | (0.043) | (0.045) |
| BR-20 | −0.021 | −0.007 | −0.065 | −0.095 | −0.029 | −0.026 |
| | (0.039) | (0.039) | (0.078) | (0.074) | (0.045) | (0.046) |
| BR-2040 | −0.008 | −0.001 | 0.044 | 0.054 | −0.040 | −0.039 |
| | (0.035) | (0.037) | (0.063) | (0.061) | (0.043) | (0.045) |
| BR-40 | 0.030 | 0.035 | 0.128** | 0.151*** | −0.037 | −0.044 |
| | (0.032) | (0.034) | (0.050) | (0.048) | (0.043) | (0.045) |
| Speed10 | 0.947*** | 0.810*** | 0.995*** | 0.619*** | 0.932*** | 0.993*** |
| | (0.036) | (0.135) | (0.067) | (0.159) | (0.045) | (0.169) |
| Feedback × Speed10 | | 0.063 | | 0.453** | | −0.477 |
| | | (0.203) | | (0.200) | | (0.377) |
| Money × Speed10 | | 0.213 | | 0.594*** | | −0.080 |
| | | (0.170) | | (0.192) | | (0.212) |
| BA × Speed10 | | 0.204 | | 0.496** | | −0.022 |
| | | (0.166) | | (0.209) | | (0.205) |
| BR-20 × Speed10 | | 0.077 | | 0.552** | | −0.172 |
| | | (0.161) | | (0.214) | | (0.202) |
| BR-2040 × Speed10 | | 0.132 | | 0.192 | | −0.014 |
| | | (0.158) | | (0.251) | | (0.191) |
| BR-40 × Speed10 | | 0.207 | | 0.391** | | 0.033 |
| | | (0.146) | | (0.193) | | (0.183) |
| Constant | 0.918*** | 0.920*** | 0.881*** | 0.897*** | 1.246*** | 1.247*** |
| | (0.103) | (0.103) | (0.130) | (0.140) | (0.110) | (0.110) |
| Observations | 526 | 526 | 215 | 215 | 311 | 311 |
| R-squared | 0.652 | 0.656 | 0.650 | 0.673 | 0.676 | 0.686 |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.11: Average accuracy after $10^{th}$ article.

| (Variable) | (1) (All) | (2) (All) | (3) (Male) | (4) (Male) | (5) (Female) | (6) (Female) |
|---|---|---|---|---|---|---|
| Feedback | −0.128** | −0.128** | −0.053 | −0.054 | −0.192** | −0.188** |
| | (0.063) | (0.063) | (0.094) | (0.112) | (0.079) | (0.075) |
| Money | −0.168** | −0.173** | −0.224 | −0.240* | −0.098 | −0.084 |
| | (0.077) | (0.080) | (0.137) | (0.142) | (0.077) | (0.077) |
| BA | −0.107* | −0.107* | 0.015 | 0.007 | −0.189** | −0.188** |
| | (0.063) | (0.063) | (0.081) | (0.089) | (0.094) | (0.092) |
| BR-20 | −0.132** | −0.129** | −0.066 | −0.089 | −0.171** | −0.177** |
| | (0.064) | (0.064) | (0.120) | (0.111) | (0.078) | (0.077) |
| BR-2040 | −0.155** | −0.154** | −0.127 | −0.113 | −0.171** | −0.170** |
| | (0.062) | (0.064) | (0.100) | (0.108) | (0.080) | (0.080) |
| BR-40 | −0.152*** | −0.152*** | −0.072 | −0.081 | −0.202*** | −0.208*** |
| | (0.055) | (0.055) | (0.090) | (0.104) | (0.070) | (0.067) |
| Accuracy10 | 0.843*** | 0.783*** | 0.836*** | 0.769*** | 0.840*** | 0.828*** |
| | (0.034) | (0.089) | (0.062) | (0.119) | (0.042) | (0.117) |
| Feedback × Accuracy10 | | 0.029 | | 0.002 | | 0.020 |
| | | (0.142) | | (0.214) | | (0.187) |
| Money × Accuracy10 | | 0.093 | | 0.134 | | −0.066 |
| | | (0.128) | | (0.179) | | (0.173) |
| BA × Accuracy10 | | 0.156 | | 0.170 | | 0.101 |
| | | (0.137) | | (0.214) | | (0.178) |
| BR-20 × Accuracy10 | | −0.002 | | −0.209 | | 0.081 |
| | | (0.123) | | (0.194) | | (0.142) |
| BR-2040 × Accuracy10 | | 0.077 | | 0.157 | | −0.021 |
| | | (0.115) | | (0.172) | | (0.151) |
| BR-40 × Accuracy10 | | 0.060 | | 0.139 | | −0.044 |
| | | (0.110) | | (0.202) | | (0.131) |
| Speed10 | 0.291*** | 0.289*** | 0.205 | 0.218 | 0.331*** | 0.338*** |
| | (0.071) | (0.072) | (0.139) | (0.136) | (0.081) | (0.083) |
| Constant | 2.991*** | 3.008*** | 3.034*** | 3.055*** | 2.967*** | 3.008*** |
| | (0.135) | (0.140) | (0.168) | (0.164) | (0.293) | (0.337) |
| Observations | 526 | 526 | 215 | 215 | 311 | 311 |
| R-squared | 0.602 | 0.604 | 0.571 | 0.581 | 0.644 | 0.646 |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.12: Longitudinal analysis: seconds (ln) and points per article.

| Variable | (1)<br>Seconds<br>All | (2)<br>Seconds<br>Males | (3)<br>Seconds<br>Females | (4)<br>Points<br>All | (5)<br>Points<br>Males | (6)<br>Points<br>Females |
|---|---|---|---|---|---|---|
| Feedback (11-15) | 0.045 | 0.009 | 0.076 | −0.034 | 0.020 | −0.085 |
| | (0.032) | (0.037) | (0.052) | (0.073) | (0.102) | (0.101) |
| Feedback (16-20) | −0.014 | −0.033 | −0.009 | −0.174** | −0.104 | −0.239* |
| | (0.040) | (0.065) | (0.049) | (0.088) | (0.115) | (0.134) |
| Feedback (21-25) | 0.001 | −0.039 | 0.024 | −0.148* | −0.136 | −0.224* |
| | (0.040) | (0.062) | (0.056) | (0.086) | (0.112) | (0.121) |
| Feedback (26-30) | −0.066 | −0.223* | 0.096** | −0.297** | −0.276 | −0.255 |
| | (0.063) | (0.131) | (0.046) | (0.142) | (0.183) | (0.255) |
| Feedback (31-35) | −0.008 | −0.098 | 0.010 | −0.461*** | −0.573*** | −0.329** |
| | (0.070) | (0.095) | (0.078) | (0.108) | (0.131) | (0.145) |
| Feedback (36-40) | −0.051 | −0.302*** | 0.123*** | 0.262 | −0.488*** | 0.487*** |
| | (0.075) | (0.071) | (0.040) | (0.222) | (0.140) | (0.186) |
| Money (11-15) | −0.025 | −0.035 | −0.022 | −0.072 | −0.091 | −0.047 |
| | (0.032) | (0.042) | (0.046) | (0.073) | (0.105) | (0.099) |
| Money (16-20) | 0.002 | −0.030 | 0.021 | −0.150* | −0.196 | −0.104 |
| | (0.041) | (0.059) | (0.058) | (0.083) | (0.137) | (0.098) |
| Money (21-25) | 0.027 | −0.032 | 0.061 | −0.147* | −0.288** | −0.054 |
| | (0.047) | (0.071) | (0.063) | (0.085) | (0.139) | (0.103) |
| Money (26-30) | −0.049 | −0.166 | 0.014 | −0.417*** | −0.574*** | −0.274*** |
| | (0.067) | (0.139) | (0.063) | (0.108) | (0.193) | (0.104) |
| Money (31-35) | −0.120 | −0.203* | −0.081 | −0.515*** | −0.719*** | −0.394*** |
| | (0.077) | (0.104) | (0.102) | (0.104) | (0.176) | (0.120) |
| Money (36-40) | −0.032 | −0.322*** | 0.105* | −0.227 | −1.332*** | 0.234 |
| | (0.074) | (0.058) | (0.054) | (0.291) | (0.246) | (0.197) |
| BA (11-15) | 0.003 | −0.013 | 0.014 | −0.026 | 0.044 | −0.080 |
| | (0.029) | (0.042) | (0.040) | (0.069) | (0.090) | (0.100) |
| BA (16-20) | −0.010 | −0.038 | 0.006 | −0.026 | 0.133 | −0.138 |
| | (0.037) | (0.062) | (0.046) | (0.082) | (0.102) | (0.116) |
| BA (21-25) | −0.002 | −0.101 | 0.059 | −0.119 | −0.014 | −0.197 |
| | (0.036) | (0.062) | (0.041) | (0.094) | (0.105) | (0.134) |
| BA (26-30) | −0.109* | −0.277** | −0.010 | −0.168* | −0.163 | −0.155 |
| | (0.063) | (0.137) | (0.048) | (0.097) | (0.150) | (0.125) |
| BA (31-35) | −0.012 | −0.192** | 0.099 | −0.120 | −0.066 | −0.186 |
| | (0.077) | (0.097) | (0.103) | (0.124) | (0.121) | (0.190) |
| BA (36-40) | 0.138 | −0.184 | 0.296*** | 0.034 | −0.450*** | 0.061 |
| | (0.102) | (0.146) | (0.055) | (0.245) | (0.155) | (0.218) |
| Constant | 4.095*** | 4.078*** | 4.106*** | 3.229*** | 3.250*** | 3.215*** |
| | (0.015) | (0.023) | (0.019) | (0.033) | (0.048) | (0.044) |
| Observations | 12,184 | 4,920 | 7,264 | 12,183 | 4,919 | 7,264 |
| R-squared | 0.322 | 0.288 | 0.358 | 0.434 | 0.451 | 0.434 |
| Number of id | 526 | 215 | 311 | 526 | 215 | 311 |

Individual and article fixed effects included in all models.
Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.13: Longitudinal analysis: seconds (ln) and points per article (cont.).

| Variable | (1) Seconds (All) | (2) Seconds (Males) | (3) Seconds (Females) | (4) Points (All) | (5) Points (Males) | (6) Points (Females) |
|---|---|---|---|---|---|---|
| BR-20 (11-15) | 0.037 | 0.098** | 0.027 | −0.033 | 0.111 | −0.088 |
| | (0.029) | (0.045) | (0.038) | (0.067) | (0.142) | (0.083) |
| BR-20 (16-20) | 0.040 | 0.140* | 0.017 | −0.135* | −0.210 | −0.121 |
| | (0.040) | (0.074) | (0.047) | (0.080) | (0.138) | (0.101) |
| BR-20 (21-25) | 0.127*** | 0.209** | 0.108** | −0.143* | −0.079 | −0.155 |
| | (0.047) | (0.101) | (0.051) | (0.083) | (0.123) | (0.108) |
| BR-20 (26-30) | 0.026 | 0.033 | 0.049 | −0.145 | −0.053 | −0.179 |
| | (0.063) | (0.150) | (0.042) | (0.089) | (0.166) | (0.108) |
| BR-20 (31-35) | 0.058 | 0.036 | 0.085 | −0.193** | −0.298* | −0.136 |
| | (0.067) | (0.101) | (0.081) | (0.095) | (0.174) | (0.104) |
| BR-20 (36-40) | 0.057 | −0.066 | 0.103** | 0.241 | −0.549*** | 0.492*** |
| | (0.058) | (0.055) | (0.041) | (0.215) | (0.104) | (0.186) |
| BR-2040 (11-15) | −0.003 | −0.041 | 0.022 | 0.040 | 0.153 | −0.033 |
| | (0.030) | (0.046) | (0.040) | (0.065) | (0.093) | (0.088) |
| BR-2040 (16-20) | −0.032 | −0.018 | −0.040 | −0.156** | −0.147 | −0.163 |
| | (0.039) | (0.074) | (0.044) | (0.079) | (0.119) | (0.105) |
| BR-2040 (21-25) | 0.028 | 0.065 | 0.009 | −0.164** | −0.303** | −0.089 |
| | (0.044) | (0.091) | (0.043) | (0.081) | (0.139) | (0.100) |
| BR-2040 (26-30) | −0.028 | −0.094 | 0.016 | −0.166* | −0.127 | −0.183 |
| | (0.065) | (0.147) | (0.052) | (0.098) | (0.151) | (0.126) |
| BR-2040 (31-35) | −0.005 | −0.134 | 0.056 | −0.118 | −0.385** | 0.008 |
| | (0.073) | (0.109) | (0.087) | (0.125) | (0.190) | (0.151) |
| BR-2040 (36-40) | 0.113 | −0.014 | 0.120** | −0.010 | −0.918*** | 0.321 |
| | (0.079) | (0.108) | (0.052) | (0.296) | (0.351) | (0.300) |
| BR-40 (11-15) | 0.005 | −0.059* | 0.048 | −0.032 | −0.063 | −0.011 |
| | (0.027) | (0.034) | (0.038) | (0.063) | (0.096) | (0.084) |
| BR-40 (16-20) | 0.004 | −0.036 | 0.029 | −0.146* | −0.016 | −0.225** |
| | (0.035) | (0.056) | (0.044) | (0.080) | (0.117) | (0.106) |
| BR-40 (21-25) | 0.025 | −0.015 | 0.049 | −0.096 | −0.105 | −0.095 |
| | (0.034) | (0.057) | (0.042) | (0.072) | (0.096) | (0.100) |
| BR-40 (26-30) | −0.010 | −0.114 | 0.047 | −0.136 | −0.107 | −0.144 |
| | (0.061) | (0.136) | (0.041) | (0.089) | (0.156) | (0.109) |
| BR-40 (31-35) | −0.016 | −0.105 | 0.027 | −0.277*** | −0.268** | −0.297** |
| | (0.064) | (0.092) | (0.078) | (0.096) | (0.117) | (0.133) |
| BR-40 (36-40) | 0.049 | −0.206*** | 0.146*** | 0.104 | −0.399*** | 0.188 |
| | (0.059) | (0.050) | (0.041) | (0.224) | (0.136) | (0.186) |
| Constant | 4.095*** | 4.078*** | 4.106*** | 3.229*** | 3.250*** | 3.215*** |
| | (0.015) | (0.023) | (0.019) | (0.033) | (0.048) | (0.044) |
| Observations | 12,184 | 4,920 | 7,264 | 12,183 | 4,919 | 7,264 |
| R-squared | 0.322 | 0.288 | 0.358 | 0.434 | 0.451 | 0.434 |
| Number of id | 526 | 215 | 311 | 526 | 215 | 311 |

Individual and article fixed effects included in all models.
Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.14: Average marginal effects per quartile: seconds (ln).

| Variable | (1) All (11-15) | (2) All (16-20) | (3) Males (11-15) | (4) Males (16-20) | (5) Females (11-15) | (6) Females (16-20) |
|---|---|---|---|---|---|---|
| **Feedback** | | | | | | |
| Feedback × Q1 | −0.049 | −0.195*** | −0.051 | −0.155* | −0.050 | −0.223*** |
| | (0.047) | (0.048) | (0.058) | (0.088) | (0.071) | (0.052) |
| Feedback × Q2 | 0.000 | −0.118* | −0.035 | −0.219* | 0.027 | −0.040 |
| | (0.045) | (0.062) | (0.068) | (0.112) | (0.058) | (0.061) |
| Feedback × Q3 | 0.107** | 0.069 | 0.039 | 0.070 | 0.177** | 0.041 |
| | (0.044) | (0.043) | (0.039) | (0.060) | (0.082) | (0.062) |
| Feedback × Q4 | 0.118* | 0.077 | 0.065 | 0.035 | 0.185 | 0.116 |
| | (0.070) | (0.054) | (0.058) | (0.074) | (0.166) | (0.085) |
| **Money** | | | | | | |
| Money × Q1 | −0.101* | −0.094 | −0.094 | −0.385*** | −0.112** | 0.029 |
| | (0.052) | (0.128) | (0.092) | (0.057) | (0.053) | (0.105) |
| Money × Q2 | −0.098* | −0.071 | −0.057 | −0.050 | −0.184** | −0.132 |
| | (0.053) | (0.057) | (0.053) | (0.064) | (0.092) | (0.107) |
| Money × Q3 | 0.085** | 0.058 | 0.053 | 0.020 | 0.105** | 0.080 |
| | (0.042) | (0.042) | (0.087) | (0.076) | (0.047) | (0.049) |
| Money × Q4 | 0.009 | 0.067 | −0.020 | 0.022 | 0.030 | 0.097 |
| | (0.033) | (0.066) | (0.053) | (0.077) | (0.040) | (0.105) |
| **BA** | | | | | | |
| BA × Q1 | −0.071** | −0.264*** | −0.094** | −0.314*** | −0.057 | −0.210*** |
| | (0.033) | (0.053) | (0.039) | (0.079) | (0.055) | (0.054) |
| BA × Q2 | 0.033 | −0.037 | 0.108 | −0.005 | −0.038 | −0.070 |
| | (0.060) | (0.054) | (0.103) | (0.080) | (0.063) | (0.071) |
| BA × Q3 | −0.010 | 0.040 | −0.045 | 0.018 | 0.015 | 0.052 |
| | (0.041) | (0.046) | (0.057) | (0.080) | (0.057) | (0.054) |
| BA × Q4 | 0.070 | 0.037 | 0.022 | 0.004 | 0.099* | 0.056 |
| | (0.047) | (0.038) | (0.093) | (0.063) | (0.054) | (0.048) |
| Observations | 9,662 | 9,662 | 3,919 | 3,919 | 5,743 | 5,743 |
| Number of subjects | 526 | 526 | 215 | 215 | 311 | 311 |

Individual and article fixed effects included in all models.
Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.15: Average marginal effects per quartile: seconds (ln).

| Variable | (1) All (11-15) | (2) All (16-20) | (3) Males (11-15) | (4) Males (16-20) | (5) Females (11-15) | (6) Females (16-20) |
|---|---|---|---|---|---|---|
| **BR-20** | | | | | | |
| BR-20 × Q1 | −0.104** | −0.185* | −0.077 | 0.356 | −0.095* | −0.171* |
| | (0.040) | (0.100) | (0.047) | (0.394) | (0.050) | (0.103) |
| BR-20 × Q2 | −0.002 | 0.050 | 0.050 | 0.185*** | 0.002 | 0.043 |
| | (0.026) | (0.047) | (0.035) | (0.051) | (0.034) | (0.053) |
| BR-20 × Q3 | 0.095* | 0.024 | 0.316** | 0.356 | 0.074 | 0.003 |
| | (0.055) | (0.056) | (0.156) | (0.394) | (0.055) | (0.049) |
| BR-20 × Q4 | 0.082** | 0.091* | 0.091** | 0.113 | 0.077 | 0.075 |
| | (0.037) | (0.049) | (0.043) | (0.072) | (0.055) | (0.065) |
| **BR-2040** | | | | | | |
| BR-2040 × Q1 | −0.109** | −0.324*** | −0.190** | −0.493*** | −0.058 | −0.111* |
| | (0.046) | (0.084) | (0.093) | (0.083) | (0.047) | (0.063) |
| BR-2040 × Q2 | −0.016 | −0.098** | −0.027 | −0.017 | −0.007 | −0.135*** |
| | (0.034) | (0.043) | (0.053) | (0.070) | (0.045) | (0.048) |
| BR-2040 × Q3 | 0.041 | 0.014 | 0.069 | 0.098 | 0.030 | −0.021 |
| | (0.053) | (0.049) | (0.093) | (0.115) | (0.065) | (0.046) |
| BR-2040 × Q4 | 0.052 | 0.074 | −0.044 | 0.044 | 0.128* | 0.091 |
| | (0.049) | (0.057) | (0.062) | (0.106) | (0.067) | (0.058) |
| **BR-40** | | | | | | |
| BR-40 × Q1 | −0.091** | −0.279*** | −0.134*** | −0.126** | −0.059 | −0.334*** |
| | (0.043) | (0.065) | (0.051) | (0.062) | (0.065) | (0.069) |
| BR-40 × Q2 | 0.009 | −0.084** | −0.040 | −0.122 | 0.042 | −0.061 |
| | (0.038) | (0.042) | (0.056) | (0.074) | (0.051) | (0.050) |
| BR-40 × Q3 | −0.001 | 0.054 | −0.068 | 0.044 | 0.036 | 0.062 |
| | (0.035) | (0.039) | (0.054) | (0.063) | (0.046) | (0.050) |
| BR-40 × Q4 | 0.069* | 0.068* | −0.018 | −0.010 | 0.131*** | 0.121** |
| | (0.035) | (0.038) | (0.045) | (0.059) | (0.048) | (0.048) |
| Observations | 9,662 | 9,662 | 3,919 | 3,919 | 5,743 | 5,743 |
| Number of subjects | 526 | 526 | 215 | 215 | 311 | 311 |

Individual and article fixed effects included in all models.
Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.16: Average marginal effects per quartile: points.

| Variable | (1) All (11-15) | (2) All (16-20) | (3) Males (11-15) | (4) Males (16-20) | (5) Females (11-15) | (6) Females (16-20) |
|---|---|---|---|---|---|---|
| **Feedback** | | | | | | |
| Feedback × Q1 | 0.315** | −0.274* | 0.458 | −0.338 | 0.220 | −0.242 |
| | (0.151) | (0.164) | (0.296) | (0.272) | (0.154) | (0.204) |
| Feedback × Q2 | −0.101 | −0.100 | −0.001 | 0.153 | −0.187** | −0.332 |
| | (0.073) | (0.133) | (0.103) | (0.113) | (0.093) | (0.215) |
| Feedback × Q3 | −0.230*** | −0.077 | −0.119 | −0.109 | −0.366*** | −0.037 |
| | (0.086) | (0.125) | (0.096) | (0.218) | (0.136) | (0.101) |
| Feedback × Q4 | −0.215*** | −0.260*** | −0.164* | −0.231* | −0.309*** | −0.249** |
| | (0.072) | (0.096) | (0.093) | (0.120) | (0.091) | (0.114) |
| **Money** | | | | | | |
| Money × Q1 | 0.157 | −0.554* | −0.044 | −0.425 | 0.483** | −0.899*** |
| | (0.186) | (0.323) | (0.226) | (0.442) | (0.218) | (0.093) |
| Money × Q2 | −0.031 | −0.111 | 0.078 | −0.255 | −0.113 | −0.011 |
| | (0.106) | (0.126) | (0.173) | (0.256) | (0.127) | (0.103) |
| Money × Q3 | −0.066 | −0.074 | −0.205 | −0.102 | 0.016 | −0.058 |
| | (0.101) | (0.083) | (0.181) | (0.133) | (0.109) | (0.107) |
| Money × Q4 | −0.299*** | −0.169** | −0.209* | −0.108 | −0.406*** | −0.231** |
| | (0.088) | (0.075) | (0.117) | (0.105) | (0.119) | (0.099) |
| **BA** | | | | | | |
| BA × Q1 | 0.143 | −0.011 | 0.310 | 0.089 | 0.058 | −0.058 |
| | (0.160) | (0.209) | (0.258) | (0.177) | (0.202) | (0.267) |
| BA × Q2 | 0.011 | 0.080 | 0.052 | 0.290** | −0.026 | −0.121 |
| | (0.084) | (0.113) | (0.103) | (0.117) | (0.134) | (0.168) |
| BA × Q3 | −0.085 | −0.054 | −0.015 | 0.055 | −0.145 | −0.144 |
| | (0.089) | (0.087) | (0.094) | (0.102) | (0.146) | (0.130) |
| BA × Q4 | −0.242*** | −0.215*** | −0.184 | −0.162* | −0.277** | −0.247** |
| | (0.084) | (0.073) | (0.126) | (0.096) | (0.111) | (0.101) |
| Observations | 9,662 | 9,662 | 3,919 | 3,919 | 5,743 | 5,743 |
| Number of subjects | 526 | 526 | 215 | 215 | 311 | 311 |

Individual and article fixed effects included in all models.
Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.17: Average marginal effects per quartile: points.

| Variable | (1) All (11-15) | (2) All (16-20) | (3) Males (11-15) | (4) Males (16-20) | (5) Females (11-15) | (6) Females (16-20) |
|---|---|---|---|---|---|---|
| **BR-20** | | | | | | |
| BR-20 × Q1 | 0.457*** | −0.324* | 0.789*** | −0.184 | 0.265** | −0.404* |
| | (0.108) | (0.186) | (0.141) | (0.326) | (0.118) | (0.228) |
| BR-20 × Q2 | −0.051 | 0.041 | −0.024 | −0.242 | −0.069 | 0.101 |
| | (0.075) | (0.120) | (0.143) | (0.282) | (0.095) | (0.132) |
| BR-20 × Q3 | −0.224*** | −0.167** | −0.324** | −0.222 | −0.206** | −0.158 |
| | (0.083) | (0.082) | (0.148) | (0.136) | (0.102) | (0.105) |
| BR-20 × Q4 | −0.211*** | −0.291*** | −0.106 | −0.349** | −0.253** | −0.287*** |
| | (0.079) | (0.082) | (0.119) | (0.145) | (0.103) | (0.101) |
| **BR-2040** | | | | | | |
| BR-2040 × Q1 | 0.242** | −0.451*** | 0.298** | −0.444** | 0.190 | −0.442** |
| | (0.098) | (0.136) | (0.142) | (0.181) | (0.130) | (0.208) |
| BR-2040 × Q2 | 0.119* | 0.014 | 0.170** | 0.065 | 0.086 | −0.017 |
| | (0.068) | (0.099) | (0.086) | (0.125) | (0.097) | (0.138) |
| BR-2040 × Q3 | −0.161* | −0.078 | −0.034 | 0.036 | −0.231* | −0.140 |
| | (0.091) | (0.076) | (0.132) | (0.110) | (0.120) | (0.101) |
| BR-2040 × Q4 | −0.320** | −0.147* | −0.524*** | −0.022 | −0.306** | −0.184* |
| | (0.131) | (0.083) | (0.062) | (0.084) | (0.152) | (0.104) |
| **BR-40** | | | | | | |
| BR-40 × Q1 | 0.216** | −0.091 | 0.113 | −0.052 | 0.271*** | −0.114 |
| | (0.099) | (0.162) | (0.213) | (0.319) | (0.099) | (0.182) |
| BR-40 × Q2 | −0.054 | −0.187 | 0.233* | 0.462*** | −0.127 | −0.315** |
| | (0.082) | (0.119) | (0.124) | (0.119) | (0.101) | (0.134) |
| BR-40 × Q3 | −0.139** | −0.096 | −0.121 | −0.055 | −0.137 | −0.128 |
| | (0.069) | (0.084) | (0.089) | (0.113) | (0.106) | (0.125) |
| BR-40 × Q4 | −0.255*** | −0.238*** | −0.324** | −0.163 | −0.178* | −0.306** |
| | (0.085) | (0.085) | (0.131) | (0.104) | (0.095) | (0.134) |
| Observations | 9,662 | 9,662 | 3,919 | 3,919 | 5,743 | 5,743 |
| Number of subjects | 526 | 526 | 215 | 215 | 311 | 311 |

Individual and article fixed effects included in all models.
Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.18: Survey 1.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | | Instructions | | | Interesting | | | Challenging | |
| Variable | All | Males | Females | All | Males | Females | All | Males | Females |
| Feedback | 0.234 | 0.199 | 0.162 | −0.333 | −0.092 | −0.630 | 0.040 | 0.240 | −0.046 |
| | (0.528) | (0.522) | (0.557) | (0.316) | (0.432) | (0.469) | (0.339) | (0.532) | (0.456) |
| Money | 0.095 | 0.105 | −0.046 | −0.077 | −0.061 | −0.155 | −0.251 | −0.284 | −0.214 |
| | (0.807) | (0.614) | (0.510) | (0.301) | (0.424) | (0.452) | (0.341) | (0.490) | (0.491) |
| BA | −0.311 | −0.757 | 0.027 | −0.538* | −0.864* | −0.271 | −0.298 | −0.230 | −0.334 |
| | (0.342) | (0.511) | (0.440) | (0.312) | (0.457) | (0.425) | (0.296) | (0.469) | (0.396) |
| BR-20 | −0.165 | −0.201 | −0.128 | −0.592* | −0.695 | −0.507 | −0.186 | 0.227 | −0.318 |
| | (0.643) | (0.655) | (0.456) | (0.339) | (0.779) | (0.393) | (0.321) | (0.556) | (0.409) |
| BR-2040 | 0.253 | 0.067 | 0.385 | −0.219 | −0.339 | −0.138 | −0.226 | −0.010 | −0.343 |
| | (0.419) | (0.471) | (0.431) | (0.282) | (0.459) | (0.378) | (0.302) | (0.474) | (0.407) |
| BR-40 | 0.041 | 0.128 | −0.069 | −0.144 | −0.221 | −0.049 | −0.142 | 0.241 | −0.329 |
| | (0.900) | (0.505) | (0.448) | (0.307) | (0.483) | (0.412) | (0.322) | (0.484) | (0.439) |
| R1 Articles | 0.031*** | 0.032* | 0.030* | −0.038*** | −0.037** | −0.037*** | −0.029*** | −0.043*** | −0.020 |
| | (0.009) | (0.018) | (0.016) | (0.009) | (0.015) | (0.013) | (0.010) | (0.016) | (0.013) |
| Observations | 526 | 215 | 311 | 526 | 215 | 311 | 526 | 215 | 311 |
| Pseudo R-squared | | | | | | | | | |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.19: Survey 2.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | Recognition | | | Compensation | |
| (Variable) | All | Males | Females | All | Males | Females |
| Feedback | −0.144 | 0.288 | −0.455 | −0.086 | 0.043 | −0.273 |
| | (0.376) | (0.559) | (0.523) | (0.327) | (0.487) | (0.459) |
| Money | 0.718* | 1.239** | 0.224 | 0.748** | 1.024** | 0.436 |
| | (0.381) | (0.610) | (0.501) | (0.323) | (0.469) | (0.475) |
| BA | −0.237 | 0.234 | −0.607 | −0.577* | −0.612 | −0.528 |
| | (0.341) | (0.519) | (0.456) | (0.308) | (0.465) | (0.422) |
| BR-20 | 0.087 | −0.062 | 0.026 | −0.506 | −0.428 | −0.598 |
| | (0.370) | (0.725) | (0.448) | (0.326) | (0.589) | (0.420) |
| BR-2040 | 0.174 | 0.490 | −0.032 | −0.174 | −0.246 | −0.216 |
| | (0.332) | (0.547) | (0.425) | (0.289) | (0.446) | (0.389) |
| BR-40 | 0.256 | 0.424 | 0.092 | −0.168 | 0.016 | −0.322 |
| | (0.336) | (0.536) | (0.445) | (0.316) | (0.472) | (0.435) |
| R1 Articles | −0.017 | −0.014 | −0.021 | −0.036*** | −0.023 | −0.044*** |
| | (0.011) | (0.018) | (0.014) | (0.011) | (0.018) | (0.015) |
| Observations | 526 | 215 | 311 | 526 | 215 | 311 |
| Pseudo R-squared | | | | | | |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.20: Additional job offer.

| Variable | (1)<br>All | (2)<br>Males | (3)<br>Females |
|---|---|---|---|
| Feedback | −0.338 | 0.282 | −1.024 |
| | (0.514) | (0.682) | (0.787) |
| Money | 0.314 | 0.518 | 0.0876 |
| | (0.603) | (0.789) | (0.953) |
| BA | −0.314 | −0.0277 | −0.663 |
| | (0.496) | (0.687) | (0.789) |
| BR-20 | −0.912* | 0.670 | −1.489** |
| | (0.467) | (0.912) | (0.692) |
| BR-2040 | −0.0373 | −0.355 | 0.422 |
| | (0.495) | (0.664) | (0.844) |
| BR-40 | −0.387 | −0.385 | −0.482 |
| | (0.479) | (0.644) | (0.744) |
| R1 Articles | 0.00231 | −0.0190 | 0.0178 |
| | (0.0155) | (0.0241) | (0.0242) |
| Constant | 0.348 | −0.403 | 2.420*** |
| | (0.875) | (1.313) | (0.841) |
| Observations | 526 | 215 | 309 |
| Pseudo R-squared | 0.039 | 0.057 | 0.083 |

Controls for gender, age, and frequency of playing video games included in all models.
Two observations are dropped from the analysis due to collinearity (females between 18 and 20 years old in our sample always accept the additional job offer).
Robust standard errors (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.21: Adjusted predictions.

| Variable | (1)<br>All | (2)<br>Males | (3)<br>Females |
|---|---|---|---|
| Control | 0.891*** | 0.833*** | 0.930*** |
| | (0.0362) | (0.0692) | (0.0400) |
| Feebdack | 0.854*** | 0.868*** | 0.828*** |
| | (0.0450) | (0.0555) | (0.0710) |
| Money | 0.918*** | 0.893*** | 0.936*** |
| | (0.0361) | (0.0587) | (0.0460) |
| BA | 0.857*** | 0.829*** | 0.873*** |
| | (0.0404) | (0.0707) | (0.0524) |
| BR-20 | 0.767*** | 0.907*** | 0.751*** |
| | (0.0529) | (0.0650) | (0.0630) |
| BR-2040 | 0.887*** | 0.777*** | 0.953*** |
| | (0.0318) | (0.0766) | (0.0255) |
| BR-40 | 0.847*** | 0.772*** | 0.892*** |
| | (0.0387) | (0.0686) | (0.0415) |
| Observations | 526 | 215 | 309 |

Controls for gender, age, and frequency of playing
video games included in all models.
Adjusted predictions obtained after all other.
covariates were fixed at the corresponding sample mean.
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.22: Articles submitted in round 2.

| Variable | (1)<br>All | (2)<br>All | (3)<br>Males | (4)<br>Males | (5)<br>Females | (6)<br>Females |
|---|---|---|---|---|---|---|
| Feedback | −0.0628 | −0.268 | 0.950 | 0.866 | −0.988 | −1.509 |
| | (0.604) | (0.788) | (0.634) | (0.663) | (1.059) | (1.474) |
| Money | 0.327 | 0.306 | 0.534 | 0.180 | 0.198 | 0.354 |
| | (0.490) | (0.484) | (0.674) | (0.711) | (0.692) | (0.653) |
| BA | 0.833* | 0.779* | 1.703** | 1.617** | 0.286 | 0.231 |
| | (0.457) | (0.470) | (0.658) | (0.694) | (0.670) | (0.661) |
| BR-20 | 0.536 | 0.471 | 0.320 | 0.170 | 0.673 | 0.647 |
| | (0.526) | (0.530) | (1.114) | (1.184) | (0.619) | (0.623) |
| BR-2040 | 1.124** | 1.109** | 1.082 | 0.850 | 1.117* | 1.151* |
| | (0.441) | (0.466) | (0.733) | (0.755) | (0.580) | (0.590) |
| BR-40 | 1.088** | 0.992** | 1.752** | 1.533* | 0.760 | 0.740 |
| | (0.447) | (0.457) | (0.710) | (0.782) | (0.596) | (0.603) |
| R1 Articles | 0.584*** | 0.639*** | 0.559*** | 0.626*** | 0.605*** | 0.644*** |
| | (0.0216) | (0.0516) | (0.0376) | (0.0626) | (0.0278) | (0.0689) |
| Feedback × R1 Articles | | −0.146 | | 0.00617 | | −0.302 |
| | | (0.177) | | (0.0892) | | (0.317) |
| Money × R1 Articles | | −0.166** | | −0.203** | | −0.141 |
| | | (0.0717) | | (0.0918) | | (0.0902) |
| BA × R1 Articles | | 0.00185 | | 0.0398 | | −0.00244 |
| | | (0.0625) | | (0.0871) | | (0.0828) |
| BR-20 × R1 Articles | | −0.0561 | | −0.0783 | | 0.00485 |
| | | (0.0690) | | (0.103) | | (0.0909) |
| BR-2040 × R1 Articles | | −0.0160 | | −0.112 | | 0.0165 |
| | | (0.0636) | | (0.0856) | | (0.0836) |
| BR-40 × R1 Articles | | −0.0436 | | −0.0496 | | −0.0368 |
| | | (0.0618) | | (0.111) | | (0.0758) |
| Constant | 15.65*** | 15.58*** | 15.16*** | 14.78*** | 16.20*** | 16.20*** |
| | (1.592) | (1.623) | (1.840) | (2.026) | (2.576) | (2.612) |
| Observations | 444 | 444 | 177 | 177 | 267 | 267 |
| R-squared | 0.762 | 0.769 | 0.769 | 0.780 | 0.769 | 0.781 |

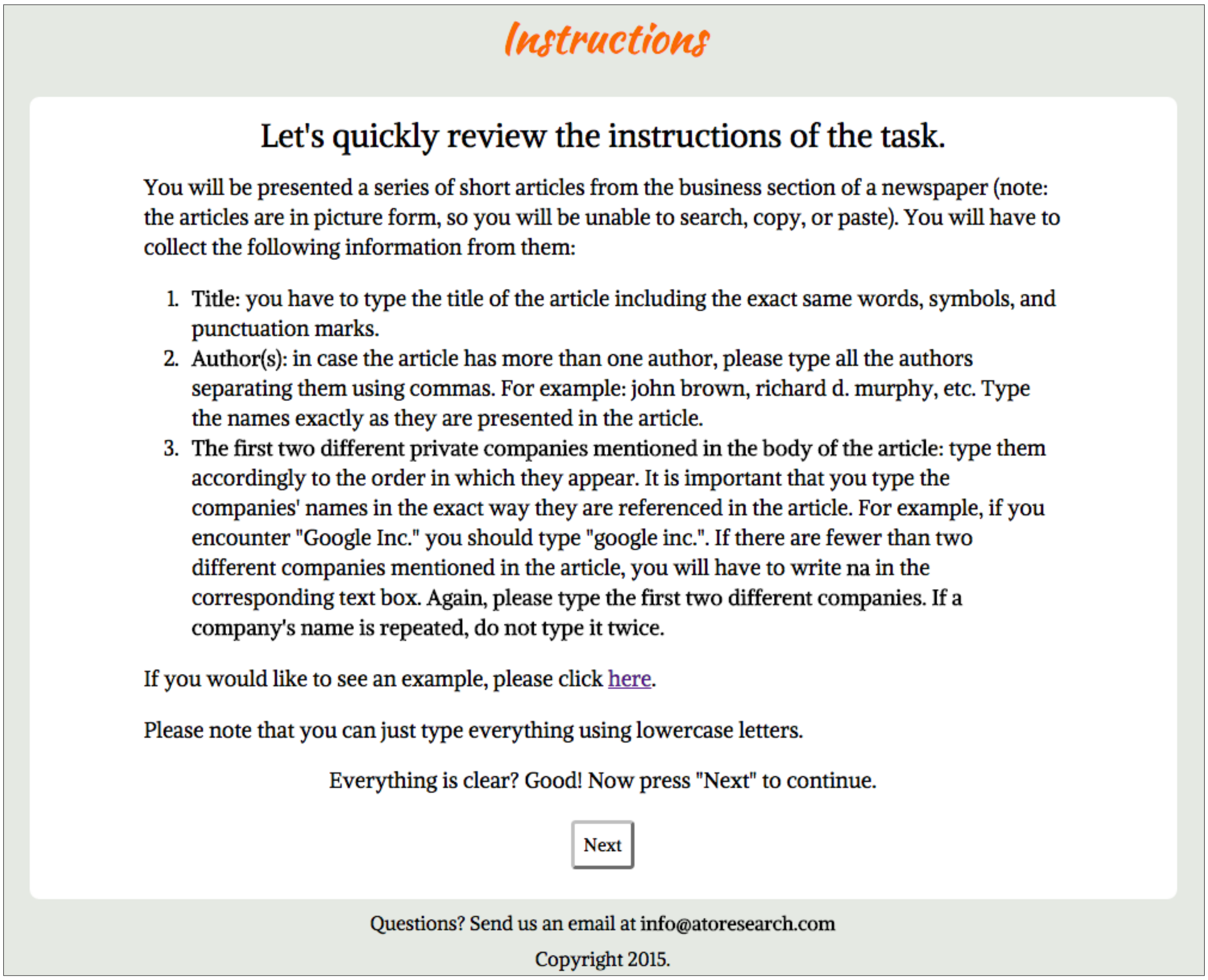
Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.23: Points obtained in round 2.

| (Variable) | (1) (All) | (2) (All) | (3) (Male) | (4) (Male) | (5) (Female) | (6) (Female) |
|---|---|---|---|---|---|---|
| Feedback | −1.566 | −2.164 | 1.977 | 1.584 | −3.687 | −6.013 |
| | (2.252) | (2.661) | (2.688) | (2.759) | (3.793) | (5.823) |
| Money | 0.00563 | −0.0457 | 1.633 | 0.550 | −1.277 | −1.029 |
| | (1.685) | (1.663) | (2.491) | (2.695) | (2.238) | (2.072) |
| BA | 2.334 | 2.081 | 6.735** | 6.326** | −0.334 | −0.480 |
| | (1.747) | (1.727) | (2.664) | (2.684) | (2.465) | (2.420) |
| BR-20 | 0.703 | 0.542 | 1.715 | 0.849 | 0.412 | 0.353 |
| | (1.847) | (1.774) | (3.547) | (3.359) | (2.195) | (2.157) |
| Br-2040 | 4.195** | 4.064** | 6.462** | 5.321* | 2.777 | 2.864 |
| | (1.664) | (1.729) | (2.810) | (2.737) | (2.099) | (2.134) |
| BR-40 | 3.049* | 2.851* | 7.627*** | 6.917** | 0.708 | 0.703 |
| | (1.685) | (1.689) | (2.867) | (3.040) | (2.108) | (2.104) |
| R1 Points | 0.641*** | 0.745*** | 0.590*** | 0.704*** | 0.674*** | 0.765*** |
| | (0.0218) | (0.0641) | (0.0347) | (0.0856) | (0.0292) | (0.0833) |
| Feedback × R1 Points | | −0.215 | | −0.0672 | | −0.349 |
| | | (0.181) | | (0.107) | | (0.332) |
| Money × R1 Points | | −0.151* | | −0.216* | | −0.114 |
| | | (0.0811) | | (0.112) | | (0.101) |
| BA × R1 Points | | −0.0466 | | −0.0114 | | −0.0511 |
| | | (0.0746) | | (0.108) | | (0.0996) |
| BR-20 × R1 Points | | −0.109 | | −0.0950 | | −0.0881 |
| | | (0.0772) | | (0.111) | | (0.0963) |
| BR-2040 × R1 Points | | −0.0991 | | −0.206* | | −0.0594 |
| | | (0.0803) | | (0.110) | | (0.102) |
| BR-40 × R1 Points | | −0.128* | | −0.176 | | −0.106 |
| | | (0.0725) | | (0.131) | | (0.0889) |
| Constant | 51.33*** | 51.36*** | 50.20*** | 48.17*** | 52.17*** | 52.57*** |
| | (4.956) | (4.955) | (8.709) | (9.809) | (3.132) | (2.961) |
| Observations | 444 | 444 | 177 | 177 | 267 | 267 |
| R-squared | 0.774 | 0.779 | 0.747 | 0.758 | 0.801 | 0.810 |

Controls for gender, age, and frequency of playing video games included in all models.
Robust standard errors (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1.

Figure 1.1: Instructions



## Instructions

### Let's quickly review the instructions of the task.

You will be presented a series of short articles from the business section of a newspaper (note: the articles are in picture form, so you will be unable to search, copy, or paste). You will have to collect the following information from them:

1. Title: you have to type the title of the article including the exact same words, symbols, and punctuation marks.
2. Author(s): in case the article has more than one author, please type all the authors separating them using commas. For example: john brown, richard d. murphy, etc. Type the names exactly as they are presented in the article.
3. The first two different private companies mentioned in the body of the article: type them accordingly to the order in which they appear. It is important that you type the companies' names in the exact way they are referenced in the article. For example, if you encounter "Google Inc." you should type "google inc.". If there are fewer than two different companies mentioned in the article, you will have to write na in the corresponding text box. Again, please type the first two different companies. If a company's name is repeated, do not type it twice.

If you would like to see an example, please click here.

Please note that you can just type everything using lowercase letters.

Everything is clear? Good! Now press "Next" to continue.

Next

Questions? Send us an email at info@atoresearch.com

Copyright 2015.

Figure 1.2: Digital badge (BA) presentation.

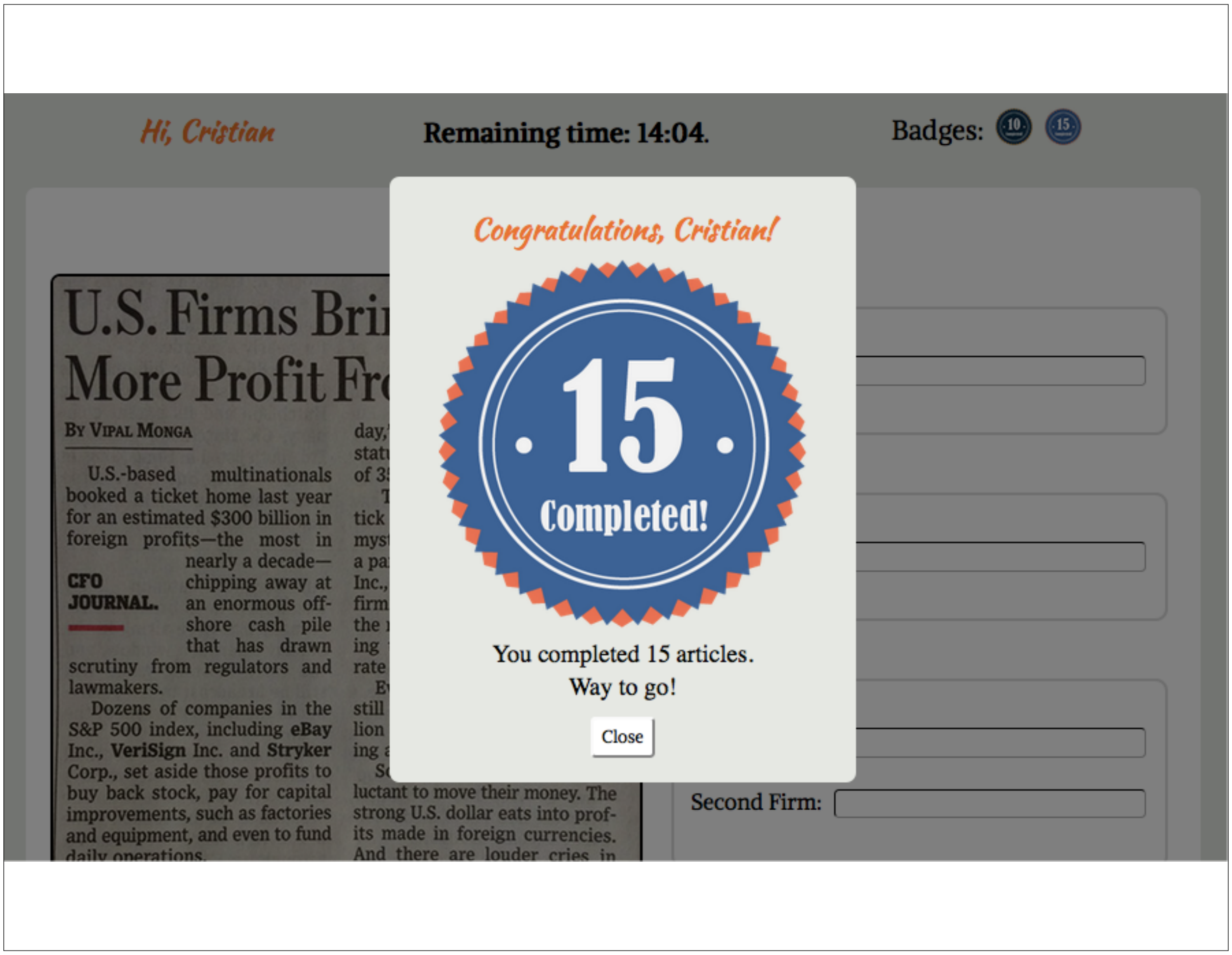


Figure 1.3: Badges

Figure 1.4: Histograms of the number of articles submitted by round.

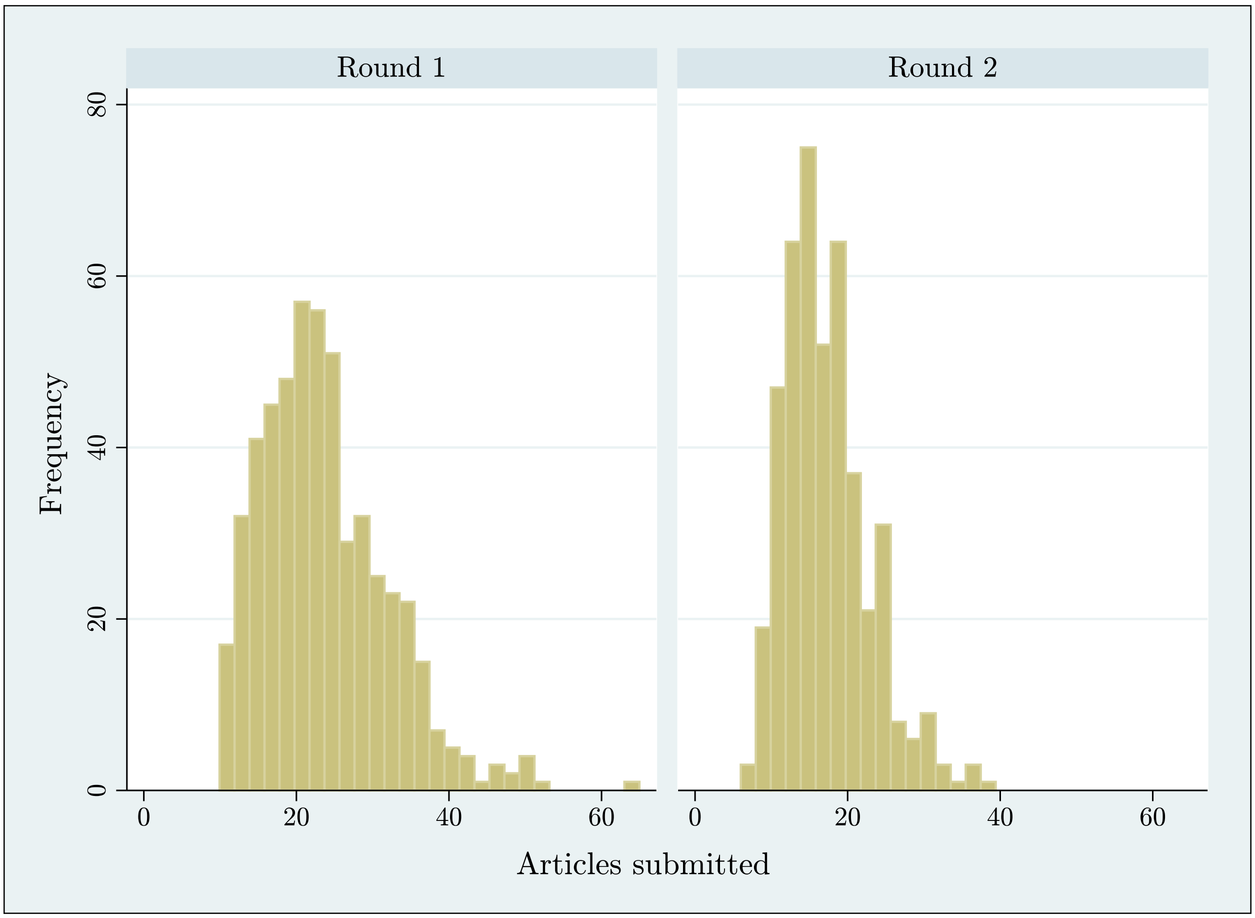


Figure 1.5: Scatterplot: articles submitted during rounds 1 and 2.

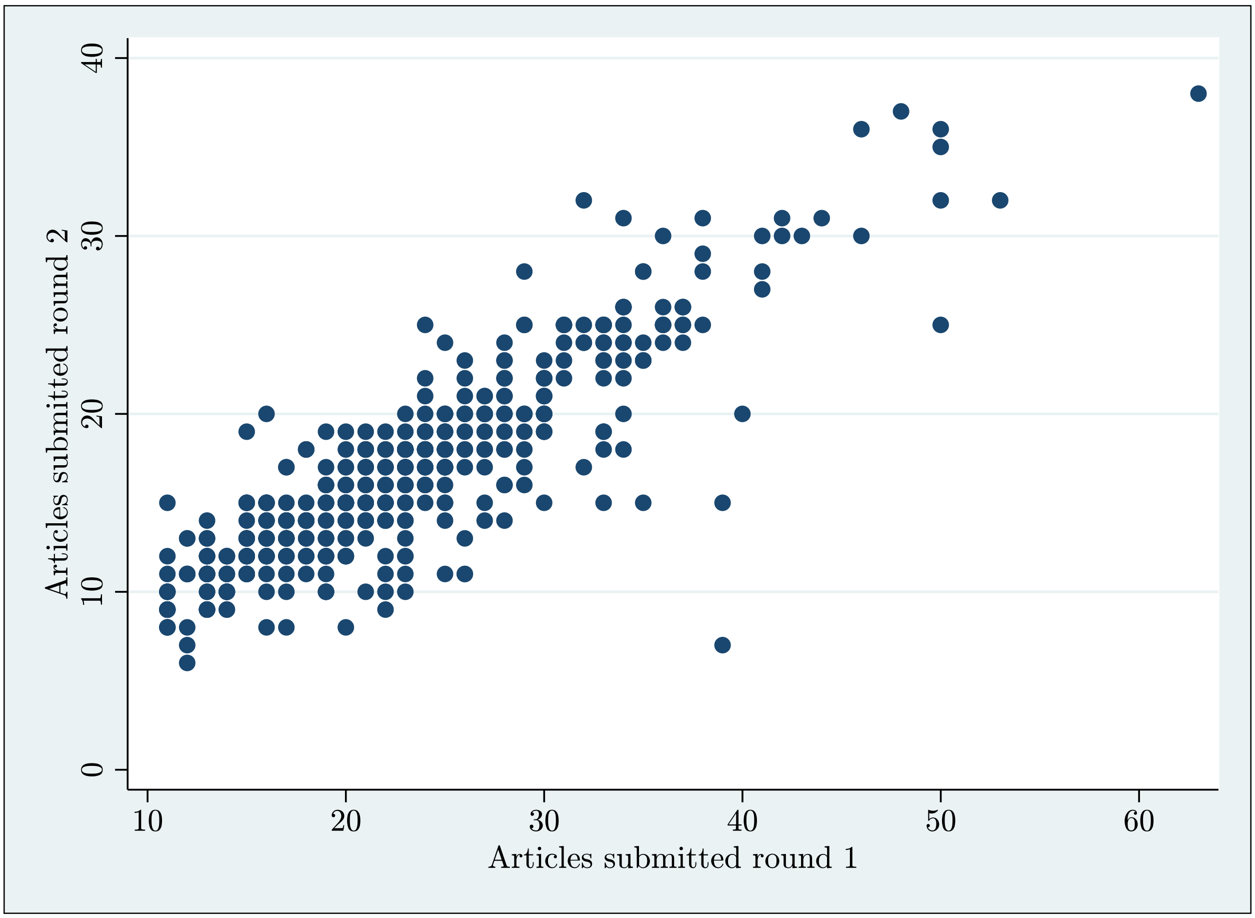
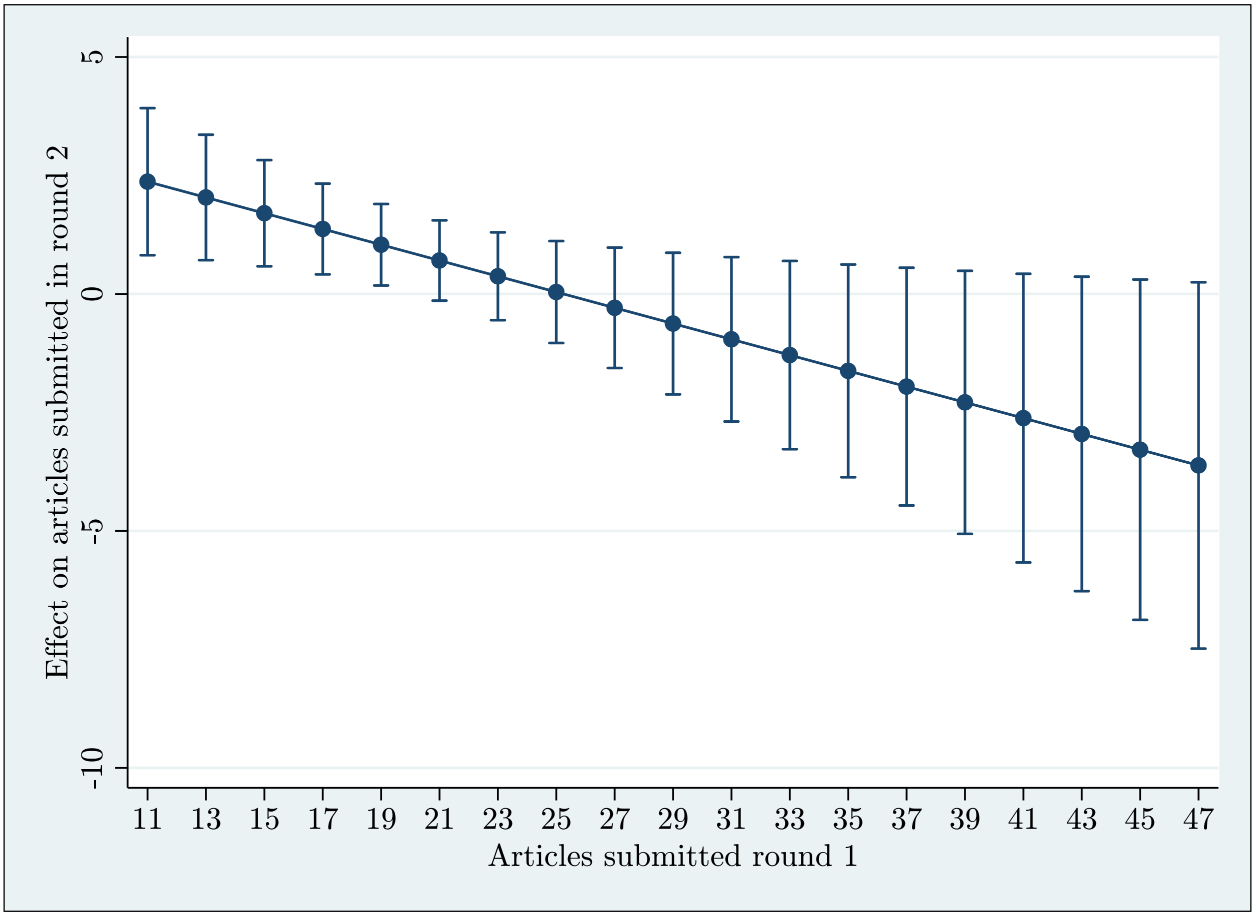


Figure 1.6: Average marginal effect of money condition on articles submitted in round 2.



## 1.8 Appendix: Survey questions

At the end of the first 25 minutes of work, subjtects had to complete a short survey in order to get the completion code for the task (which they will subsequently paste into a box on the AMT website as a proof that they worked on the task). To get the code, every question had to be answered. Below we present the questions and their potential answers in the same order that subjects received them.

1. **How clear were the instructions of this HIT?**
   - Extremely clear.
   - Very clear.
   - Moderately clear.
   - Slightly clear.
   - Not at all clear.
2. **How interesting was the HIT?**
   - Extremely interesting.
   - Very interesting.
   - Moderately interesting.
   - Slightly interesting.
   - Not at all interesting.
3. **How challenging was the HIT?**
   - Extremely challenging.
   - Very challenging.
   - Moderately challenging.
   - Slightly challenging.
   - Not at all challenging.
4. **How often do you read a daily newspaper, either online or in print?**
   - Almost every day.
   - At least once a week.
   - A few times a month.
   - A few times a year.
   - Never.
5. **The requester of this HIT recognizes good job performance.**
   - Strongly agree.

- Agree.
- Neutral / Neither agree nor disagree.
- Disagree.
- Strongly disagree.

6. **I am satisfied with the overall compensation I received for working on this HIT.**

   - Strongly agree.
   - Agree.
   - Neutral / Neither agree nor disagree.
   - Disagree.
   - Strongly disagree.

7. **Are you male or female?**

   - Male.
   - Female.

8. **What is your age?**

   - 18-20.
   - 21-29.
   - 30-39.
   - 40-49.
   - 50 or older.

9. **What is the highest level of school you have completed or the highest degree you have received?**

   - Less than high school degree.
   - High school degree or equivalent (e.g., GED).
   - Some college but no degree.
   - Associate degree.
   - Bachelor degree.
   - Graduate degree.

10. **Which of the following categories best describes your employment status?**

    - Employed, working full-time.
    - Employed, working part-time.
    - Not employed, looking for work.
    - Not employed, NOT looking for work.
    - Retired.
    - Disabled, not able to work.

11. **Are you White, Black or African-American, American Indian or Alaskan Native, Asian, Hispanic or Latino, Native Hawaiian or other Pacific Islander, or some other race?**
    - White.
    - Black or African-American.
    - American Indian or Alaskan Native.
    - Asian.
    - Hispanic or Latino.
    - Native Hawaiian or other Pacific Islander.
    - From multiple races.
12. **What is your approximate annul income?**
    - $0 - $24,999.
    - $25,000 - $49,999.
    - $50,000 - $74,999.
    - $75,000 - $99,999.
    - $100,000 - $199,999.
    - $200,000 and up.
13. **How many hours per week do you usually work on HITs?**
    - 1 to 5 hours.
    - 6 to 9 hours.
    - 10 to 20 hours.
    - 21 to 29 hours.
    - 30 hours or more.
14. **What is the main reason you complete HITs on Amazon Mechanical Turk?**
    - Income purposes (primary source of income).
    - Pocket change / extra cash.
    - Fruitful way to spend free time.
    - It is fun.
    - To kill time.
15. **How often do other Amazon Mechanical Turk requesters give you feedback about your work?**
    - Extremely often.
    - Very often.
    - Moderately often.
    - Slightly often.
    - Not at all often.

16. **How often do you play video games?**

    - Almost every day.
    - At least once a week.
    - A few times a month.
    - A few times a year.
    - Never.

# CHAPTER 2

# Motivating truck drivers: The role of points, badges, prizes, and t-shirts on fuel efficiency

## 2.1 Introduction

Gamification has been mainly understood as the use of game elements in contexts different from full games (Deterding et al., 2011), with the main objective of altering user behavior (Hamari and Koivisto, 2013). Other definitions accentuate the affordance of gameful experiences or the use of a design that resembles games with the objective of fostering the value creation process (Hamari, Huotari, and Tolvanen, 2015; Huotari, 2012). Although researchers have not reached a complete agreement with respect to what constitutes gamification, or even the minimum characteristics that a *gamified* application should contain (Seaborn and Fels, 2015), this has not stopped firms from implementing and developing gamification applications to motivate and engage consumers and employees. The evidence on the effect of gamification applications on subject behavior has been mostly positive, although results vary depending on the context of the implementation and the final users of it (Hamari, Koivisto, and Sarsa, 2014; Seaborn and Fels, 2015). Studies on the effectiveness of digital badges and achievements have been carried out in topics such as Internet forums (Grant and Betts, 2013), social media websites (Easley and Ghosh, 2013), photo-sharing services (Montola et al., 2009), online newspapers (Jones, Altadonna, and Lindsey, 2012), marketing (Huotari, 2012), and educational platforms (Abramovich, Schunn, and Higashi, 2013; Domínguez et al., 2013; Gibson et al., 2013; Haaranen et al., 2014; Landers and Callan, 2011).. Gartner, an information technology research company, forecasted that by 2014 around 70 percent of the Forbes Global 2000 firms would have applied gamification to some extent (Gartner Inc., 2012a). However, Gartner also predicted that about eighty percent of those applications would fail as consequence of poor design (Gartner Inc., 2012b).

In order to understand the possibilities gamification brings for firms, a first step is to turn our attention to what a game is and its essential elements. Although gamification does not consider the introduction of a full game, as is the case with has been called in the literature serious games or edutainment in the case of games that are developed with educational goals in mind (Domínguez et al., 2013), knowing the characteristics aspects of games will help us understand what gamification is and how it can be used inside businesses. According to Mcgonigal (2011), all games share the following four common components: *goals*, *rules*, *feedback system*, and *voluntary participation* (p. 21). The goals are the objectives that players intend to reach. The rules are the norms that limit the ways goals can be achieved. The feedback system provides players with information on how close they are to accomplish the goals. Finally, voluntary participation means that each player accepts the goals, rules, and the feedback system in place. According to Koivisto and Hamari (2014), gamification distinguishes itself from a game in the fact that gamification is used to achieve an objective unrelated to a game while playing a game is an intrinsically motivated action that has a purpose in itself. Although a carefully designed game can have some positive consequences for an organization that incentivizes its member to play it (Smith, 2012), we do not consider this kind of intervention here. An example of this latter situation is depicted in Mollick and Rothbard (2014), where the authors study how consent moderates the relationship between the application of a full game in a work environment —an approach that Mollick and Rothbard call gamification— and its impact on worker performance. Some authors have discussed the benefits of using games inside a company, highlighting how through games employees can learn to trust each other and develop cooperative behavior (Smith, 2012). Gamification, on the other hand, does not make use of full games but of game elements to increase, for example, the engagement of subjects with the tasks they are performing.

In general, what differentiates gamification from a full game is the idea that the creator of the gamified application uses elements from games in its system instead of a complete game. However, the presence of mere game elements —and the absence of a complete game— does not, necessarily, preclude the users of the platform to experience it as *gameful* (Deterding et al., 2011). According to some authors, there is no clear definition of what a 'game element' is, since most of the typical components that one finds in games such as badges, levels, marketplaces, etc. already exist outside games and these do not appear 'gameful' when considered independently (Deterding et al., 2011).[1]

[1] Badgeville, one of the leader companies in the gamification industry, indicates in one post on their online blog that gamification is the use of game mechanics and rewards in non-game settings (`https://badgeville.com/gamification-101-gaming-the-system/`). Badgeville gives the following examples of mechanics: points, progress bars, badges, and leaderboards. If we compare this to the list of 51 items

Although most of the research on gamification do not discuss theoretical foundations behind their interventions (Seaborn and Fels, 2015), the majority of authors who do mention them usually refer to the literature on intrinsic and extrinsic motivation (Deci, Koestner, and Ryan, 1999; Seaborn and Fels, 2015) as the central theory of motivation behind gamification. Although other authors (Ferrell et al., 2016) have also related gamification with other motivation theories such as flow theory (Csikszentmilhalyi and Lefevre, 1989) and goal-setting theory (Locke and Latham, 2002), it seems that the actual condition of the research in gamification favors the idea of gamification focusing on the intrinsic values and needs of the users of the applications (Seaborn and Fels, 2015). Przybylski, Rigby, and Ryan (2010), when analyzing the appealing of video games, claim that they interest generated by video games in people can be explained by the satisfaction the psychological needs of competence, autonomy, and relatedness.

Critics of gamification have been questioning the focus on leaderboards, points, and badges instead of the other attributes of games, an approach that has been called pointsification. Robertson (2010) mentions that gamification just takes what is the least interesting thing about games (e.g., points or badges) and presents it as the essential element of the experience of playing games. However, it is hard to argue that by simply adding some of these features to an external activity where rewards are predictable, worker performance —or other behaviors— would change persistently (Fizek, 2014). According to Seaborn and Fels (2015), around 60 percent of the papers on gamification they surveyed show positive effects of gamification (studies were about various topics such as sustainability, education, marketing, etc.). Seaborn and Fels also recognize that a 'file-drawer' effect might be in place, in which only 'good' results made it to a publication while other never see the light. Additionally, various studies have found that the effects of gamification tend to be temporal and mostly based on 'novelty' (Koivisto and Hamari, 2014; Seaborn and Fels, 2015). Therefore, it is important to analyze empirical evidence of the effect of gamification to see whether the sole introduction of game elements is sufficient to elicit a response from subjects that does not fade away after a while, but that can help modify behaviors in more permanent ways. Also, scholars have mentioned the need to integrate the study of gamification with other areas of knowledge, since most of the discussion of gamification has occurred in game studies

that the website boardgamegeek.com, one of the largest online communities on board gaming, has for the category game mechanics, we can see a huge difference. While for Badgeville game mechanics are just ways to transmit feedback to the user, for board gamers mechanics refer to the procedures that designers choose to implement in their games and determine the way players interact with it.

(Hamari, Huotari, and Tolvanen, 2015).

The paper is organized as follows. Section 2.2 describes our empirical setting while Section 2.4 describes the database and our empirical strategy. Section 2.5 shows the results of our empirical analyses. Section 2.6 discusses the main findings, the limitations of the present study, and future lines of research. Section 2.7 concludes.

## 2.2 Empirical setting: The trucking industry

In the literature, the relationship between drivers and trucking companies has been characterized as one between agents and a principal, with drivers interested in maximizing their utility, which depends positively on income and other benefits and negatively on effort, while trucking companies maximize profits (Hubbard, 2000). As explained by Hubbard (2000), drivers are one of the main determinants of the cost of transporting goods since the way drivers operate trucks affect the time at which they arrive at their destination and the wear of the equipment. Baker and Hubbard (2004) discuss how the cost of wear and tear of trucks is kept low when drivers maintain a steady speed, but Baker and Hubbard recognize how drivers might have different preferences. For example, driver might prefer to driver faster and then rest for a longer period of time. Also, different driving techniques affect one of the main costs for trucking companies (besides salary): fuel consumption. According to Hubbard, the costs for the trucking company increase (convexly) in the speed at which workers decide to drive their trucks. Additionally, the costs (of effort) of arriving on time and driving efficiently are born by drivers alone (Hubbard, 2000). Then, in an industry affected by a shortage of drivers for the last 15 years and with high rotation (Costello and Suarez, 2015), trucking firms are looking for ways to make workers drive more efficiently (without putting more pressure on wages) and keep them satisfy or engage so to decrease the chances they leave the organization.

Different approaches have been tested. Tulusan, Staake, and Fleisch (2012) show that eco-feedback technologies can have a positive impact on driving behavior (in terms of fuel consumption) even when drivers do not bear the cost of fuel or do not have financial incentives to become more efficient drivers. Tulusan, Staake, and Fleisch (2012) use a smartphone application that provides real-time feedback regarding three main areas: acceleration, braking, and speed using a score from zero to one-hundred (a higher score represents a more 'ecological' driving style). Tulusan, Staake, and Fleisch (2012) use a self-selected sample of corporate car drivers (42 in total) from a single company in Switzerland. Car drivers that had access to the smartphone application showed an increment in their fuel efficiency, measured as liters of fuel per 100 kilometers, of 3.2 percent. Tulusan, Staake, and Fleisch (2012) do not offer an explanation to this effect besides mentioning the idea that persuasive feedback technologies can have an effect on driving behavior. They also mention that acceleration has a strong effect on fuel consumption. Siero et al. (1989) evaluate the effects of a program with the objective of reducing fuel consumption. The initiative included the following features: information, feedback, and control. Siero et al. have a quasi-experimental design, with two groups of drivers (one control and treatment) Siero et al. claim that compared to the control group, the experimental group reduced its consumption by 7.3 percent. In this case, it seems that management involvement was key, as some qualitative data suggest that instructions and talks by managers increased fuel efficiency. Five months after the end of the program, there was still a difference between the experimental group and the treatment group (5.5 percent). Harvey, Thorpe, and Fairchild (2013) show results from focus groups in which drivers expressed that feedback might be an incentive to drive more efficiently and that saving time was more important than saving money. Based on these results, it is possible that an application that considers feedback and other game elements might prove effective on fuel efficiency.

## 2.3 The platform: *FuelOpps*

*FuelOpps* is an application developed by Propel IT, Inc. (Propel IT hereafter) that helps improve the performance of truck drivers. The algorithm that powers *FuelOpps* analyzes multiple data inputs and translate them into scores that provide information to drivers regarding their driving skills and consumption of fuel as a consequence of idling. Scores are in a zero to ten scale and are accompanied by information about a driver's relative position among their peers. By improving their driving skills, operators will drive more efficiently, reducing the vehicle's wear and tear and the consumption of fuel. The premise of *FuelOpps* is that by focusing on the training of some specific driving skills, the industry's key metric for fuel consumption —miles per gallon (MPG)— will improve automatically.

### 2.3.1 How does *FuelOpps* work?

Besides providing performance information to drivers in the form of scores tied to their driving skills, *FuelOpps* also offers real-time support to drivers in the form of coaches that can be reached via phone, e-mail, or Qualcomm messaging system for feedback on driving habits and ways to improve scores.[2] Trucking companies can determine whether they allow drivers to be contacted by coaches, and if that is the case, coaches are usually very active in terms of providing scores to drivers, offering help to discuss practical ways to polish skills, and answering their questions. Also, *FuelOpps* supplies fleet managers with data about drivers and their scores so to help with the supervision and administration of drivers.

#### 2.3.1.1 The RPM matrix

The core of *FuelOpps* is the revolutions-per-minute (RPM) matrix. The RPM matrix is a visual representation of the possible combinations of ranges of speed and RPM that could occur while the engine of a truck is operating and the actual time spent on each RPM-speed cell. To minimize fuel consumption, an engine should be running at the minimum RPM possible, since any other combination represents just fuel lost. In the 2.0 version of the algorithm, Propel IT has divided the RPM matrix into zones (green, yellow, and red). The evaluation of each of the driving skills in *FuelOpps* is determined using the time drivers spend in each of the zones. More time spent on green zones indicate better driving skills, while time spent on red zones mean that fuel was consumed unnecessarily. In the 1.0 version of the algorithm, which was the one available during the period covered by the database analyzed in this study, there were only two zones: a green zone and a 'bad' zone. For any given speed, RPMs higher than a certain threshold were considered 'bad' and penalized with lower scores. In the 1.0 version, thresholds were the same for all firms, independent of the nature of their operation. This was later changed in the 2.0 version of the algorithm, and this rendered scores obtained in versions 1.0 and 2.0 incomparable between each other.

[2]The Qualcomm messaging system is directly connected to the truck's computer and allows trucking companies to send messages to drivers while they are operating the trucks. Also, the Qualcomm device comes with a GPS, so trucking companies can monitor where the truck is at any moment in time.

#### 2.3.1.2 Driving skills and scores

Based on the RPM matrix, up to four different scores are computed for each driver depending on the type of truck she drives. For manual transmission trucks, the dimensions scored are: progressive shifting, high RPM, highest gear, and highway speed. Progressive shifting evaluates how efficient a driver changes shifts while accelerating. High RPM assesses hard acceleration (i.e., events in which a driver accelerates her truck in a very short time span, sharply increasing the engine's RPM and consuming more fuel than necessary if the driver had chosen to increase speed slowly). Highest gear grades whether the truck is kept at the lowest possible RPM given the current speed. Finally, highway speed checks the maximum speed at which the trucks is operated (speeds over sixty-two or sixty-three miles per hour are related to higher RPM and fuel consumption). On the other hand, for automatic transmission trucks, the dimensions are acceleration, kickdown, and highway speed. The definition of these dimensions is analogous to the ones for manual-transmission trucks, with the exception to kickdown, which evaluates a driver's ability to avoid her truck to downshift by slowing down when possible. The scores for each of the dimensions are computed using the RPM matrix, and they are a function of the time a driver spent on the green and red zones described above. As a rule of thumb, the higher the time spent on green zones, the better a driver's score for each dimension would be.

Under the 1.0 version of *FuelOpps*, the range for the score of each dimension was from zero to 100. The scores for the four dimensions in manual-transmission trucks (and the three dimensions considered in the case of automatic-transmission trucks) are further summarized in an overall score that aims to represent an aggregate measure of a driver's driving skills using the dimensions available. This overall score plays a large role in the determination of the points allocated to each driver at the end of each day.

#### 2.3.1.3 Messages and calls

*FuelOpps* coaches can contact drivers directly using the trucks' on-board computers. Also, by sending messages to the truck, coaches make sure drivers will receive the message since the driver needs to be logged in for the message to arrive (otherwise it is not delivered). The degree to which coaches can interact with drivers varies from company to company, although in general coaches can respond to drivers questions or other inquiries. When *FuelOpps* coaches have permission to contact drivers by either text messages or phone calls, coaches are generally very active; they would send information about the program and scores obtained in the last period, details about any new or running marketing campaigns, and some general advice on how to improve certain scores. Additionally, coaches may also send congratulation messages that can be either directed to the driver (e.g., "We really like how much better your high RPM score got this week") or with the aim of giving public recognition to a specific driver in a group (e.g., "Congratulations to *first_name last_name*. He had the most improved FuelOpps score in your group during the past week..."). Drivers can respond to the messages sent by coaches. Drivers also have the chance to opt out of receiving more messages if they want to.[3]

Drivers can also call coaches in case they have questions about the *FuelOpps* program or if they need advice to improve their driving skills. Besides using the smartphone application or the website to check their scores, driver can also get them through an interactive voice response (IVR) system.

### 2.3.2 Gamification features of ***FuelOpps***

#### 2.3.2.1 Points

When companies decide to implement *FuelOpps*, they can choose whether to enable a point system tied to drivers' scores and their number of miles driven. In some cases, the points a driver accumulates can be redeemed for prizes (if the truck company also agrees to have a catalog of items managed by Propel IT). Trucking companies have the freedom to create their own campaigns. For example, a company might decide to give pecuniary prizes to the top drivers according to their overall ranking at the end of a specific date.

If a company agrees to use a point system, Propel IT would suggest an exponential function to allocate points to each driver ($i$) on each day she drives ($t$).[4] The point allocation function for a driver rests on six different variables; with some of them being relative to the group ($j$) a driver belongs to: the lowest ($r_{0j}$) and highest ($r_{1j}$) percentile of drivers that will receive points, the minimum ($p_{0j}$) and maximum ($p_{1j}$) number of points per thousand miles

[3] Although the database contains the messages sent to drivers, there is no way to actually determine whether the messages were read.

[4] This way to allocate points was used during the 1.0 version of the *FuelOpps* algorithm. Under the 2.0 version, the exponential function was replaced by a step function that, similarly, awards an increasing proportion of points to the top drivers.

that a driver could obtain, the total number of miles driven ($m_{it}$), and a driver's relative rank ($r_{it}$) among members of her group. To rank a driver in its group on a given day, Propel IT uses the driver's overall score, which is stated to be independent of the total number of miles she drove on that day. A score for a given day is generated only if the number of miles driven by a driver is higher than a pre-defined threshold, which is usually around 25 miles for most companies. This assures that there is a minimum timeframe under which performance can be measured and evaluated. After all scores for a group are computed, a driver's measure of relative ranking is given by the percentile rank of her score.[5]

The point allocation function is presented below.

$$f(m_{it}, r_{it}, r_{0j}, r_{1j}, p_{0j}, p_{1j}) = \begin{cases} \left(\frac{m_{it}}{1{,}000}\right) p_{1j} \left[\exp\left\{\ln\left(\frac{p_{1j}}{p_{0j}}\right)\left(\frac{r_{it}-r_{1j}}{r_{1j}-r_{0j}}\right)\right\}\right], & \text{if } r_{0j} \leqslant r_{it} \leqslant r_{1j} \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Companies typically choose to award points to drivers in the top 45 ($r_{0j} = 55$) or 35 percent ($r_{0j} = 65$) within each group. The highest percentile of workers is usually fixed across companies and groups and it is set to one-hundred ($r_{1j} = 100$), so all top drivers —specially those on the right tail of the distribution— receive points.

Figure 2.1 shows the point allocation function for a group $j$ in a company with $r_{0j} = 55$, $r_{1j} = 100$, $p_{0j} = 10$, and $p_{1j} = 1,000$ for four different values of $miles$ traveled (100, 300, 500, and 700). We can see how there is an increasing amount of points allocated to the top-ranked drivers. Given the way the allocation function is defined, the percentage of points allocated to the top ten and top five percent of drivers is the same under any assumption of miles driven (sixty-five and forty percent, respectively). The high proportion of points awarded to top performers is meant to increase competition by drivers. Assuming a convex disutility of effort for drivers, rewards should increase exponentially in order to incentivize employees to drive better.

[5]According to Propel IT, once the relevant data is obtained from the telematics company, the scores and points for each driver are computed overnight and become available to drivers in the first hours of the following day. However, in the rare cases where the data from a specific driver is not received, which could happen as consequence of faulty equipment on the truck or because the telematics company releases various days of data in a row instead of providing daily observations, the score calculations cannot be performed. This is because of the relative nature of the way points are computed. When a situation like this occurs, Propel IT notifies drivers of their points as soon as the data for all drivers in the group is received. As stated by Propel IT, this is a rare event and it is estimated to affect around one percent of the observations in the database.

#### 2.3.2.2 Badges

Drivers can obtain digital badges at the end of each month if their relative rank in at least one of the skill dimensions measured by *FuelOpps* is higher than a pre-defined threshold. This is conditional on having a minimum number of extractions (i.e., observations) and total miles traveled in that month. The thresholds are specific to each company and to each dimension and vary according to the color of the badge, which can be either blue or gold. Gold badges are harder to obtain, since their thresholds are higher than the ones for blue badges. In general, blue badges are awarded to the drivers ranked in the top thirty or forty percent, while gold badges are given to drivers in the top ten or twenty percent. Blue and gold badges are mutually exclusive in a given month; if a driver receives a gold badge for his relative ranking in the progressive shifting skill, she will not get a blue badge and vice versa.

Companies that implement *FuelOpps* can opt to include points per badges obtained as well. In general, only a couple of companies have chosen to assign points to badges, so in most cases badges represent just a non-pecuniary way to highlight positive driver behavior.

#### 2.3.2.3 ***FuelOpps*** levels

According to Propel IT, the idea of creating *FuelOpps* levels (FL hereafter) was motivated by the interest of introducing an additional *gamification* feature on the platform. FL are not tied to any pecuniary reward or incentive, and the information regarding a driver's level is only privately available to each driver. The objective behind FL was to give drivers a way to brag about their levels, which are correlated with their driving skills.

All drivers start at level one (F1). They can advance to the next level through accumulation of experience points, which are provided by badges. Blue badges give one experience point, while gold badges add five experience points. Table 2.1 presents the number of experience points needed to reach each level. The number of points required to level up is increasing with each level, reaching a maximum of 2,000 for F16. Reaching F15 would require at least 50 months for the top drivers of manual transmission trucks, assuming these drivers get gold badges in all four dimensions during each of the months in which the *FuelOpps* is operative.

A FL cannot be lost once reached. Also, experience points are cumulative. If, for example, at the end of the first month a driver ends up with two experience points (given by two blue badges she obtained in that month), she only needs to get one more blue badge in the next period to achieve F2. If she had obtained one gold badge instead, she would have advanced to F3 directly at the end of the second month.

### 2.3.3 Company-specific interventions and campaigns

Trucking companies have the last word when it comes to which features of *FuelOpps* to implement. The *FuelOpps* program allows for customization in terms of the features available and the dimensions of performance that will be measured and rewarded. For example, firms can choose to tie a portion of the drivers' compensation to their scores. On the other side, trucking companies can even leave scores as a purely symbolic evaluation of drivers' performance, without any way to translate them into goods or higher earnings. Clients of *FuelOpps* can also decide to run their own campaigns, such as bonus payments for drivers in the top of the distribution according to their scores. This also applies to the way companies use the information generated by *FuelOpps* to provide public recognition to their best drivers. An example of this is the use of printed leaderboards posted at terminals so drivers can see where they stand among their peers.

### 2.3.4 Graphical user interface

*FuelOpps* offers different ways to driver to access their scores and other pieces of information such as items in the catalog or messages. One of the main ones is to do it via website, using a login and a password provided by Propel IT.

Figure 2.2 presents a screenshot of the web version of *FuelOpps* 1.0 for a driver of an automatic-transmission truck. On the top of the screen, the driver can choose the date range that she would like to visualize. On the left of these boxes, the driver can see her actual FL and her progression to the next level. Below this, the driver can observe her overall score and her relative position in her group, with the same type of information also available for each of the other driving skills. Blue regions in the graphs show the range of actual scores obtained by other operators in the driver's group. Information about idling (in minutes) is also presented on the bottom of the screen. On the right-hand side, data on badges obtained and points earned are displayed. A link to the catalog (if available) is located on the bottom-right corner, where drivers can check the items that they can purchase using their points. All products in the catalog are expressed in points rather than in their actual monetary values. In general, a *FuelOpps* point is defined in such a way that it is worth, approximately, \$0.01.

As mentioned above, drivers and coaches can interact with each other via messages that are sent to the Qualcomm device connected to the truck's computer. Drivers can also visualize these messages on the smartphone application of *FuelOpps*, which drivers can download for free and use their credentials to log in, check their scores and points, and read any messages sent by coaches. The website version of *FuelOpps* did not offer the ability to read or send messages.

## 2.4 Data

### 2.4.1 Company 1050

The database consists of 96,742 daily observations between June 25th, 2014, and January 5th, 2016 for 706 drivers in Company ID #1050 (Company 1050 hereafter). Daily information collected for drivers include (among other variables): distance covered, fuel consumed, an identificator code for the truck driven on each day, *FuelOpps* scores, badges obtained, points awarded, messages sent and received by drivers, calls made by coaches, and points redeemed for goods or services. All drivers in Company 1050 operate manual-transmission trucks.

Company 1050 is one of the few firms in the database of historical customers of Propel IT that has measures of performance for its drivers for at least a month prior to the introduction of the *FuelOpps* platform. Even though collecting data from drivers is a standard procedure before launching a trial, an activity that is done in order to test the communication services between Propel IT and the telematic companies, the length of the collected pre-trial information varies greatly among implementations of *FuelOpps*. The pre-trial data available for Company 1050 cover from June 25th until August 3rd, 2014, making it a suitable candidate to test the impact of the *FuelOpps* platform on driver performance. In addition, data of Company 1050 also include observations for the period after the implementation of *FuelOpps* was finished, giving us the opportunity to test the impact on fuel efficiency after the application was removed.

Regarding its activity, Company 1050 operates in the midwest zone of the United States and most of its operation is about transporting different derivatives of petroleum products such as gas or ethanol.

#### 2.4.1.1 Timeline of events in Company 1050

In general, trucking companies first agree to test the *FuelOpps* system on a subset of their workers (i.e., a trial) to evaluate its impact before deciding whether to implement it for all their workforce. The trial in Company 1050 formally started on August 11th, 2014 and it was originally expected to last until October 31st of the same year. One group of drivers —which we call treatment— was chosen to have access to the *FuelOpps* platform during the trial. The drivers in the treatment group received the first message relative to their participation in the *FuelOpps* program on August 4th. The messages stated acknowledged drivers that they would be part of the trial and that they would have access to the *FuelOpps* platform starting on August 11th.[6]

After the trial had ended on October 31st, 2014, Company 1050 decided to extend the trial starting again on February 3rd, 2015 and ending on October 31st, 2015.[7] In addition to drivers in the treatment condition, a second group of drivers —which we call internal— got access to *FuelOpps* during the second wave of the trial. Between the end of the first wave of the trial and the start of the second, drivers in the treatment group stopped receiving support from Propel IT.

With the inclusion of drivers the internal group, Company 1050 had its entire workforce of full-time drivers in the *FuelOpps* system. The only subset of drivers working for Company 1050 that never were part of the *FuelOpps* programs was the one comprised of independent drivers (operators that generally own their trucks and pay for their fuel).

In summary, we can distinguish three different types of drivers in Company 1050.[8]

- **External**: drivers in the external group never had access to the *FuelOpps* platform

[6]The messages sent to drivers during the seven days prior to the launch of the trial are presented in Appendix 2.8.

[7]Besides just extending the trial, Company 1050 added two subsidiary companies to the trial starting on March 11 and until June 9th, 2015. The data of these two subsidiary companies are not analyzed in this study.

[8]As explained above, we identify three groups of drivers according to when they got access to *FuelOpps*. However, the data provided to us by Propel IT does not directly recognize if a driver is a direct employee or a contract operator. We were able to establish whether a driver was a direct employee of Company 1050 based on the date they were added to the program. For drivers whose first measure of performance is obtained after the start of the trials, and for which there is no message sent to by coaches, we are unable to determine if they were full-time drivers or external contractors. We opt to eliminate all drivers for which we cannot determine their status in the company.

during the pre-trial, trial, or post-trial phases. This subset of drivers is comprised of independent contractor drivers that are hired by Company 1050 to cover some routes but are not considered full-time employees. In some cases, these independent contractors also own their trucks, so they are responsible for the fuel expenses and the depreciation of their vehicles.

- **Treatment**: drivers in this group had access to the *FuelOpps* platform between August 11th and October 31st, 2014 and then again from February 3rd, until October 31st, 2015.
- **Internal**: drivers in this group did not have access to the *FuelOpps* platform until February 3rd, 2015. They had access through October 31st, 2015.

Figure 2.3 presents a timeline of the events related to implementation of *FuelOpps* in Company 1050 and the dates for which we have information about the performance of drivers for each of the groups described above. Dashed lines represent time intervals in which groups did not have access to the *FuelOpps* platform, while solid lines represent periods in which drivers had access to their *FuelOpps* scores and coaching support from Propel IT. We divide the whole period for which we have information in five different phases in order to facilitate the analysis. **Phase 1**, the pre-trial phase, includes performance information from June 25th until August 10th, 2014. **Phase 2** goes from August 11th until October 31st, 2014, which is when the first trial ended. **Phase 3** comprises between November 1st, 2014 and February 2nd, 2015, which is the period when drivers in Group 1 stopped receiving support from Propel IT. **Phase 4** starts on February 3rd and lasts until the end of the trial on October 31st, 2015. Finally, **phase 5** includes all observations from November 1st until January 5th, 2016, a period in which no driver from Company 1050 had access to *FuelOpps*.

During most part of the trial phase, scores were computed under the algorithm 1.0. On August 1st, 2015, Company 1050 requested that scores were presented under the new version of the algorithm (2.0). As mentioned before, scores under different versions of the algorithms are not comparable between each other.

Phases do not only differ in terms of the groups of drivers that had access to the *FuelOpps*, but also on the specific features of the *FuelOpps* application that were available. For example, while in phase 4 drivers could exchange their points for goods available in the catalog, this feature was not accessible and points did not have any pecuniary value during phase 2. Moreover, there are a few differences in the characteristics of *FuelOpps* that drivers experimented in each phase. While phase 4 can be understood as a somewhat full implementation of *FuelOpps* in terms of the overall characteristics of the system explained above, phase 2 gradually added some features that allows us to compare their effects on driver performance. Figure 2.4 presents a summary of the main characteristics of *FuelOpps* during Phase 2.

Phase 2 is composed of four different sub-phases. As mentioned previously, between August 4 and August 10, 2014 drivers in the treatment condition received messages that alerted them of the start of the trial on August 11. Although these messages did not offer much detail regarding what the drivers could expect from *FuelOpps*, the message sent on August 7th does mention that rewards would be given at the end of the trial to those who did well. Starting August 11th drivers in the treatment condition got access to the main features of *FuelOpps* such as skill scores, points, badges, and messages, attributes that were available until the end of the trial on October 31, 2014. On September 15, drivers were informed that at the end of the trial, those on the top 30, 20, and 10 percent would receive \$75, \$150, and \$250, respectively. Lastly, on Octobre 15, drivers received a message that communicated them of a t-shirt campaign, which implied that all drivers who reached a target score — determined specifically for each driver— during the last two weeks of the trial would get a Nike *FuelOpps* shirt. We use this gradual addition of features as our main strategy to study the effects of *FuelOpps* on fuel efficiency over the duration of the trial. Given that the implementation of *FuelOpps* for the internal group of drivers (those who got access to *FuelOpps* during 2015) included characteristics not available to those during the first part of the trial such as a catalog of items that they could purchase using their points, we decide to focus only on phases 1, 2, and 3 so to have a clear vision of the effects of *FuelOpps* on driver behavior as new features were made available.[9] Also, since during phase 3 drivers stopped receiving support from Propel IT, we can still evaluate the impact of *FuelOpps* during a time frame of almost three months in which one of the main services provided by *FuelOpps*, the coach support, was not available and no other campaign was active.

It is interesting to note that the information available for Company 1050 allow us to estimate (i) the impact of the introduction of the *FuelOpps* application on driver performance and (ii) whether drivers that had access to *FuelOpps* show a better performance after the trial has ended.

[9]Even though drivers in the treatment group were able to collect points as a result of their daily performance during phase 2, there was no way for them to redeem those points for goods or services since Company 1050 chose to not have a catalog active during this phase of the trial.

### 2.4.2 Characteristics of ***FuelOpps*** during phase 2

During Phase 2, all drivers in the treatment condition were assigned to the same group for ranking purposes. With respect to the parameters of the point allocation function (Equation 2.1), the following ones were used: $r_0 = 55$, $r_1 = 100$, $p_0 = 10$, and $p_1 = 1,000$.

In terms of non-monetary incentives, badges and FL were present during phase 2. Additionally, physical leaderboards were posted at the terminals from which drivers were dispatched.[10] With respect to feedback, coaches were very active in terms of reaching out to drivers and congratulate them for increasing their scores from one week to the next or to ask why scores have dropped.

Two additional campaigns were added to the trial:

- Monetary prize: On September 15, drivers were alerted that the top 30 percent of drivers would obtain a financial prize at the end of the trial. The details were as follows: the top 10 percent would get $250, the top 20 percent $150, and finally the top 30 percent only $75.

- T-shirt: On October 15, drivers received a message that told them that they could obtain a Nike *FuelOpps* shirt if they were able to reach a specific target score for the last two weeks of the trial. Each target was uniquely determined by Propel IT's former CTO, and it was a combination of measures of past performance and relative ranking. As we discuss in Section 2.5.4, this way of computing target scores implied that some drivers got goals that were below their actual levels of performance.

### 2.4.3 Dependent variables

The most well-known measure of fuel efficiency in the trucking industry is MPG. However, MPG is not the most appropriate metric to evaluate fuel efficiency as it can lead to erroneous conclusions if not used carefully. The issue with MPG is that, for a given distance, the relationship between fuel consumption and MPG is curvilinear (Larrick and Soll, 2008). Therefore, comparing the reduction in fuel consumption for an increase in MPG is not independent of the actual MPG of the vehicle. Let's consider a somewhat extreme example (see Table 2.2). Assume you drive 10,000 miles each year, what would be more beneficial in

[10] These leaderboards were only available during phase 2.

terms of fuel efficiency: (A) replacing an old car with 1 MPG for one that offers 2 MPG or (B) replacing a car with 25 MPG for one that gives 40 MPG? The first alternative would allow us to save 5,000 gallons of fuel per year, while the second one would imply a reduction in fuel consumed of just 150 gallons.[11] With this in mind, we use the inverse of MPG or gallons per mile driven (GPM) as our main dependent variable. GPM is a linear measure of fuel efficiency and it can be used in a straightforward way to compare the performance of drivers.

Driving is not the only task performed by truck drivers that is related to fuel consumption. Idling, leaving the engine on while the vehicle is stopped, has been identified as an activity that greatly affects the fuel used by a vehicle. Besides being bad for the environment, in the form of the additional $CO_2$ that is emitted to the atmosphere, idling has been associated with the consumption of up to one gallon of fuel per hour and higher wear-and-tear expenses (Omnitracs, 2012). While there is some idling that is inevitable (i.e., having to stop at a traffic light), there are many situations in which leaving the engine on is under the direct control of the driver. Given that the objective of *FuelOpps* is to help reduce the total fuel consumed by drivers, we also evaluate the impact of its introduction on fuel used differentiating between driving and idling.

We construct different versions of our fuel efficiency variables. Throughout this study we use $i$ and $t$ to denote drivers and days, respectively.

- **GPM**$_{it}$: represents the ratio of total fuel used to distance traveled by driver $i$ during day $t$. By total fuel used we mean the fuel used while driving and also the fuel consumed during idling.
- **GPM-D**$_{it}$: represents the ratio of fuel used while driving to distance traveled by driver $i$ during day $t$.
- **GPM-I**$_{it}$: represents the ratio of fuel used as consequence of idling to distance traveled by driver $i$ during day $t$. Since idling might be more likely when drivers need to cover long distances, we choose to express the fuel consumed idling as a ratio to the total distance driven, keeping it consistent with the other two variables defined above.

[11] The number of gallons consumed by the vehicle with 25 MPG is $400 = 10,000/25$, while the number of gallons used by the car with 40 MPG is $250 = 10,000/40$.

### 2.4.4 Graphical examination of GPM

Figure 2.5 presents the average fuel efficiency —$\text{GPM}_{it}$— per group on each of the days between June 25th, 2014 and January 5th, 2016. As can be seen, the daily average fuel efficiency of group 1 and group 2 behave rather similarly during phase 1. However, we can notice how both lines differ after the introduction of *FuelOpps* during the first days of phase 2. In phase 3, the difference in fuel economy between group 1 and group 2 becomes less pronounced as time goes by, giving the impression that right after the start of phase 3 there is little difference, if any, between the average fuel efficiency of those groups. During phase 4, which is the phase when both groups of internal workers have access to *FuelOpps*, we can observe that their average fuel efficiency follows almost the exact same evolution over time from the start of phase 4 and even after the trial ended (phase 5). For external drivers, their daily average fuel efficiency during any of the phases is consistently worse than the one observed for internal workers (groups 1 and 2).

Since the time covered by our database includes almost 18 months, it is easy to observe the seasonality present in our fuel efficiency variable (GPM). During the summer season, fuel economy is around 0.15 GPM, while in the winter season fuel efficiency 'decreases' to levels around 0.165 (note that the lower the number of gallons consumed per mile driven, the better the fuel efficiency). The spike in GPM during winter can be explained by, among other reasons, an increased friction in engine and transmission, denser air (which affects aerodynamics by increasing resistance), and lower tire pressure (DOE, 2017; EPA, 1976).

### 2.4.5 Independent variables

#### 2.4.5.1 Control variables

One of the first variables that could affect the fuel efficiency of drivers is the total distance that they drive in a single day. Engines consume less fuel when they are used in highest gears, and that is something that is possible to achieve when driving long distances. Therefore, we would expect, on average, a higher fuel efficiency when drivers cover longer distances. However, we do not anticipate this positive effect of distance on fuel efficiency to be linear, since the ability and concentration of drivers to change gears as soon as possible (so to reduce the amount of energy the engine uses) might be negatively related to the amount of time the driver has been operating the vehicle. This could be specially relevant for trucks that use manual transmission, as it is the case with Company 1050. Thus, we include both the distance driven, defined as hundreds of miles driven in a single day, and its square term in our regressions as control variables.

Additionally, weather conditions might also affect the performance of drivers (EPA, 1976). Operators need to be extra careful when driving trucks in bad weather conditions, since they could easily lose control of their trucks if they do not take precautionary measures such as reducing their top speed and the gear they use.[12] In order to control for all these day-specific characteristics, we include day fixed effects in our models in the form of dichotomous variables for each of the days in our database. Similarly, we control for machine fixed effects; since we have a variable that allows us to identify the truck an operator is driving, we use this variable to control for machine-specific effects that might affect the performance of drivers. As for day-specific effects, we use a set of dichotomous regressors to capture for the average effect a specific truck may have on a driver's measure of fuel efficiency. Finally, we add driver fixed effects to control for drivers' time-invariant attributes (e.g., a driver's 'quality') that can be associated with better performance.

### 2.4.6 Filters

According to Propel IT, the raw version of Company 1050's data may not be error-free. In order to eliminate some potential outliers from our data, we drop all observations that: (i) contain missing values (or zero) for either the fuel consumed or the total distance driven; (ii) have a value for the distance driven in a single day that is less than 25 or greater than 800 miles; (iii) have a value for GPM, GPM-D, or GPM-I that is greater (less) then the mean plus (minus) four times its standard deviation; (iv) are not directly related to a specific driver; and (v) correspond to a multi-day extract (i.e., drivers that were dispatched more than once on a specific day). Table 2.3 presents the descriptive statistics for the final sample of drivers.

### 2.4.7 Empirical strategy

Given the nature of the data and the way the *FuelOpps* application was introduced, we use a differences-in-differences approach to measuring the impact of *FuelOpps* on fuel performance.

[12] Additionally, the actual composition of fuel changes between the winter and summer seasons; in the United States, the fuel produced in the summer contains less energy than the fuel produced in winter (DOE, 2017).

A differences-in-differences estimator compares the changes in a variable of interest between a group that receives a treatment at some point and a group that does not (hence its name).[13]

In our setting, a generalized version of the simplest diff-in-diff model with $t = 1, ..., T$ ($T > 2$) that explains fuel efficiency ($GPM$) as a function of 1) the introduction of *FuelOpps*, 2) a set of dichotomous variables to control for common factors ($d_t$) that might affect fuel performance such as weather conditions and/or general events that take place in the geographical zone where drivers operate,, 3) additional covariates to control for time–varying variables ($X_{it}$), and 4) and dichotomous variables to control for fixed characteristics of the trucks that drivers operate ($m_{it}$) is presented in Equation 2.2.

$$GPM_{it} = \alpha + \beta FuelOpps_{it} + X_{it}'\delta + d_t + m_{it} + v_i + u_{it}. \tag{2.2}$$

Although the estimate of $\beta$ in Equation 2.2 cannot be expressed as the exact difference between the changes experimented by the treatment and control groups when a vector of time-varying variables is included in the model (see Appendix 2.9), the causal interpretation of the effect of *FuelOpps* on fuel performance remains the same (Wooldridge, 2010).[14]

One potential modification to Equation 2.2 would be to include a driver's lagged fuel efficiency as an additional regressor. Although this idea has intuitive appeal since a driver's past performance is usually the best indicative of what her future performance would be, incorporating this variable into our model will make the OLS estimation inconsistent (Angrist and Pischke, 2008).

### 2.4.8 Parallel trends

One of the assumptions of a diff-in-diff approach is that the pre-intervention trends should be similar for both the control and treatment groups. To test this, we run the following regression in a fashion similar to Aouad, Brown, and Whaley (2017), but only considering

[13]Please see Appendix 2.9 for a more detailed explanation of the differences-in-differences methodology.

[14]Equation 2.2 includes a constant term, which we denote $\alpha$ just for practical purposes. Since we include driver fixed effects to control for subjects' time-invariant characteristics, we need to either drop the constant term or one of the drivers' dichotomous fixed effects to actually estimate our models. We use Stata 13 to fit our models, and Stata by default presents a constant for fixed-effects panel data models that is an average of all the drivers' fixed effects. This way, we keep the constant term in our equations with the caveat that it does not represent an actual individual's fixed effect but rather an average of all the estimated individual effects.

the data from the pre-trial period:

$$GPM_{it} = \alpha + \sum_{t=25Jun}^{02Aug} \beta_t \left(d_t \times treated_i\right) + X_{it}'\delta + d_t + m_{it} + v_i + u_{it}. \tag{2.3}$$

In Equation 2.3, we regress GPM on a group of driver-day specific variables ($X_{it}$) that includes distance covered, its squared term, and short idle time. We also add dichotomous variables that control for day ($d_t$), vehicle ($m_{it}$), and time-invariant driver ($v_i$) characteristics.[15] Variable $treated_i$ takes the value of 1 if driver $i$ is part of the treatment group and the term $\beta_t$ captures any difference, on average, that could exist between these drivers and the ones in the control group (internal and external drivers) relative to fuel efficiency (GPM) in the pre-trial period. August 3, 2014 —the day before drivers received the first message about *FuelOpps*— is the base category for the computation of the effects of interest.

Figure 2.7 presents the estimates for the $\beta_t$ parameters in Equation 2.3. As we can see, the estimates and their 95% confidence intervals include zero in each of the days. According to the figure, the performance of the treatment group before the trial started does not seem to be systematically different from the control group (internal and external drivers), especially during the week before drivers started receiving the first pre-trial messages.

## 2.5 Results: does ***FuelOpps*** help improve driver performance?

As explained in Section 2.4.7, we evaluate the effect of the introduction of *FuelOpps* on fuel efficiency by regressing our measure of fuel efficiency (GPM$_{it}$) against a dichotomous variable that becomes 1 if the driver has access to *FuelOpps* on day $t$ plus our set of control variables (explained above) along day, vehicle, and driver fixed effects. We expand Equation 2.2 by including two additional binary variables to help us capture the influence of the pre-trial messages (PT-Messages$_{it}$) and the effect *FuelOpps* had on the post-trial performance (Post-Trial$_{it}$) of drivers. Our main regression model appears below.

$$GPM_{it} = \alpha + \beta_1 \text{PT-Messages}_{it} + \beta_2 FuelOpps_{it} + \beta_3 \text{Post-Trial}_{it} + X_{it}'\delta + d_t + m_{it} + v_i + u_{it}. \tag{2.4}$$

[15]This same set of control variables is used throughout all the models estimated in this paper. All models are estimated considering driver fixed effects except when noted otherwise.

Column (1) in Table 2.4 presents the estimates for the *FuelOpps* program as a whole, without differencing among the different stages in phase 2. As we can see, the average daily effect of *FuelOpps* on GPM is a reduction of almost $-0.00281$ in gallons of fuel consumed per mile driven, considering both driving and idling (a result that is also highly significant in statistical terms). Since the average driver in the treated group covered around 6,000 miles during August, 2014, we can approximate the total average reduction of fuel consumed per driver as 16.86 gallons per month. Interestingly, there is also an effect on GPM during the period when drivers received the pre-trial messages, although the messages did not contain any tip on how to improve scores. Since drivers could access the *FuelOpps* website and check their scores since the day the login credentials were sent (August, 8th), we expect them to show curiosity and be excited about this new system that promised to bring them the recognition they deserved as it was mentioned in the first pre-trial message on August 4th (Appendix 2.8). This 'anticipation' effect is smaller than our estimate for the *FuelOpps* application, which also makes sense since little was know about all the characteristics of this new program and the way it would work. The estimate for the post-trial effect of *FuelOpps* is also significant both statistically and substantively, implying that during the three months after the end of the trial drivers in the treatment group consumed less fuel per mile than drivers that did not get access to *FuelOpps*. The post-trial effect is also lower than the effect of *FuelOpps* when the trial was running, but not for much: during the post-trial period drivers in the treatment group consumed, on average, 0.00041 gallons per mile more than while the trial was running (or 2.46 gallons per month more assuming 6,000 miles driven per month). Our results suggest that most of the newly acquired skills (if any) and/or behavior changes that occurred as consequence of *FuelOpps* were 'sticky' and that they did not vanish in the absence of external incentives (either monetary or non-monetary). This seems to be the case, at least, for the period covered by our data.

Control variables behave in the expected ways in all models estimated in Table 2.4. Specifically for results under column (1), the total number of miles driven appear to have a positive effect on fuel efficiency (i.e., reduction of fuel consumed), but this effect seems to be decreasing as drivers cover a longer distance. Our estimates suggest that the maximum GPM is reached when workers have driven around 483 miles in a day, and any distance greater than that negatively affects a driver's fuel efficiency.[16] Also, once a driver has surpassed the 965 miles in a day, the total effect of distance becomes detrimental to GPM. Driving

[16] $483 = 0.01081/(2 \times 0.00112)$.

483 miles in a day is actually possible for a driver to achieve given the restrictions on the number of hours that a driver can be behind the wheel. This result helps us visualize that there is a limit to the fuel efficiency that a driver can achieve even if she is able to drive a long distance (using a high gear and low RPMs) as a consequence of fatigue and exhaustion. Finally, short idle time has a negative impact on fuel efficiency, which makes sense, since every time that a truck has to stop because of a red traffic light or other events on the road, the driver needs to accelerate the truck again incurring in a higher energy consumption.

The next question is whether the estimated impact of *FuelOpps* under column (1) is constant over the different stages involved in phase 2. Under column (2), we show the results of our main model but now differentiating among the three sub-phases in phase 2 as depicted in Figure 2.4. The savings are greater during the first month of the trial, when there was no information about the composition or the criteria that would be used to distribute the prizes promised at the beginning of the trial. The decrease in fuel savings between the first two stages of the trial is of almost 25 percent, while the difference between the first and third stages the reduction is even more noticeable (38 percent).

Columns (3) to (6) show the results of our main models for GPM-D and GPM-I (gallons of fuel consumed per mile as a consequence of driving and idling), respectively. Our estimates in columns (4) and (5) indicate that the total fuel savings during the trial can be explained in almost entirely for better driving techniques (although the estimate for the effect of *FuelOpps* on idling is not significatively different from zero at the 90 percent of confidence). The estimates for the effects of *FuelOpps* by each sub-phase are similar to those under column (2), which seems natural given the importance of driving in the determination of fuel efficiency. On the other hand, the estimates for the GPM-I show an interesting pattern. Although drivers could see their score for idling when they accessed the *FuelOpps* platform, in all the messages sent by *FuelOpps* coaches to drivers, idling was never mentioned.[17] The decrease in fuel consumed as consequence of idling during the first sub-phase (after the pre-trial messages) can be attributed to the vague description of how prizes would be determined at the end of the trial. Once it was made clear that prizes would be awarded in base to the overall score (which did not include the idling score), idling returned to its pre-trial levels for the rest of the trial. However, after the trial ended, drivers in the treatment group exhibited a drop in their GPM-I relative to other drivers in the company. We do not have

[17]There is only one message sent by a *FuelOpps* coach where she discussed idling, but it was in response to a driver that had asked about it. Idling was never mentioned in any other message sent to drivers during phase 2.

any information on additional initiatives taken by Company 1050 after the end of the trial with the aim of improving fuel efficiency among drivers in the treatment group.

### 2.5.1 Behavioral response

Once we have determined the impact of *FuelOpps* on performance, our next step is to analyze the potential causes, besides idling, that could explain the observed drop in fuel consumption during and after the trial.[18] As mentioned above, *FuelOpps* provide drivers with scores on progressive shifting, high RPM, highest gear, and highway speed. All these dimensions are also further summarized in an overall score, and a driver can see all her scores (computed for different time frames) on the smartphone application or on the *FuelOpps* website.[19] Although drivers in the internal and external groups were not part of the trial, Propel IT still computed their scores. However, drivers in those groups never got access to them. Since drivers in the treatment condition could track their scores and their evolution over time, it is natural to think that they would try to improve their scores. Then, assuming that the *FuelOpps* scores do capture important determinants of fuel consumption, then we should observe a positive correlation between scores and fuel efficiency. Of the four scores, we only consider three in our regression models: progressive shifting, highest gear, and highway speed. High RPM is excluded because most of the drivers obtained perfect scores in that dimension. Figure 2.8 presents histograms for each score in each of the sub-phases considered in our analyses (from the pre-trial to the post-trial periods). As can be appreciated, most of the distribution of the scores shift to the right during the trial (especially when we compare the pre and post-trial histograms). The introduction of *FuelOpps* helped increase *FuelOpps* scores.

We complement our analysis considering alternative measures to the *FuelOpps* scores. Since one of the objectives of *FuelOpps* is to decrease the top speed of trucks on highways, we look at both the average speed and move time and their evolution during the trial. A reduction in top speed should be reflected in both the average speed and in the number of minutes a worker is driving the truck, since a lower average speed implies more time driving (keeping distance constant). We consider these variables as an alternative to the 'highway speed' score.

[18]We already tested for the potential effect of *FuelOpps* on idling in Section 2.5, concluding that there was almost no reduction in idling during the period under analyses. Therefore, the entire effect must be consequence of better driving techniques or a reduction of average speed.

[19]The IVR system only provides information on the most recent scores.

Each truck in Company 1050 is equipped with a specific sensor that records the number of minutes the engine has been operated above a threshold, which is truck specific and it is not defined by Propel IT. Lower RPMs are related to less engine wear, so the less time an engine is operated above that threshold the better for the truck. We use this variable as a replacement for the 'high rpm' score which we do not include in our analyses.

Table 2.5 shows the results for *FuelOpps* scores and our select set of variables that capture changes in behavior mentioned above.[20] In general, all scores increase during the trial. Additionally, the effect of the trial on scores seems to be persistent, since positive differences in scores for drivers in the treatment group are observed after the end of the trial. However, some of the post-trial differences are not as large as the ones observed during the trial. For example, the highest gear score shows an average increment during the post-trial phase of almost 8.7 points with respect to the pre-trial, but the difference was as large as 11.9 points when the t-shirt campaign was running. Interestingly, the effect on the scores is also perceived, although to a lesser extent, during the days when drivers were alerted of the trial. To the extent that the *FuelOpps* scores capture important dimensions of fuel efficiency, we can interpret these estimates as evidence that drivers know how to perform better (to some extent), but that they lack the motivation to do it. If drivers had not known how to improve their driving, and new skills were necessary, we hardly would have observed an effect of the pre-trial messages on the dimensions measured by *FuelOpps*.

The score for the highway speed dimension improves during and after the trial as well. Although this implies that drivers were better at the keeping the speed on highways around 62 MPH, it does not tell us how this impacted their average speed or move time (e.g., time the truck is in movement). Columns (4) and (5) in Table 2.5 present the effect of *FuelOpps* on speed and move time, respectively. As we can appreciate, the average speed is reduced in almost 0.87 MPH during the first phase of the trial. This reduction in average speed translates into an increase of almost 6.58 minutes of extra move time per day. Assuming a worker drive for 23 days in a month, this estimate implies that —on average— this driver would work for almost 2.5 extra hours in a month. After the trial has ended, workers did not return to their previous levels of speed, but stayed at a somewhat lower level (-0.26 MPH in comparison to the pre-trial levels) and took around two more minutes extra each day to make the deliveries.

[20] Scores for 100 observations were not available in the database. This is the reason the total number of observations is lower for the models that use *FuelOpps* scores as a dependent variable.

Column (6) in Table 2.5 presents the result of a driver fixed-effects Poisson model in which the dependent variable (over RPM) is the number of minutes the engine is operated above a threshold that is specific for each truck.[21] Our estimates indicate that the likelihood of operating the engine above the threshold is lower while the *FuelOpps* trial is running, especially during the t-shirt campaign. However, for the post-trial phase we cannot reject the null hypothesis that drivers return to their original levels of operating the trucks above their thresholds as much as they did before the trial.

### 2.5.2 Impact of *FuelOpps* by quartiles

Since we have already measured the average impact of *FuelOpps* on GPM for the sample as a whole, the next step in understanding the way *FuelOpps* affected drivers is to analyze whether the impact varied for different 'quality' of drivers. To accomplish this, we divide our sample of drivers in quartiles according to the drivers' fuel efficiency, which we obtain from the simplified regression model below.

$$GPM_{it} = \alpha + X_{it}'\delta + d_t + m_{it} + v_i + u_{it}. \tag{2.5}$$

The regression model in Equation 2.5 was estimated using only pre-trial observations for drivers in the control and treatment groups. Driver fixed effects were extracted and divided into quartiles according to their values. Those drivers with an estimated fixed effect below the 25 percentile were categorized as 'top' drivers, since a low fixed effect implies better fuel efficiency. Drivers with values between the 25 percentile and the median were considered 'good' drivers, and so on (the last two groups were named 'bad' and 'worst' just for clarity). Table 2.6 presents the results of our main model when drivers are separated in quartiles.

The estimates in Table 2.6 present a very different story for each of the quartiles. While the best three quartiles reacted to the pre-trial messages, the worst quartile did not show any change in its average fuel efficiency during the days when the *FuelOpps* program was

[21] The number of observations is greatly reduced when compared to the other models in Table 2.5 because all drivers that never had an 'event,' i.e. that never operated their trucks above the thresholds, were not considered in the estimation. Additionally, we do not consider vehicle fixed effects in this analysis since (i) the dependent variable is measured for each truck differently and (ii) the fixed-effect Poisson model, as usually occurs with other non-linear models that include a large set of independent variables, had trouble converging when we incorporated our set of dichotomous variables to control for vehicle fixed effects. Results from an OLS regression that considers the whole sample, and that includes vehicle fixed effects, are similar.

announced. The contrasting differences by group to the announcement of the program could be explained by the chances each driver assumed she had to get a prize at the end, since pre-trial messages only said there would be "rewards for doing well." The results for the top three quartiles give support to this idea since the effect of the announcement of *FuelOpps* is greater for better drivers. During the first sub-phase of the trial, all groups react positively to *FuelOpps*, implying that the set of points, badges, levels, and scores helped drivers improve their skills. However, all this was under the promise of future rewards that were not fully disclosed at the time. Once the criteria were defined, the effect on the lowest-quality drivers was quite dramatic: they reverted to their pre-trial levels of performance. Since rewards were only targeted to the top 30 percent of drivers, workers in the bottom quartile probably felt they had almost no chance to get them, so they just stopped trying. Drivers in the other groups also lowered their fuel efficiency, although they still showed a better performance in comparison to the one they had before the trial started. It is interesting how drivers in the top quartile lowered their effort levels after the specifics of the prizes were disclosed; the $250 offered as prize for the top 10 percent might not have been enough to compensate the additional effort and time (more on this in later) that they put on to decrease their fuel consumption.

The estimates of the effects of the control variables are similar in sign and significance to the ones in Table 2.4. However, some differences remain. Distance and its squared term keep displaying the quadratic relationship we found previously. For all groups, the maximum increase in fuel efficiency happens around the 500 miles driven, with each additional mile having a negative marginal impact on GPM. However, the overall impact on GPM is different for each group. While lowest-quality drivers can lower their GPM by almost 0.03 when driving the optimal number of miles, top drivers can only achieve a GPM that is better by 0.02. This difference might reflect their different set of skills if we assume that driving a higher number of miles implies spending more time driving on highways where it is, in general, easier to keep a steady speed.[22] With respect to short idle time, all quartiles present a lower GPM the longer the short idle time. Nevertheless, the estimates for the impact of short idle time also reflect the distinction between quartiles: while each additional minute of short idle time in a day increases GPM in 0.00034 (or 10.2 percent of gallon assuming a driver covers 300 miles on that day) for low-quality drivers, it only increases GPM in 0.00028

[22] Also, low-quality drivers have more room to improve than top quality drivers, and this is also being reflected in the estimates.

(8.4 percent of a gallon) for top drivers. As discussed above, upshifting at the right number of RPM and avoiding hard acceleration can have a big impact on fuel consumption, and those are characteristics of high-quality drivers. Our estimates reflect the distinct skill levels of drivers.

### 2.5.3 *FuelOpps* and novelty

Koivisto and Hamari (2014) present the idea of 'novelty effects' in a gamified application, which they develop from the fact that users' reported levels of satisfaction and usefulness of a gamified platform decrease with the use of the service. After observing how the effects of *FuelOpps* dissipated for the some of the quartiles after the prizes were made public on September 15, 2014, we can question the efficacy of the gamification elements of *FuelOpps* when considered in isolation. Although all drivers could have been genuinely motivated by the promise of rewards at the end of the trial, we can go one step further and use the data until September 15 to analyze how the reaction of drivers to *FuelOpps* evolved. To do this, we define a new variable —$\text{days}_{it}$— that measures the number of days (divided by 30) that a driver has been in the *FuelOpps* trial, similarly to Lazear (2000).[23] The idea of $\text{days}_{it}$ is to help us control for the number of days drivers have 'experienced' *FuelOpps*, once the trial started (August 11, 2014). We include both $\text{days}_{it}$ and its squared term in our regression models to capture any potential non-linearity in the effect of the days under *FuelOpps* on fuel efficiency.[24]

Table 2.7 presents the results for the whole sample and for each of the quartiles. As we can see the effect of both days and its squared term are statistically significant at the 99 and 95 percent of confidence, respectively. The sign of the coefficients implies that a quadratic function is in place in terms of the effects of *FuelOpps* on fuel consumption. Similarly to what occurs with the distance driven, our results for the period before September 15, 2014 indicate that the effect of one extra day of *FuelOpps* is not independent of how many days a driver has already been in the trial. From our results we can compute that approximately after the 22$^{\text{st}}$ day of *FuelOpps*, the effect of one additional day of the trial —keeping everything else constant— negatively impacted the fuel performance of drivers in the treatment group. The

[23] We divide it by 30 just to avoid presenting estimates that are too small for the tables. Another way to understand the variable days is to think of it in terms of 'months' rather than days.

[24] Please note that here we are not separating between active and non-active users of *FuelOpps*. We address this in Section 2.5.5.

presence of this 'novelty' effect has been found before in the evaluation of other gamified applications (Seaborn and Fels, 2015) and it has been identified as one important challenge to overcome in order to create successful implementations of gamification.

The results for the different quartiles (columns (2) to (5)) are similar to the estimates found for the sample as a whole, although most of the estimates for days and its squared term are not statistically significant, with exception of the 'top' drivers, who seem to be driving the effect for the most part.

### 2.5.4 Target scores and t-shirts

As explained above, once the t-shirt campaign was launched, drivers in the treatment group received a message that stated their current score and the target score they needed to achieve during the last two weeks of the trial to get a t-shirt. One interesting feature about this target score is that it was determined by the Propel IT's former CTO and he used a procedure based on past performance, percentile rank, and other inputs. Figure 2.9 shows a histogram of the score gap, defined as the difference of the reported ranking and the target ranking when the t-shirt campaign was announced to drivers via text message. As can be appreciated, the score gap for most of the drivers was not very different from their actual scores (mean of the score gap variable is 0.23), but for a significant portion of workers the target was different from the current score. Surprisingly, for some drivers the target scores were even lower than their actual scores. According to goal-setting theory (Latham, 2012), we would expect higher performance when challenging goals are in place than when workers are just asked to do their best. Table 2.8 presents our main regression model including the score gap variable.

As we can see under column (2), the effect of the score gap variable on fuel efficiency goes in the opposite direction; i.e., the higher (lower) the target score relative to the current score, the lower (higher) the effect on fuel efficiency. In fact, according to our estimates, if the score gap is closer to one, the positive effect of the campaign on fuel efficiency is completely neglected. On the other hand, for drivers that were given targets below their current scores, their performance got even better. There is evidence in goal-setting theory that without goal commitment, imposed goals might not have any noticeable impact on behavior. Also, without a sense of self-efficacy, goals can also be demotivating to workers since they might question themselves if they have the tools to achieve what is being asked from them (Latham and Locke, 2007). Another possibility is that drivers who got challenging goals thought of the t-shirt as a reward more difficult to obtain, or that the t-shirt was not worth the extra effort that reaching the target would imply (Latham, 2012). For drivers on the other side of the spectrum, receiving a message with a target lower than their current score might have worked as an element of feedback, increasing the drivers' perceived degree of mastery which can also have a positive effect on their self-efficacy (Latham, 2012). Kanfer and Ackerman (1989) mention the idea that goals might help improve complex task performance when they are introduced after workers have gained the required skills. In our setting, it might have been the case that workers knew how to improve their fuel efficiency (to an extent), but did not understand how to —quickly— rise their scores so to have a chance to actually win the t-shirt.

### 2.5.5 Interaction with *FuelOpps*

Consent, which is defined in the management literature as the "the active cooperation of workers with managerial goals" (Mollick and Rothbard, 2014, p. 14) can play a central role in the effectiveness of a gamified application or game that is imposed by the management team. Scholars have started to show interest in how consent moderates the relationships between gamification and employee affect and performance (Mollick and Rothbard, 2014). In our case, *FuelOpps* was brought in by Company 1050's management as a way to improve fuel efficiency among drivers and, indirectly, keep its driver engaged an interested, with the objective of reducing the high turnover ratios that affect the trucking industry. Although we do not have information about drivers' thoughts on *FuelOpps* —so to have a clear measure of their consent— we do have data on the interaction they exhibited with the platform in the form of logins and score checks via the IVR system that *FuelOpps* had in place during the trial. Drivers with low consent should interact less (or show no interaction at all) with *FuelOpps* than drivers that have higher consent. Therefore, we can approximate a driver's consent by looking at the actual measures of interaction she had with the platform and see how her performance changed after controlling for our proxy of consent. We consider two different measures of driver engagement; a dichotomous variable ($\text{active}_i$) that takes a value of 1 if driver $i$ logged in on the *FuelOpps* application, website, or if she checked his scores via the IVR system at least once during the trial, and variable $\text{sessions}_{it}$ which measures the number of times a driver ($i$) logged in or checked his scores on the system on day $t$. We also analyze the effect of messages and calls from *FuelOpps* coaches to drivers ($\text{messages}_{it}$ and $\text{calls}_{it}$) on each day of the trial. Table 2.9 presents the results of our models including the new variables and the interactions of active with each of the sub-phases of the trial.

As in other gamified applications in the literature (Farzan et al., 2008), *FuelOpps* does not seem to be attractive to all potential users. Almost 45 percent of the drivers in the treatment group (53 out of 118) never logged in or checked their scores during the entire trial. This is a significant proportion of users for whom, apparently, *FuelOpps* had nothing interesting to offer or were reluctant even to try it. In any case, these users still received messages and might haver heard about *FuelOpps* by either their coworkers or by looking at the physical leaderboards that were available at the terminals.[25]

Results under column (1) in Table 2.9 indicate that most of the effect of *FuelOpps* on fuel performance is being driven by those drivers who showed at least some interest in the platform. Although both active and non-active drivers (in terms of their engagement with the *FuelOpps* platform) seemed to be equally excited about the announcement of the initiative, in the period when the trial was effectively launched we can clearly observe a difference between both groups; while non-active drivers are more efficient than during the pre-trial period ($-0.00216$ GPM), active users show a change in their fuel consumption of almost $-0.00401$. Once the prizes are announced on September 15, the non-active users return to their pre-trial levels of fuel efficiency. Then, all the effects observed for the last two sub-phases of the trial are entirely consequence of the efforts of the active users. However, after the trial has ended, there is no statistically significant difference between both groups. The post-trial effect for non-active users is estimated to be a reduction in fuel consumed of almost 11 gallons per month assuming 6,000 miles driven in a given month ($-0.00182 \times 6,000$), a result that is statistically significant only at the 90 percent of confidence. The estimate of the interaction term of variables active and the score gap is positive, although it fails to achieve statistical significance at the 90 percent of confidence. Column (2) presents the same model as in column (1) plus the addition of the number of messages and calls received by driver $i$ on day $t$ from the *FuelOpps* coaches. As we can see, there is no effect on fuel efficiency during the days when messages were received. Additionally, the calls seem to decrease fuel efficiency by a great amount. According to the estimates under column (3), there is no evidence that messages or calls had a different effect on active drivers. Column (4) considers the variable sessions which measures the number of logins by driver $i$ on day $t$. This variable is highly significant, even when all other variables discussed so far are included in the model (column (5)). Although how frequently to log in on the smartphone application or website is

[25]Drivers had the chance to opt out of the *FuelOpps* notification system in case they wanted to.

an endogeneous variable, we can safely conclude that there is a positive correlation between logins and fuel efficiency. In other words, those drivers that were actively engaged during the trial show better performance in terms of fuel consumption. Interestingly, the inclusion of the sessions variable only affected the effect of *FuelOpps* for active users once the trial started (and before the prizes were made public). This could mean that the positive effect of engaging with the platform are more important while users are figuring out how the whole system works and how scores change as a function of different driving patterns. Once that is understood, the impact of additional interactions with the system might not improve fuel efficiency by a relevant extent.

## 2.6 Discussion

User reviews of one implementation of the *FuelOpps* application have not been very supportive. On an Internet forum known as *The Truckers Report*, drivers discussed the features and the "fairness" of the program.[26] In general, drivers seemed critical with the implementation of the *FuelOpps* application and mentioned having issues accessing the *FuelOpps* website. Participants in this forum even calculated the value of a *FuelOpps* point and estimated its value around $0.01. One forum participant computed the benefits for her employer if drivers adopted better driving techniques and concluded that the savings for the company were approximately $8,000 per year per driver just considering fuel consumption. From the opinions expressed on the forum, one could argue that the general sentiment towards *FuelOpps* was negative; drivers thought that most of the benefits from the program would be captured by the trucking company and that the extra effort exerted by them would not be tied to higher levels of compensation. This general idea is captured by one of the messages sent by a driver from Company 1050 as a response to a coach who was asking why her scores had dropped lately.

> "There are no challenges causing my score to drop, I've simply stopped caring about it. I kept track for 5 weeks, and my score fluctuated wildly. So much that I don't think that my data was (sic) processed properly. Furthermore, I spent approximate 7 extra hours on duty as a result of driving well below my governor for a return of $10. To call that a waste of my time would be a colossal

[26]To see the comments in the forum, please click **here**.

understatement."

Not all driver-coach interactions show the same level of animosity towards *FuelOpps* as the one above. In fact, as the results presented in this study show, some drivers were genuinely interested in the system and the ways available to them to get better scores. However, it is true that higher levels of effort were required to improve scores. Being more attentive to the engine's RPM and accepting to drive at a lower speed are two examples of costly efforts that drivers must bear if they want to get better ratings.[27]

According to the information we have, drivers were alerted about the *FuelOpps* program via text messages and phone calls. In general, little detail was given to workers regarding the actual process used to compute the points, which might not be an easy process to explain and/or understand. This, added to the fact that some drivers experimented high variation in their scores when they had to drive different machines could have been damaging to the validity of *FuelOpps* in terms of the fairness of the application and the clarity of its rules (Mollick and Rothbard, 2014). Additionally, the imposition of a goal during the t-shirt campaign without a rationale behind describing why the goal was chosen in the first place could be a potential reason why drivers reacted negatively to it (Latham, 2012).

During the time algorithm 1.0 was operative, the objective determined by Propel IT was for *FuelOpps* to help minimize RPMs as much as possible, with the idea that similar threshold of RPM/speed combinations could be used to evaluate the performance of drivers in different trucking companies. However, since trucking companies operate in different settings and in the transportation business of dissimilar goods such as cereal, asphalt, natural gas, and other hazardous materials like fuel, toxic, and corrosive and radioactive materials (BTS, 2016), metrics that apply to one firm might not work another. Starting with the 2.0 version of the algorithm, Propel IT took these differences into account and now the scores provided by *FuelOpps* should be more accurate and provide a better representation of the actual performance of drivers, which should also increment the effect of the platform on fuel efficiency.

It is also important to recognize how some demographic factors might be in play and affect the effectiveness of a gamified application (Seaborn and Fels, 2015). Gender and age are some are some of the characteristic that have been discussed in the gamification (Koivisto

[27]According to the results discussed in previous sections, at some point during the trial workers in the treatment group were driving, on average, an extra 2.5 hours per month.

and Hamari, 2014), economics (Azmat and Iriberri, 2012) and psychology (Deci, Cascio, and Krusell, 1973) literatures regarding their moderating effect on some interventions such as feedback, which as we have discussed above plays an important role in all gamification initiatives. Unfortunately, the data collected by Propel IT do not include gender or any other type of demographic information from drivers. In order to explore whether the effect found in this paper might be different for males and female drivers, we use Gender-API, a web application that identifies genders from first names, to classify 95 percent of the names in our database as either male or female with at least 92 percent of accuracy (for the rest of the names, we manually decided whether it corresponded to a male or female based on information available on the Internet). Only four drivers in the treatment group were identified as female, which is in line with some statistics that say that around six percent of truck drivers in the United States are female. However, given the small number of female drivers relative to male drivers, we decide not to add this set of results to this study.[28] In short, our estimates suggest that females react similarly to males to the introduction of *FuelOpps* and that female workers drive slower than males during the t-shirt campaign. However, female drivers increase (decrease) their average speed (move time) after the end of the trial in comparison to their pre-trial levels, which indicates that, in some way, female drivers could be trying to recover the extra time they spent driving slowly during the trial.

Even though the database facilitated by Propel IT contains multiple variables related to the performance of drivers, we miss information on Company 1050 so as to have a more comprehensive view of the context in which *FuelOpps* is being used. For example, Company 1050's compensation policy is one of the characteristics that could affect the impact of the *FuelOpps* platform on fuel consumption and truck depreciation that we do not have access to. Even more, the premise of behavioral change imbued in *FuelOpps* depends, implicitly, on driver, road, load, and compensation aspects that are not included in our database and that might prove relevant to the evaluation of the effects of *FuelOpps* on driver behavior. Although we control for these characteristics in our estimates via fixed effects, we miss the opportunity of exploring the moderating effect of these elements on the impact of *FuelOpps*.

In some sense, what *FuelOpps* does is aggregating information that trucking companies already receive. Virtually all trucks operated by Company 1050 have systems on board that keep track of the data relevant to the operation of the truck (distance traveled, RPMs, fuel consumed, etc.). According to agency theory, workers should increase their effort when

[28] They are available upon request.

incentives are in place —a fact that has been corroborated empirically (Larkin, Pierce, and Gino, 2012; Prendergast, 1999)— or firms should be able to write contracts that make wages contingent to certain performance metrics given the current availability of information. However, this does not seem to be the case, at least for Company 1050. We can think of two potential explanations. First, given the current shortage of drivers Costello and Suarez (2015), one of the main challenges faced by trucking companies is the attraction of talented workers and the low retentions rates that plague the industry. Then, if a company starts implementing policies that restrict the time drivers can idle or put some pressures on wages according to the operation of the trucks based on the information recorded by the on-board systems, some drivers would probably walk away and start driving for a different company unless their compensation increases considerably. Second, anecdotal evidence indicates that driver managers lack the skills to make use of the information generated by the computer systems on the trucks. Therefore, although the information is available, they require a simpler solution —such as the *FuelOpps* score— that summarizes all relevant information in a straightforward way.[29]

## 2.7 Conclusion

In summary, our results show that drivers in Company 1050 react, in general, positively to *FuelOpps*. On average, a driver in the treatment group saved around 15.6 gallons of fuel per month during the trial (considering 6,000 miles driven per month).[30] And this effect did not vanish completely after the coach support from *FuelOpps* stopped on October 31 (the day the trial ended); drivers in the treatment group reduced their consumption of fuel in around 12.4 gallons in the months after the trial. This effect seems to be unrelated to the time drivers spent idling, which does not show any variation while the *FuelOpps* program was in place. Therefore, all the improvement in fuel efficiency came from a combination of better driving skills and or/lower average speed. Our estimates indicate that, overall, drivers in the treatment group improve their ability to upshift using the lowest number of RPMs (progressive shifting) and keep a constant speed at the minimum RPMs possible (highest gear). Additionally, our results indicate that after the trial had ended, drivers in

[29] According to an industry expert, driver managers "cannot do spreadsheets."

[30] Considering currently prices of gas of around \$3 per gallon, the overall savings per driver per month are about \$50.

the treatment group decreased their average speed in almost -0.26 MPH, which translates into almost an extra 2.25 minutes of driving each day. Considering a driver that is on the road for 23 days in a month, this implies that she drives almost 52 minutes extra each month for no additional compensation.

Given the way rewards were added to the trial, we were able to study the effect of the gamification elements of *FuelOpps* during a period where drivers were only 'promised' a future reward without any indication of the conditions under which it would be awarded. This way, we could quantify the effect of one extra day under in the *FuelOpps* program conditional on the promise of future rewards. According to our estimates, drivers began to lose interest in the system around 20 days after the introduction of the program, with each day after that mark negatively affecting the fuel efficiency of drivers. Although this is not a test of the effect that a gamified platform with no promise of external rewards would have on fuel efficiency, it would be hard to argue that *FuelOpps* would have been remotely as effective as our results show without pecuniary incentives. As discussed in the literature review on gamification at the beginning of this paper, there is little evidence that workers or consumers would constantly be motivated to engage in certain behaviors when the consequences of their acts translate in the same set of non-monetary rewards —such as points or badges— being awarded time after time. Similar to what occurs with games, the feedback provided by a game, although an essential part of it, is not what makes a game interesting in the long term. The novelty effect (Koivisto and Hamari, 2014; Seaborn and Fels, 2015) generated by the introduction of a gamified platform reflects the fact that as time goes by there is little incentive for a user to come back to the platform once its major characteristics have already been discovered and understood.

Although the premise of gamification is simple, its implementation is not. User experience with a platform is not increased automatically by the inclusion of game-like features, as is documented in the literature (Haaranen et al., 2014). An important research effort remains to be done in order to fill our knowledge gaps with respect to the game element/mechanics that workers value and the types of employees more likely to react to the gamified incentive system (Hamari, Koivisto, and Sarsa, 2014). In order to have a successful gamification, it is important not only to focus on the goals of the firm implementing the gamified application, but also on the experience of the individuals that will engage with the platform (Hamari, Huotari, and Tolvanen, 2015). We need to remember that points and badges, although useful in their own terms to communicate relevant information to individuals, are not ends by themselves. They just play a secondary role in making an experience more gameful.

Table 2.1: *FuelOpps* levels and experience points.

| Level | Exp. Points | Level | Exp. Points |
|---|---|---|---|
| F1 | 0 | F9 | 90 |
| F2 | 3 | F10 | 120 |
| F3 | 7 | F11 | 155 |
| F4 | 13 | F12 | 200 |
| F5 | 20 | F13 | 250 |
| F6 | 30 | F14 | 500 |
| F7 | 45 | F15 | $1,000$ |
| F8 | 65 | F16 | $2,000$ |

Table 2.2: MPG and GPM.

| Case | Car | MPG | $\Delta$ MPG | GPM | $\Delta$ GPM |
|---|---|---|---|---|---|
| A | A1 | 1 | | 1 | |
| A | A2 | 2 | 1 | 0.5 | $-0.5$ |
| B | B3 | 25 | | 0.04 | |
| B | B4 | 40 | 15 | 0.025 | $-0.015$ |

Table 2.3: Descriptive statistics.

| Group | Variable | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|---|
| Treatment | Observations | 3,665 | 6,290 | 6,152 |
| | Drivers | 118 | 118 | 107 |
| | GPM | 0.149 | 0.148 | 0.161 |
| | Fuel (gallons per day) | 46.230 | 45.043 | 49.621 |
| | Fuel: driving | 45.406 | 44.299 | 48.625 |
| | Fuel: idling | 0.824 | 0.744 | 0.997 |
| | Move time (minutes) | 383.399 | 379.256 | 377.916 |
| | Idle time (minutes) | 26.960 | 22.275 | 45.671 |
| | Distance (100 miles) | 3.134 | 3.080 | 3.101 |
| | Short idle time | 24.454 | 20.173 | 37.321 |
| Internal | Observations | 8,641 | 15,562 | 16,468 |
| | Drivers | 285 | 283 | 284 |
| | GPM | 0.151 | 0.152 | 0.164 |
| | Fuel (gallons) | 45.251 | 45.633 | 49.018 |
| | Fuel: driving | 44.416 | 44.828 | 47.978 |
| | Fuel: idling | 0.835 | 0.804 | 1.040 |
| | Move time | 359.941 | 360.825 | 358.743 |
| | Idle time (minutes) | 40.109 | 35.661 | 53.493 |
| | Distance (100 miles) | 3.031 | 3.043 | 3.022 |
| | Short idle time (minutes) | 31.411 | 28.732 | 38.978 |
| External | Observations | 9,479 | 16,053 | 14,432 |
| | Drivers | 303 | 302 | 268 |
| | GPM | 0.155 | 0.156 | 0.168 |
| | Fuel (gallons) | 35.645 | 36.232 | 38.930 |
| | Fuel: driving | 34.678 | 35.288 | 37.682 |
| | Fuel: idling | 0.967 | 0.944 | 1.249 |
| | Move time | 313.288 | 315.069 | 314.098 |
| | Idle time (minutes) | 40.172 | 39.795 | 69.303 |
| | Distance (100 miles) | 2.343 | 2.365 | 2.339 |
| | Short idle time | 32.847 | 32.123 | 44.528 |

Table 2.4: GPM: main results.

| Variable | (1) GPM | (2) GPM | (3) GPM-D | (4) GPM-D | (5) GPM-I | (6) GPM-I |
|---|---|---|---|---|---|---|
| Pre-Trial Messages | −0.00214*** | −0.00214*** | −0.00191*** | −0.00191*** | −0.00023** | −0.00023** |
| | (0.00049) | (0.00049) | (0.00048) | (0.00048) | (0.00010) | (0.00010) |
| FuelOpps (WP) | −0.00260*** | | −0.00260*** | | 0.00001 | |
| | (0.00056) | | (0.00054) | | (0.00009) | |
| FuelOpps | | −0.00318*** | | −0.00306*** | | −0.00012 |
| | | (0.00054) | | (0.00054) | | (0.00008) |
| FuelOpps + Prize | | −0.00230*** | | −0.00239*** | | 0.00009 |
| | | (0.00064) | | (0.00061) | | (0.00011) |
| FuelOpps + Prize + T-Shirt | | −0.00177** | | −0.00191*** | | 0.00014 |
| | | (0.00077) | | (0.00074) | | (0.00012) |
| Post-Trial Phase | −0.00207** | −0.00202** | −0.00192** | −0.00188** | −0.00016 | −0.00015 |
| | (0.00086) | (0.00087) | (0.00085) | (0.00086) | (0.00014) | (0.00014) |
| Distance | −0.00992*** | −0.00992*** | −0.00644*** | −0.00644*** | −0.00348*** | −0.00348*** |
| | (0.00051) | (0.00051) | (0.00046) | (0.00046) | (0.00011) | (0.00011) |
| Distance (Squared) | 0.00101*** | 0.00101*** | 0.00065*** | 0.00065*** | 0.00036*** | 0.00036*** |
| | (0.00007) | (0.00007) | (0.00006) | (0.00006) | (0.00002) | (0.00002) |
| Short Idle Time | 0.00030*** | 0.00030*** | 0.00021*** | 0.00021*** | 0.00009*** | 0.00009*** |
| | (0.00001) | (0.00001) | (0.00001) | (0.00001) | (0.00000) | (0.00000) |
| Constant | 0.14955*** | 0.14949*** | 0.14234*** | 0.14230*** | 0.00721*** | 0.00719*** |
| | (0.00510) | (0.00509) | (0.00451) | (0.00451) | (0.00092) | (0.00092) |
| Observations | 96,742 | 96,742 | 96,742 | 96,742 | 96,742 | 96,742 |
| R-squared | 0.46329 | 0.46337 | 0.41966 | 0.41973 | 0.35773 | 0.35779 |
| Drivers | 706 | 706 | 706 | 706 | 706 | 706 |

Day, vehicle, and driver fixed effects included in all models.
Standard errors (in parentheses) clustered at the driver level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2.5: Behavioral response

| Variable | (1) Progressive Shifting | (2) Highest Gear | (3) Highway Speed | (4) Speed | (5) Move Time | (6) Over RPM |
|---|---|---|---|---|---|---|
| Pre-Trial Messages | 1.50622*** | 3.33785*** | 2.67906*** | −0.52136*** | 4.22446*** | −0.14229 |
| | (0.33317) | (0.69651) | (0.61868) | (0.15188) | (1.06290) | (0.23362) |
| FuelOpps | 3.42566*** | 9.26657*** | 4.25720*** | −0.86941*** | 6.57870*** | −0.38358* |
| | (0.35336) | (0.97517) | (0.86755) | (0.13139) | (0.92854) | (0.23146) |
| FuelOpps + Prize | 3.85398*** | 11.34872*** | 5.23853*** | −0.72875*** | 5.82923*** | −0.77972** |
| | (0.40078) | (1.13867) | (0.92793) | (0.16141) | (1.17816) | (0.38691) |
| FuelOpps + Prize + T-Shirt | 3.82721*** | 11.89595*** | 5.51320*** | −0.77596*** | 5.99873*** | −0.93077** |
| | (0.42522) | (1.25146) | (0.99291) | (0.16489) | (1.14189) | (0.45608) |
| Post-Trial Phase | 2.89449*** | 8.73452*** | 3.57827*** | −0.26315** | 2.24686** | −0.47070 |
| | (0.32531) | (0.88254) | (0.59623) | (0.12979) | (0.90597) | (0.34992) |
| Distance | 0.55703*** | 1.70654*** | −3.63590*** | 6.66317*** | 117.06373*** | 0.35084*** |
| | (0.08318) | (0.21305) | (0.30856) | (0.17039) | (0.74107) | (0.11073) |
| Distance (Squared) | −0.07164*** | −0.11264*** | 0.38178*** | −0.59325*** | −2.53101*** | −0.03772** |
| | (0.01234) | (0.03116) | (0.04175) | (0.02271) | (0.11865) | (0.01510) |
| Short Idle Time | 0.02051*** | −0.04798*** | 0.01982*** | −0.15273*** | 1.08122*** | 0.01172*** |
| | (0.00161) | (0.00385) | (0.00474) | (0.00242) | (0.01490) | (0.00165) |
| Constant | 86.06473*** | 56.38521*** | 55.97143*** | 38.57019*** | 8.34385** | |
| | (1.19025) | (3.02409) | (2.02449) | (0.56084) | (3.32708) | |
| Observations | 92,642 | 92,642 | 92,642 | 92,742 | 92,742 | 36,092 |
| R-squared | 0.10976 | 0.24650 | 0.15691 | 0.58328 | 0.96714 | |
| Log pseudologlikelihood | | | | | | -18,975 |
| Number of drivers | 706 | 706 | 706 | 706 | 706 | 262 |

Day, vehicle, and driver fixed effects included in all models.
Standard errors (in parentheses) clustered at the driver level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2.6: Fuel efficiency (GPM) by quartile.

| Variable | (1) Worst | (2) Bad | (3) Good | (4) Top |
|---|---|---|---|---|
| Pre-Trial Messages | −0.00092 | −0.00247** | −0.00252*** | −0.00289*** |
| | (0.00104) | (0.00103) | (0.00094) | (0.00098) |
| FuelOpps | −0.00346*** | −0.00222* | −0.00428*** | −0.00273*** |
| | (0.00125) | (0.00121) | (0.00097) | (0.00077) |
| FuelOpps + Prize | −0.00369** | −0.00163 | −0.00331*** | −0.00109 |
| | (0.00153) | (0.00116) | (0.00119) | (0.00107) |
| FuelOpps + Prize + T-Shirt | −0.00277* | −0.00058 | −0.00254* | −0.00209 |
| | (0.00156) | (0.00184) | (0.00145) | (0.00137) |
| Post-Trial Phase | −0.00108 | −0.00125 | −0.00381** | −0.00244 |
| | (0.00195) | (0.00177) | (0.00156) | (0.00169) |
| Distance | −0.01205*** | −0.00949*** | −0.01030*** | −0.00732*** |
| | (0.00099) | (0.00101) | (0.00102) | (0.00099) |
| Distance (Squared) | 0.00123*** | 0.00091*** | 0.00110*** | 0.00073*** |
| | (0.00013) | (0.00013) | (0.00014) | (0.00013) |
| Short Idle Time | 0.00034*** | 0.00030*** | 0.00027*** | 0.00028*** |
| | (0.00001) | (0.00001) | (0.00002) | (0.00001) |
| Constant | 0.17147*** | 0.16972*** | 0.14224*** | 0.14054*** |
| | (0.00174) | (0.00189) | (0.00214) | (0.00181) |
| Observations | 23,967 | 24,348 | 24,582 | 23,701 |
| R-squared | 0.43001 | 0.48749 | 0.48981 | 0.51275 |
| Number of drivers | 175 | 177 | 176 | 177 |

Day, vehicle, and driver fixed effects included in all models.
Standard errors (in parentheses) clustered at the driver level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2.7: Novelty effects.

| Variable | (1) All | (2) Worst | (3) Bad | (4) Good | (5) Top |
|---|---|---|---|---|---|
| Pre-Trial Messages | −0.00191*** | −0.00070 | −0.00215** | −0.00220*** | −0.00271*** |
| | (0.00044) | (0.00095) | (0.00091) | (0.00080) | (0.00085) |
| Days | −0.00972*** | −0.00798* | −0.00668 | −0.01562*** | −0.00791*** |
| | (0.00195) | (0.00447) | (0.00427) | (0.00355) | (0.00295) |
| Days (Squared) | 0.00667*** | 0.00516 | 0.00374 | 0.01250*** | 0.00492* |
| | (0.00175) | (0.00400) | (0.00386) | (0.00321) | (0.00269) |
| Distance | −0.01127*** | −0.01465*** | −0.01003*** | −0.01123*** | −0.00902*** |
| | (0.00060) | (0.00120) | (0.00137) | (0.00112) | (0.00110) |
| Distance (Squared) | 0.00120*** | 0.00161*** | 0.00098*** | 0.00123*** | 0.00098*** |
| | (0.00008) | (0.00017) | (0.00018) | (0.00015) | (0.00015) |
| Short Idle Time | 0.00035*** | 0.00038*** | 0.00035*** | 0.00032*** | 0.00034*** |
| | (0.00001) | (0.00002) | (0.00002) | (0.00002) | (0.00002) |
| Constant | 0.16002*** | 0.17693*** | 0.16156*** | 0.15745*** | 0.13781*** |
| | (0.00175) | (0.00186) | (0.00239) | (0.00179) | (0.00171) |
| Observations | 38,017 | 9,570 | 9,605 | 9,556 | 9,235 |
| R-squared | 0.32124 | 0.36481 | 0.32096 | 0.31744 | 0.33815 |
| Number of drivers | 706 | 175 | 177 | 176 | 177 |

Day, vehicle, and driver fixed effects included in all models.
Standard errors (in parentheses) clustered at the driver level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2.8: Target score

| Variable | (1)<br>All | (2)<br>SDS |
|---|---|---|
| Pre-Trial Messages | -0.00214*** | -0.00215*** |
| | (0.00049) | (0.00049) |
| FuelOpps | -0.00318*** | -0.00319*** |
| | (0.00054) | (0.00054) |
| FuelOpps + Prize | -0.00230*** | -0.00231*** |
| | (0.00064) | (0.00064) |
| FuelOpps + Prize + T-Shirt | -0.00177** | -0.00229*** |
| | (0.00077) | (0.00081) |
| Shirt Giveaway Score Gap (Target - Current) | | 0.00243*** |
| | | (0.00083) |
| Post-Trial Phase | -0.00202** | -0.00202** |
| | (0.00087) | (0.00087) |
| Distance | -0.00992*** | -0.00993*** |
| | (0.00051) | (0.00051) |
| Distance (Squared) | 0.00101*** | 0.00101*** |
| | (0.00007) | (0.00007) |
| Short Idle Time | 0.00030*** | 0.00030*** |
| | (0.00001) | (0.00001) |
| Constant | 0.14949*** | 0.14949*** |
| | (0.00509) | (0.00510) |
| Observations | 96,742 | 96,537 |
| R-squared | 0.46337 | 0.46365 |
| Number of drivers | 706 | 706 |

Day, vehicle, and driver fixed effects included in all models.
Standard errors (in parentheses) clustered at the driver level.s
*** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Interaction

| Variable | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| PT Messages | −0.00246∗∗∗ | −0.00246∗∗∗ | −0.00246∗∗∗ | −0.00246∗∗∗ | −0.00246∗∗∗ |
| | (0.00067) | (0.00067) | (0.00067) | (0.00067) | (0.00067) |
| PT Messages × Active | 0.00057 | 0.00057 | 0.00057 | 0.00060 | 0.00060 |
| | (0.00086) | (0.00086) | (0.00086) | (0.00086) | (0.00086) |
| FuelOpps | −0.00216∗∗∗ | −0.00219∗∗∗ | −0.00210∗∗∗ | −0.00215∗∗∗ | −0.00210∗∗∗ |
| | (0.00054) | (0.00054) | (0.00053) | (0.00054) | (0.00054) |
| FuelOpps × Active | −0.00185∗∗ | −0.00186∗∗ | −0.00202∗∗ | −0.00087 | −0.00106 |
| | (0.00092) | (0.00092) | (0.00091) | (0.00094) | (0.00093) |
| FuelOpps + Prize | −0.00050 | −0.00054 | −0.00045 | −0.00050 | −0.00045 |
| | (0.00068) | (0.00068) | (0.00069) | (0.00068) | (0.00069) |
| FuelOpps + Prize × Active | −0.00321∗∗∗ | −0.00322∗∗∗ | −0.00337∗∗∗ | −0.00255∗∗ | −0.00274∗∗∗ |
| | (0.00109) | (0.00109) | (0.00108) | (0.00103) | (0.00103) |
| FuelOpps + Prize + T-Shirt | 0.00033 | 0.00026 | 0.00039 | 0.00033 | 0.00039 |
| | (0.00137) | (0.00138) | (0.00139) | (0.00137) | (0.00139) |
| F + P + T × Active | −0.00374∗∗ | −0.00375∗∗ | −0.00395∗∗ | −0.00297∗ | −0.00322∗ |
| | (0.00168) | (0.00168) | (0.00170) | (0.00166) | (0.00168) |
| Post-Trial | −0.00182∗ | −0.00182∗ | −0.00181∗ | −0.00182∗ | −0.00181∗ |
| | (0.00106) | (0.00106) | (0.00106) | (0.00106) | (0.00106) |
| Post-Trial × Active | −0.00038 | −0.00039 | −0.00042 | −0.00028 | −0.00032 |
| | (0.00153) | (0.00153) | (0.00152) | (0.00153) | (0.00153) |
| Shirt Giveaway Score Gap | −0.00039 | −0.00035 | −0.00039 | −0.00039 | −0.00038 |
| | (0.00196) | (0.00198) | (0.00196) | (0.00197) | (0.00197) |
| SGSC × Active | 0.00304 | 0.00299 | 0.00302 | 0.00242 | 0.00238 |
| | (0.00214) | (0.00215) | (0.00213) | (0.00208) | (0.00208) |
| Messages | | 0.00016 | −0.00021 | | −0.00020 |
| | | (0.00026) | (0.00041) | | (0.00041) |
| Messages × Active | | | 0.00062 | | 0.00077 |
| | | | (0.00050) | | (0.00049) |
| Calls | | 0.00226∗∗ | 0.00004 | | 0.00002 |
| | | (0.00106) | (0.00543) | | (0.00543) |
| Calls × Active | | | 0.00237 | | 0.00309 |
| | | | (0.00552) | | (0.00553) |
| Sessions | | | | −0.00098∗∗∗ | −0.00099∗∗∗ |
| | | | | (0.00032) | (0.00031) |
| Distance | −0.00993∗∗∗ | −0.00993∗∗∗ | −0.00993∗∗∗ | −0.00992∗∗∗ | −0.00992∗∗∗ |
| | (0.00051) | (0.00051) | (0.00051) | (0.00051) | (0.00051) |
| Distance (Squared) | 0.00101∗∗∗ | 0.00101∗∗∗ | 0.00101∗∗∗ | 0.00101∗∗∗ | 0.00101∗∗∗ |
| | (0.00007) | (0.00007) | (0.00007) | (0.00007) | (0.00007) |
| Short Idle Time | 0.00030∗∗∗ | 0.00030∗∗∗ | 0.00030∗∗∗ | 0.00030∗∗∗ | 0.00030∗∗∗ |
| | (0.00001) | (0.00001) | (0.00001) | (0.00001) | (0.00001) |
| Constant | 0.14953∗∗∗ | 0.14953∗∗∗ | 0.14955∗∗∗ | 0.14918∗∗∗ | 0.14919∗∗∗ |
| | (0.00510) | (0.00510) | (0.00510) | (0.00505) | (0.00505) |
| Observations | 96,537 | 96,537 | 96,537 | 96,537 | 96,537 |
| R-squared | 0.46395 | 0.46396 | 0.46397 | 0.46445 | 0.46449 |
| Number of driver | 706 | 706 | 706 | 706 | 706 |

Day, vehicle, and driver fixed effects included in all models.
Standard errors (in parentheses) clustered at the driver level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Figure 2.1: Point allocation function for a subset of number of miles traveled.

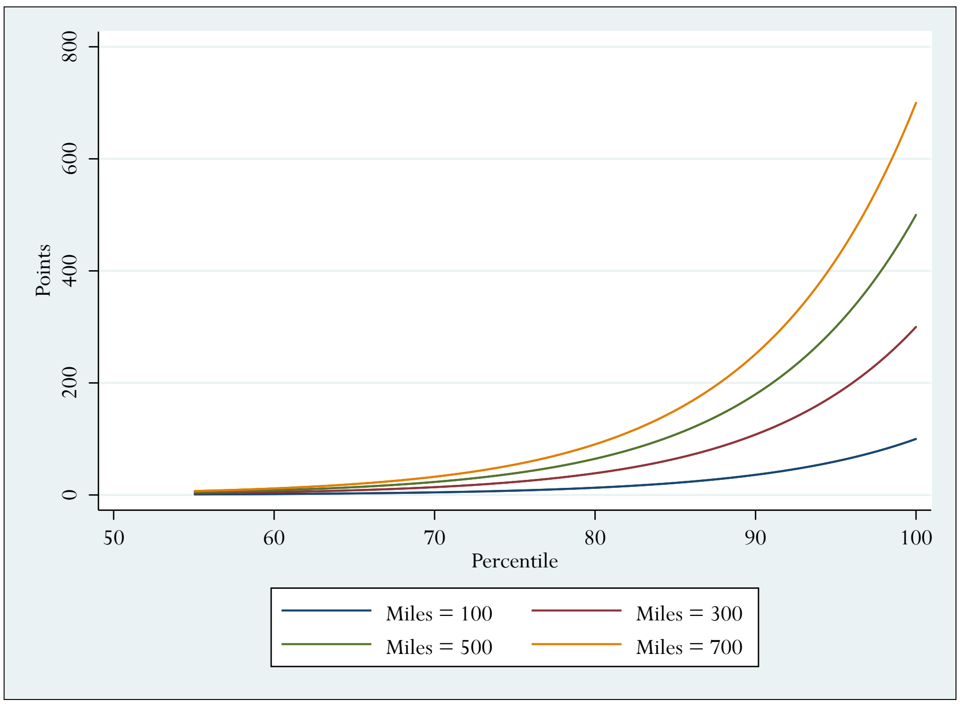


Figure 2.2: Website screenshot of *FuelOpps* version 1.0.

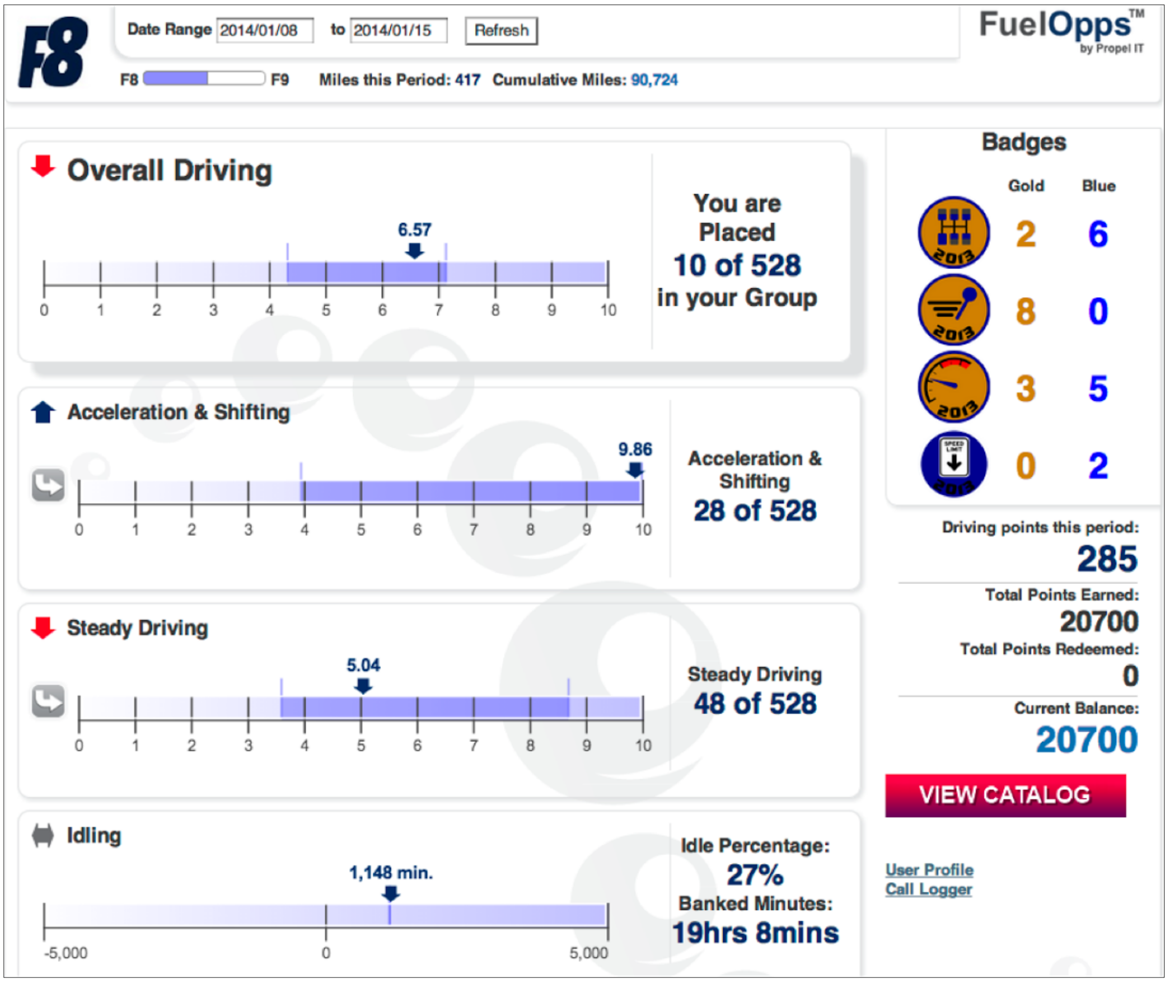


Figure 2.3: Timeline of events for company 1050.

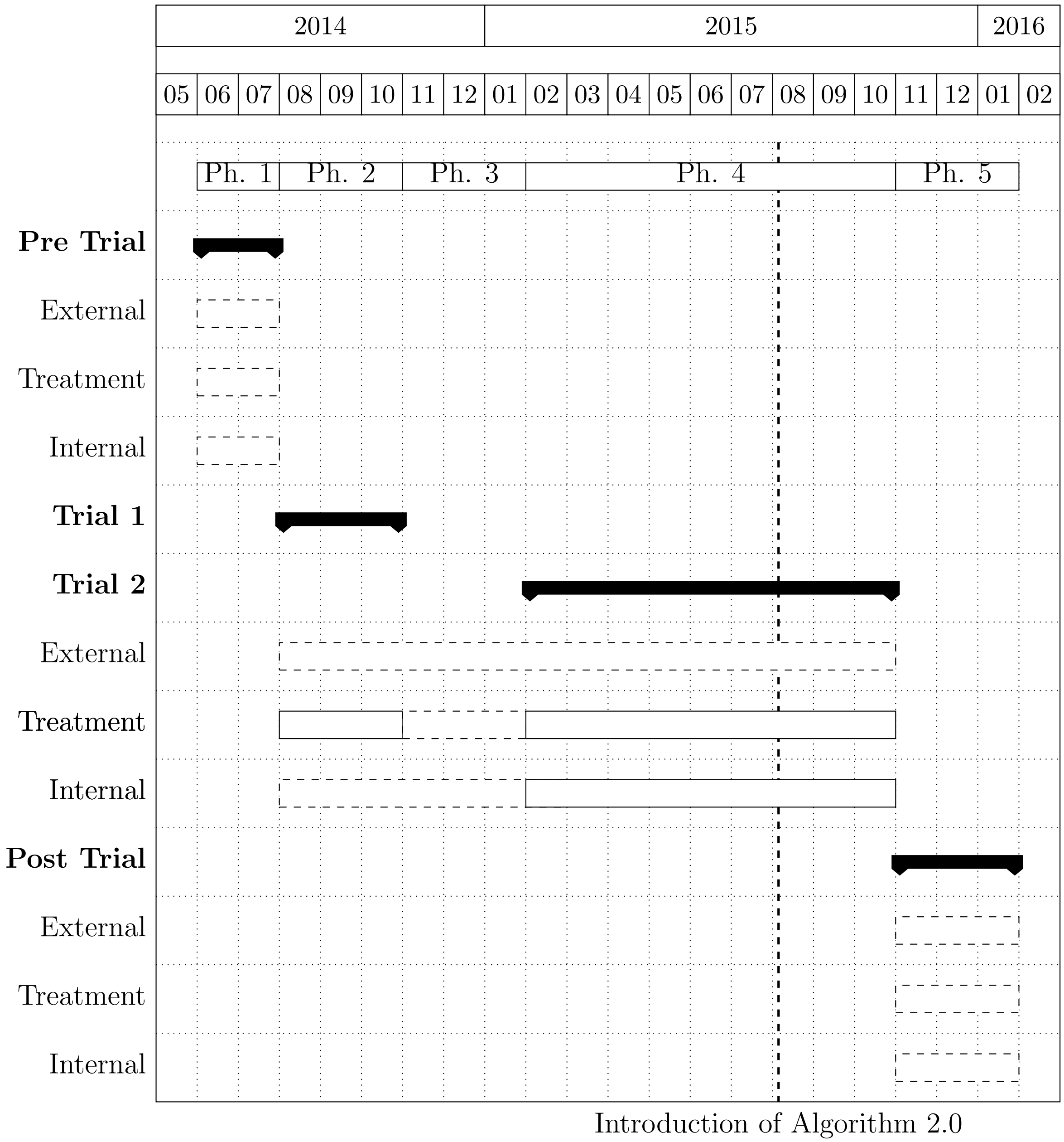


Figure 2.4: Characteristics of *FuelOpps* during phase 2.

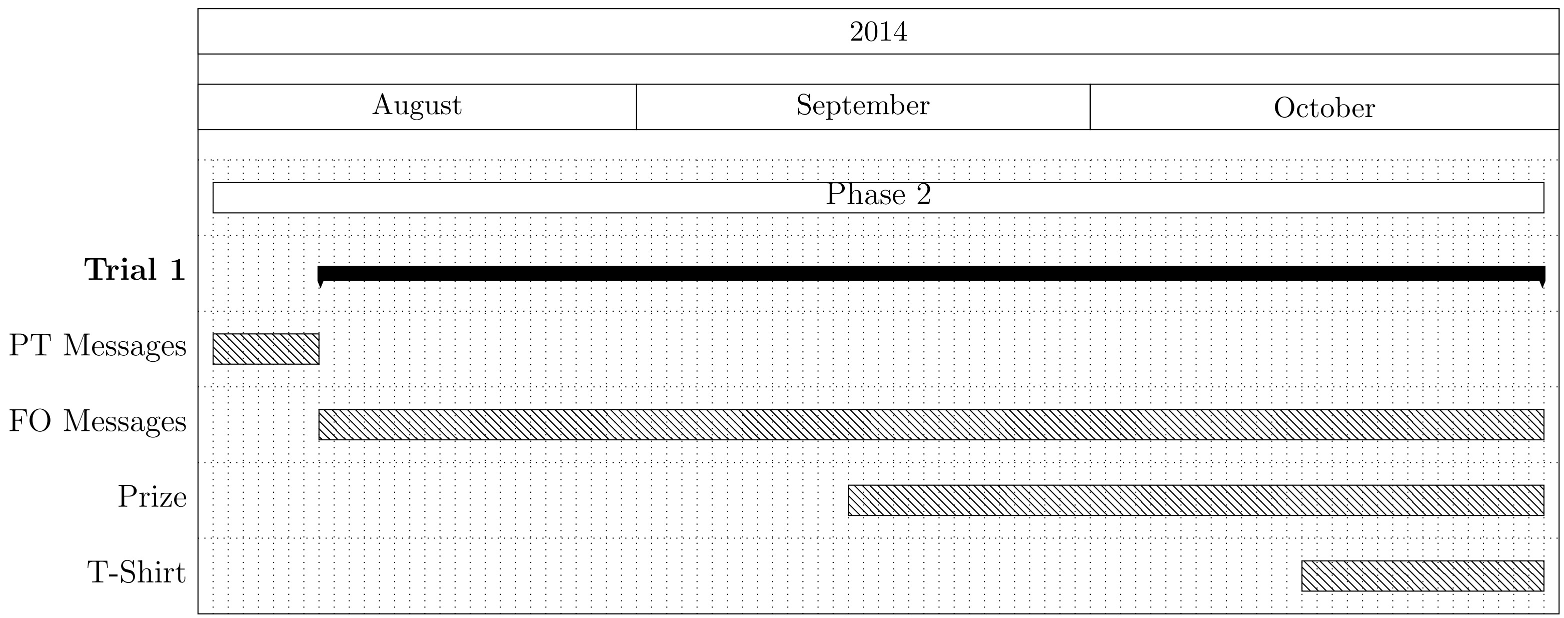


Figure 2.5: Average fuel efficiency (GPM) by group per day.

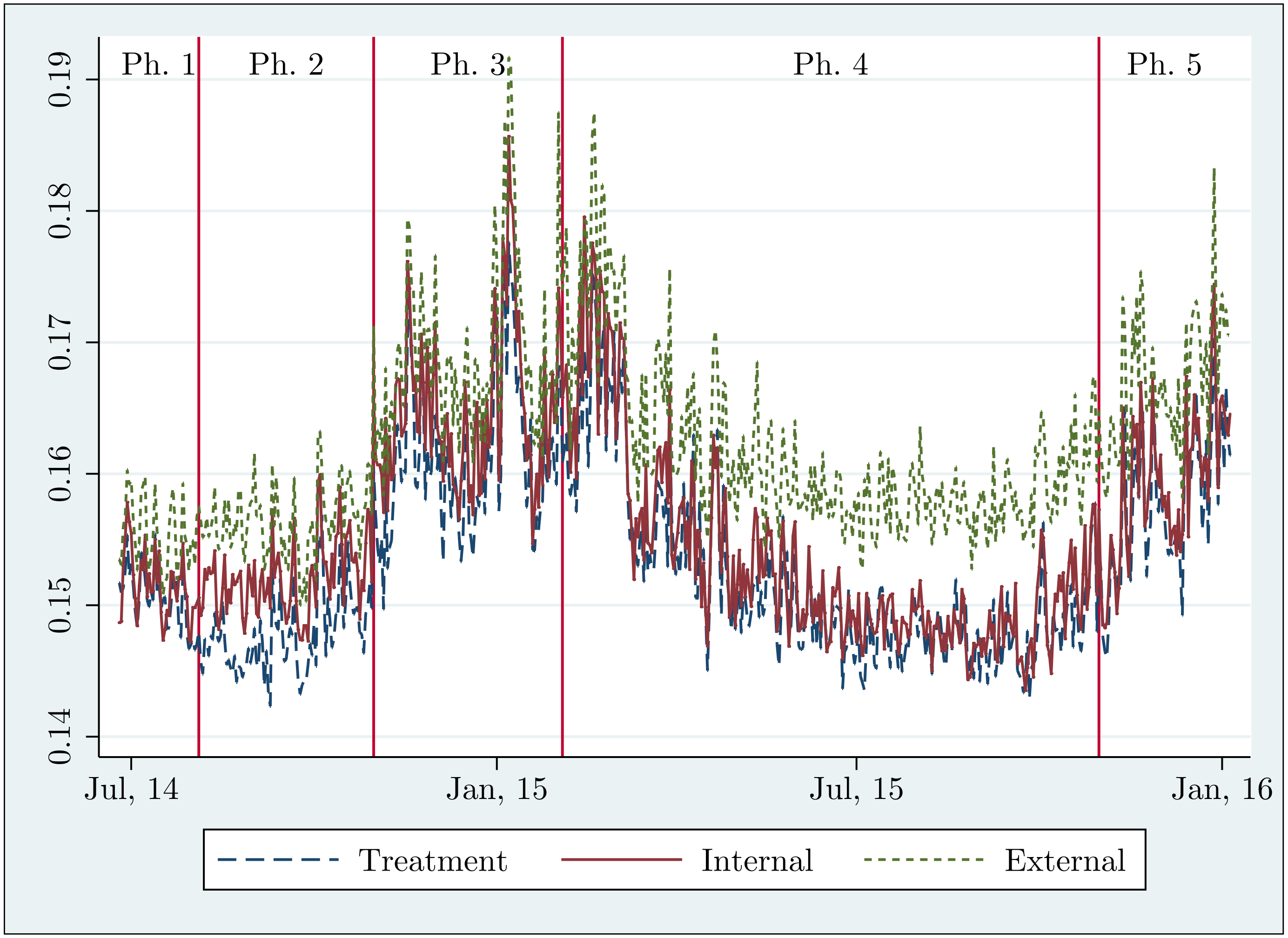


Figure 2.6: Fuel efficiency (GPM) across groups and phases.

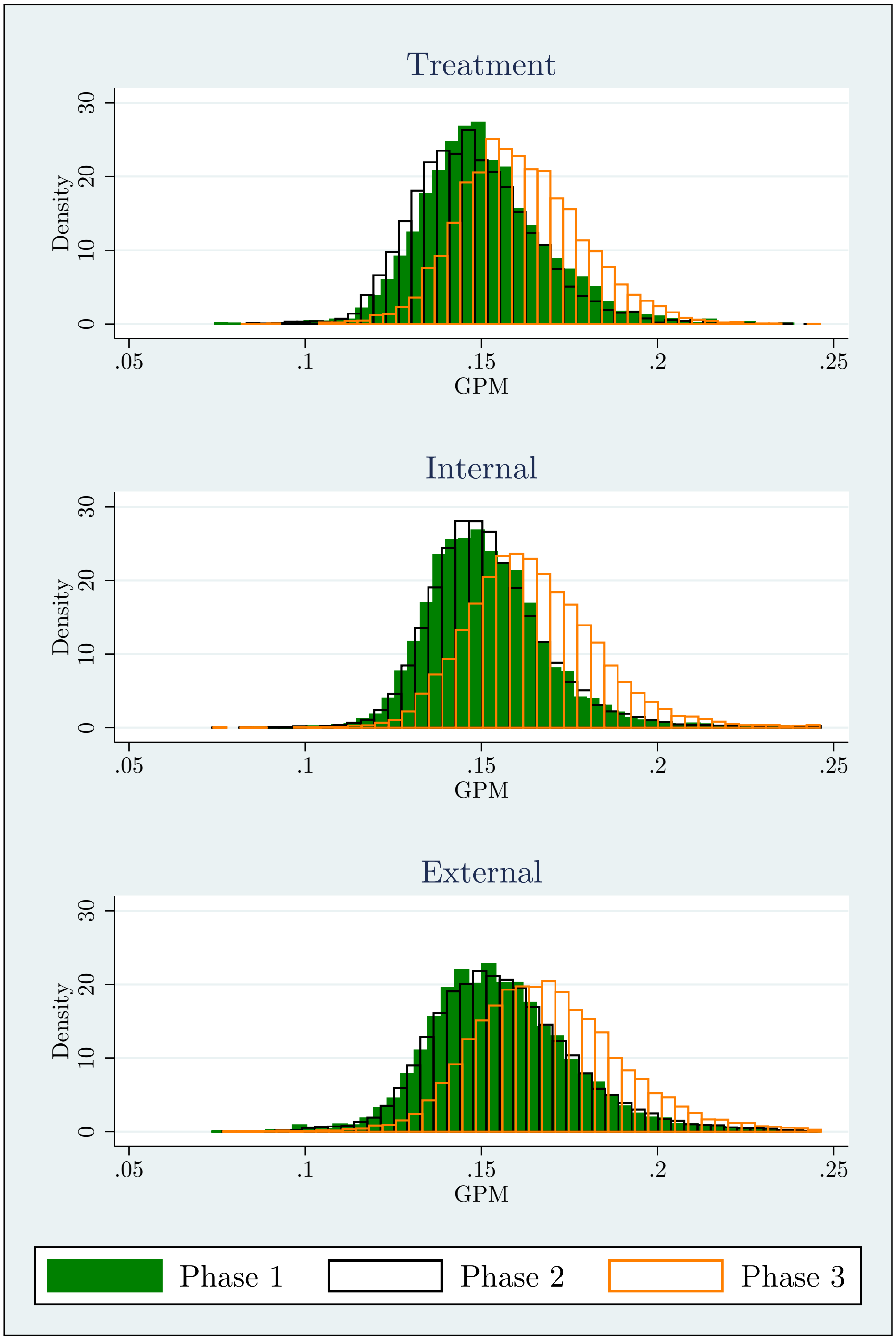


Figure 2.7: Pre-intervention trend (treatment group).

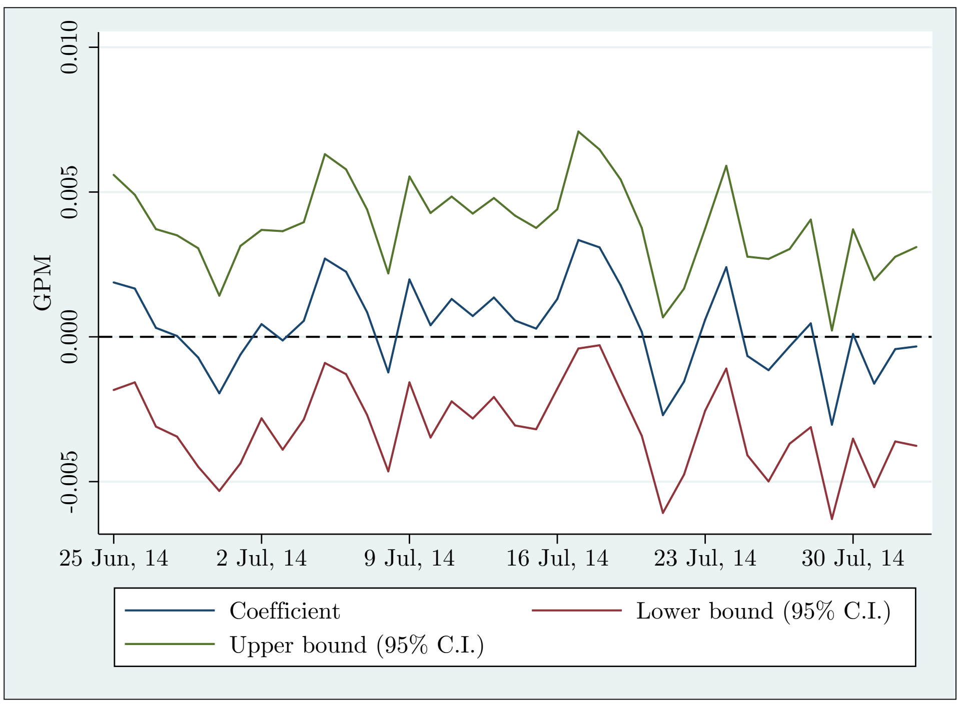


Figure 2.8: Evolution of *FuelOpps* scores over time for the treatment group.

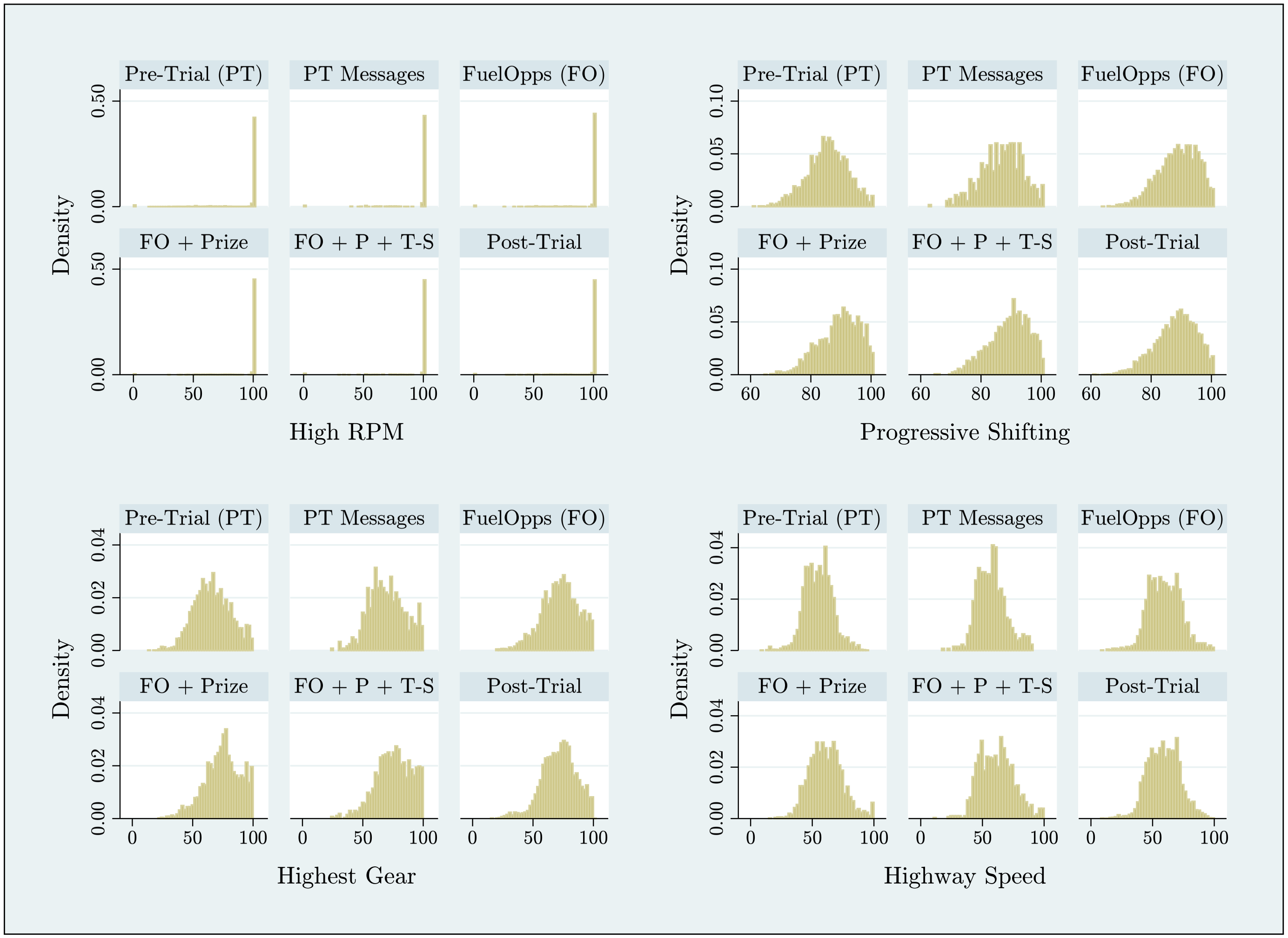
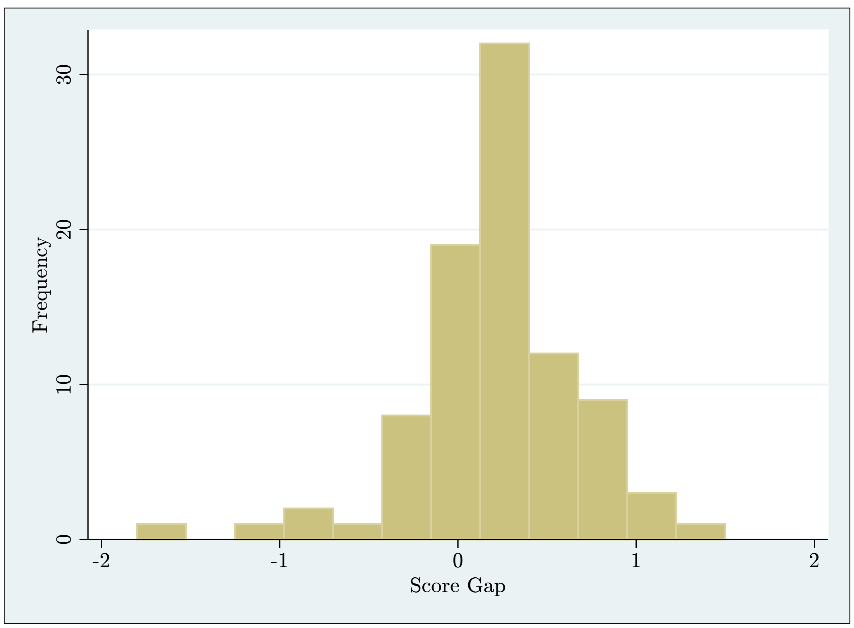


Figure 2.9: Histogram of the difference between target and current score.



## 2.8 Appendix 1: Pre-trial messages

The messages sent to the drivers that participated in the first phase of the trial are below. Text in italics such as *first name* or *company id* was replaced with a driver's specific values when the text was sent.

- **08/04/2014 - Monday**
  Hi *first name*, in 8 days, *company name* will be launching a trial of the FuelOpps Program. The goal of the FuelOpps program is to recognize and reward the best drivers for fuel efficient driving skills NOT their MPG. Over the next several days you will receive your login information and a link to the FuelOpps website, which will enable you to check your daily scores and track your point total. Thanks, The FuelOpps Team.

- **08/05/2014 - Tuesday**
  Hi *first name*, you can receive the recognition you deserve for driving fuel efficiently. Your driver ID identifies you as the driver in the FuelOpps program, so it doesn't matter what or when you are driving. Watch for your login information coming soon. Sincerely, The FuelOpps Team.

- **08/06/2014 - Wednesday**
  Hi *first name*, you can access FuelOpps any anytime. Just go to www.*company code*.fuelopps.com and login by using *driver id* as both your Username and Password. You can also download the FuelOpps mobile app on your smartphone search for the word Fuelopps in the Android or Apple stores. Have any questions? Just reply back with a question or call 1-XXX-XXX-XXX and we will be happy to help. Thanks, The FuelOpps Team.

- **08/07/2014 - Thursday**
  Hi *first name*, Do you wonder how FuelOpps works? The driving components that you are scored on are Progressive Shifting, Keeping RPMs down, Highway Speed, and being in the Highest Gear you can when possible. NOT your MPG. There will be rewards at the end of the trial for doing well. If you would like to know more about the program and the rewards, just reply to this message or give us a call. Thanks, The FuelOpps Team.

- **08/08/2014 - Friday**
  Hi *first name* The FuelOpps Trial starts on Monday. To reach a FuelOpps team member

to learn more about the program and gain the secrets of scoring well, please give us a call at 1-XXX-XXX-XXXX. Your personal Phone Pin number to use when first calling in is *pin number* Thanks again for participating, FuelOpps.

- **08/11/2014 - Monday**
  Hi *first name* The FuelOpps Trial begins today. If you have not already had a chance to login and see what FuelOpps is all about, please give it a quick try. Also, if you prefer you can download the FuelOpps Smartphone app for either Android or IOS devices or go to www.*company code*.fuelopps.com and use *driver id* as both your username and password. Thanks and good luck during the trial, FuelOpps.

## 2.9 Appendix 2: About the differences-in-differences approach

The simplest estimate of a differences-in-differences model that considers multiple observations of every individual ($i$) across time ($t = 1, 2$) is presented below (Wooldridge, 2010).

$$y_{it} = \alpha_0 + \alpha_1 D\_2_t + \alpha_2 D\_TG_i + \beta treatment_{it} + u_{it}. \tag{2.6}$$

In Equation 2.6, $D\_2$ and $D\_TG$ are dichotomous variables that control for characteristics in the second period that could affect the outcome we are measuring ($y$) and for general attributes present in the treatment group that could explain differences in our dependent variable, respectively. $\beta$ represents the effect of the treatment, and it can be understood as the difference in change of the mean responses between periods one and two for the treatment and control groups (Wooldridge, 2010). This simple model does not include covariates that control for time-invariant individual characteristics, but including them will not change the estimate or the interpretation of coefficient $\beta$.

$$y_{it} = \alpha_0 + \alpha_1 D\_2 + \beta treatment_{it} + v_i + u_{it}. \tag{2.7}$$

Equation 2.7 includes a set of $v_i$ terms that capture time-invariant differences across individuals (the inclusion of these terms implies that we need to drop the constant, $\alpha_0$, that appear in Equation 2.6 or one of the $v_i$). For simplicity, assume these $v_i$ terms represent the coefficients of a set of dichotomous variables (one for each individual). Adding these covariates to Equation 2.6 implies that the coefficient of $D\_TG$ cannot be identified and needs to be dropped from our model ($D\_TG$ is a perfect linear combination of the set of dichotomous variables for the individuals in the treatment group). However, even with the inclusion of this new set of variables, the estimate and interpretation of $\beta$ remains unchanged. To show this, we can first-difference Equation 2.7, which leave us with $\triangle y_i = \alpha_1 + \beta \triangle treatment_i + \triangle u_i$. The OLS estimate of $\beta$ can be expressed as $\hat{\beta} = \overline{\triangle y}_{treatment} - \overline{\triangle y}_{control}$, which is the difference in our dependent variable for the treatment and control groups between the two periods (Wooldridge, 2010). Equation 2.7 can be easily extended to include additional regressors to control for other potential confounding variables that might affect our estimate of $\beta$.

# CHAPTER 3

# Stakeholder value appropriation: The case of labor in the worldwide mining industry

## 3.1 Introduction

The question of who appropriates the value generated by a firm is central to understanding differences in performance across organizations (Brandenburger and Stuart, 1996). Despite its importance, the strategic management literature has given little attention to how the value generated by a firm is allocated among its stakeholders and to the determinants of this allocation (Asher, Mahoney, and Mahoney, 2005). The few studies that have empirically evaluated value appropriation have focused on analyzing whether the firm or its customers appropriates value (Bennett, 2013; Grennan, 2014) rather than on studying the determinants of value appropriation at the firm level.

In this article, we evaluate how the value appropriated by employees varies in response to an exogenous shock to the price of the firm's product and how this variation depends on institutional and ownership structures. Because employees are major stakeholders that support the firm's value generation process (Molloy and Barney, 2015), our research informs an issue that interests both scholars and practitioners: the relationship between the stakeholder view of strategy and the dynamics of value appropriation (Amit and Zott, 2001; Brandenburger and Stuart, 1996; Garcia-Castro and Aguilera, 2015).

The effect of a change in the value generated by the firm on the value appropriated by employees depends on the employees' relative bargaining power (Brandenburger and Stuart, 1996). An important determinant of this power is labor regulation (Botero et al., 2004; DiNardo, Fortin, and Lemieux, 1996; Rodrik, 1999). Regulations that favor firms over employees may decrease the amount allocated to the latter in response to an increase in the value generated by a firm, whereas the opposite is true for regulations that favor employees over firms. Another determinant of employee bargaining power is the type and concentration of ownership (Sapienza, 2004). Because of a more diffuse objective function, the management of stated-owned enterprises may be more sympathetic than the management of privately owned enterprises to employee demands when there are more resources to be distributed (Bertrand and Mullainathan, 2003; Inoue, Lazzarini, and Musacchio, 2013), whereas a more concentrated ownership can improve the monitoring capacity of owners over managers, decreasing the intensity of principal-agent problems and the likelihood of larger transfers to employees (Shleifer and Vishny, 1986).

The empirical analysis of the moderating effects of labor regulations and type and concentration of ownership on the value appropriated by employees seeks to assess the effects of institutional determinants and ownership configurations on employees' ability to appropriate value (Greckhamer, 2016) and to inform managers about characteristics that are important to consider when determining the countries and types of firms in which to invest.

Our setting is the mining industry. We have very detailed information about the performance and characteristics of the largest copper mines in the world, which are located in different countries and have different ownership structures. The mining industry provides a good setting in which to study how value is distributed between employees and employers for various reasons. First, copper mines sell a commodity, and no mine accounts for a large share of the total quantity sold in the market; thus, prices are exogenously determined. Unlike other industries in which different buyers can negotiate distinct prices for the same product from the same supplier (Grennan, 2014), price exogeneity allows us to isolate the effects of negotiations between firms and consumers in the determination of the value appropriated by employees when there is a change in the financial prospects of the firm. Second, our dataset includes a period in which a large exogenous shock to copper prices is observed, providing a good opportunity to explore variation over time in our variables of interest. Third, because copper mines are located in a large number of countries and have different ownership structures, we can have a clean measure of the effects of regulatory and institutional variables on the amount of value appropriated by employees.

Our main results show that the value appropriated by employees rises in response to an exogenous increase in the price of the firm's product. Our results also indicate that the value captured by employees as a consequence of a positive shock to the firm's product price is larger in stated-owned companies, when labor regulations promote productivity-based payments, when wages are determined through a centralized bargaining process, and when regulations associated with hiring and firing are more flexible. Overall, the results presented in this article, which are robust to the use of different measures and specifications, suggest that employees benefit from higher firm cash flows driven by an exogenous shock, but the magnitude of this benefit greatly depends on the institutional setting and the firm's ownership structure.

## 3.2 Theory and hypotheses

### 3.2.1 General considerations

Brandenburger and Stuart (1996) introduced formally the concepts of value creation and appropriation, and since then, these concepts have become keystones in competitive strategy research ("Value creation, competition, and performance in buyer-supplier relationships" 2010). An increasing interest in stakeholders as claimants and appropriators of value in their interactions with the firm has opened up opportunities to investigate further the returns to stakeholders different from shareholders (Garcia-Castro and Aguilera, 2015).

The distribution of the value generated by the firm among its different stakeholders affects managerial decisions (Obloj and Sengul, 2012). For instance, shareholders of firms with positive price prospects for their products may prefer to enter markets in which they expect to appropriate all or most of the additional value generated by the higher price and avoid markets in which other stakeholders are expected to appropriate this value. An important determinant of the division of value between shareholders and other stakeholders is the relative bargaining power of each party (MacDonald and Ryall, 2004). This bargaining power depends on regulatory and institutional variables (Coff, 1999) and ownership structures that can affect managerial motivations (Inoue, Lazzarini, and Musacchio, 2013; Shleifer and Vishny, 1986).

In one of the few empirical studies of the dynamics of value creation and stakeholder value appropriation, Bennett (2013) analyzes how the organizational structure of the sales process affects the portion of value appropriated by customers and firms in U.S. car dealerships, concluding that firms that organize their sales process serially, where customers are first served by less experienced salespeople and are then approached by more senior clerks as the difficulty of closing the deal increases, are able to increase their bargaining power and reduce customers' outside alternatives. In another article, Grennan (2014) uses a model of buyer demand and buyer-supplier bargaining to estimate both firm and customer bargaining abilities in the U.S. coronary stent market, wherein different buyers (hospitals) pay different prices for the same product provided by a common supplier. Grennan concludes that supplier costs, buyer willingness-to-pay, and competition determine only the range of potential prices, with the final price depending on the relative bargaining ability of each party. Unlike previous analyses, we focus on the value appropriated by employees, one of the most important firm stakeholders (Coff, 1999).

### 3.2.2 The effect of an exogenous positive shock on employees' value appropriation

Baker (2002) defines the value created by employees as their contribution to the firm's objective and the value appropriated by them as the value they retain. Employees should not appropriate more than their contribution to the transaction (i.e., their marginal product); otherwise, the other stakeholders would be better off without them (MacDonald and Ryall, 2004). Additionally, employees should not receive less than the value provided by the best opportunity they are not taking because of this transaction, i.e., their opportunity cost; otherwise, they would be better off by choosing not to work for the firm (Gans, 2005).

The difference between an employee's marginal product and her outside option defines a range that contains all the potentially acceptable payoffs she could receive from the relationship. The exact payoff received by employees depends on the aforementioned range and their bargaining ability (Brandenburger and Stuart, 2007; Obloj and Sengul, 2012). A positive exogenous shock to the firm's product price increases the value of the employee contribution to the firm's objective function (i.e., the same amount and quality of work delivers more value to the firm) and broadens the range of potential payoffs they could obtain. Alternatively, a negative exogenous shock to the firm's product price decreases the value of the goods or services produced by employees and their marginal contribution to the firm's profits.[1]

Briefly, an exogenous increase in the price of the firm's product increases the (i) employee

[1] A real-life example provides anecdotal support for the relationship between the capacity of employees to appropriate value and their contribution to firm resources. On July 22, 2015, the president of Codelco, a Chilean state-owned company and one of the largest copper mining companies in the world, alleged that copper market conditions did not allow for an expansion of worker benefits (La Tercera, 2015). In 2013, in bilateral negotiations between the same company and its employees, the firm was much more flexible in granting additional financial benefits to its employees. Between mid-2013 and mid-2015, the price of copper decreased approximately by 30 percent.

contribution to the firm's objective function and (ii) firm financial resources. These two consequences facilitate the value appropriation process for employees, leading to hypothesis 1 (H1).

H1: *Employees appropriate more value when there is an exogenous increase in the price of the firm's product.*

### 3.2.3 The moderating effects of labor regulations

Numerous studies indicate that regulations and institutional factors affect the determination of employee compensation. For example, DiNardo, Fortin, and Lemieux (1996) show that de-unionization and minimum wages affect countries' wage distributions. Card, Kramarz, and Lemieux (1999) compare labor market rigidities in Canada, France and the United States, finding that they impact the wage distributions of these countries. Blau and Kahn (1999) emphasize the decentralized nature of wage bargaining in the United States relative to other countries in shaping wages at the bottom end of the wage distribution. Greckhamer (2016) finds that the ratio of CEO compensation to worker pay depends on a number of interdependent institutional factors, including labor regulations. Rodrik (1999) finds that the size of the wage gap across countries is explained by differences in both labor productivity and labor regulations.

Regulations that favor firms over employees may decrease the value appropriated by employees, whereas the opposite is true for regulations that favor employees over firms. By exploring whether an exogenous change in the price of the firm's product affects the value appropriated by employees in different regulatory environments, we can argue in favor or against an institution-based approach to value appropriation.

Given that regulations affect market outcomes, every country has established a set of laws intended to protect the interests of employees and the functioning of the economy (Botero et al., 2004). Countries may provide legal protection for collective bargaining rights to seek a balance of power between labor unions and employers; for example, by empowering labor unions to represent employees collectively (Coff, 1999), the wage bargaining process becomes more centralized (e.g., at the industry level) rather than centered on the firm level.[2] A more centralized wage bargaining process is believed to increase the bargaining

[2] Scholars tend to agree that employees are able to enhance their power relative to that of organizations and capital holders by acting as a collective (Coff, 1999; Greckhamer, 2016).

ability of employees, whereas a more individualized negotiation process is thought to increase the bargaining ability of employers; thus, firms usually express concern about changes in regulations that shift power to unions or that improve collective bargaining (Mason, 2015).

Regulations can also affect flexibility in hiring and firing employees. More flexibility favors employment under fixed-term contracts and allows for easier dismissals, which are usually associated with "employer-friendly" regulations. Conversely, less flexibility usually implies higher costs of firing, including higher severance payments and mandatory penalties, and stricter guards against temporary work, which are both usually associated with "employee-friendly" regulations. Thus, regulations that favor hiring and dismissal of employees may increase the relative bargaining ability of employers with respect to employees. The literature has focused on the impact of employee protection laws on employment rather than on wages, leading to scarce evidence regarding the impact of protection laws on employees' earnings (Betcherman, 2012). According to the OECD (2011), there is a negative relationship between employee protection regulation and wage inequality in OECD countries. Other scholars have found similar results (Koeniger, Leonardi, and Nunziata, 2007; Checchi et al., 2008).

Laws and regulations can also affect the extent to which employees' earnings are associated with their productivity. Because a stronger relationship between pay and productivity implies that wages are more closely correlated with workers' output, we should expect a positive relationship between the value captured by workers and the degree to which wages are sensitive to productivity levels. In a setting in which productivity is higher (in dollar terms), for instance, because of an increase in the price of the firm's product, employees should increase their effort levels when their compensation is more closely associated with productivity. Empirical evaluations of the relationship between pay-for-performance schemes and wages have generally concluded that the higher the dependence of wages on employee productivity, the higher the level of output (Lazear, 2000; Prendergast, 1999).

Because the capacity of employees to appropriate value may depend on the characteristics of the labor regulations, it is important to analyze the moderating effects of these regulations on the employees' ability to appropriate value when there is an exogenous shock to the price of the firm product. A positive shock increases the difference between the financial resources available to the firm and the minimum amount the firm will be willing to accept to remain in business, raising the amount of value available for distribution among the firm's stakeholders. In a setting of increasing resource availability, employees' bargaining ability becomes relatively more important, and thus, regulations that are thought to increase this ability, such as more centralized bargaining processes, less flexible hiring and dismissal processes, and more intense pay-for-performance schemes, should become relatively more important in explaining the value appropriated by employees. This discussion leads to hypotheses 2a (H2a), 2b (H2b), and 2c (H2c).

H2a: *The positive effect of an exogenous increase in the firm's product price on the value appropriated by employees is larger when wages are determined by a more centralized bargaining process.*

H2b: *The positive effect of an exogenous increase in the firm's product price on the value appropriated by employees is smaller when labor regulations of hiring and firing are more flexible.*

H2c: *The positive effect of an exogenous increase in the firm's product price on the value appropriated by employees is larger when labor regulations are more favorable to productivity-based pay.*

### 3.2.4 The moderating effects of firm ownership type

Firms can be state owned (SOEs), privately owned (POEs), or a combination of both. SOEs are companies over which the state has significant control through full, majority, or significant minority ownership (OECD, 2015). State ownership is an influential global force in many industrial sectors and countries (Inoue, Lazzarini, and Musacchio, 2013; Tian and Estrin, 2008; Musacchio and Lazzarini, 2014). For example, in the copper mining industry, 24 percent of world production was controlled by states in 2008 (Ericsson and Massey, 2011). Major differences between SOEs and POEs appear to be driven by their different motivations and objectives. CEOs of SOEs are generally focused on creating wealth in the economy and ensuring the well-being and employment of citizens, whereas managers of POEs primarily focus on creating wealth for their shareholders.

There are different views concerning the role of state ownership in firms (Sapienza, 2004). The social view (Atkinson and Stiglitz, 1980) suggests that SOEs are created to address market failures whenever the social benefits of SOEs exceed their costs. According to this view, POEs and SOEs differ because the former maximize profits and the latter maximize broader social objectives. The political view (Shleifer and Vishny, 1994) posits that SOEs are a mechanism for pursuing the goals of politicians, such as maximizing employment or financing favored companies. The agency view shares with the social view the idea that SOEs are created to maximize social welfare but stresses that SOEs can also generate corruption and misallocation because of weak managerial incentives, low effort, and diversion of resources for personal benefit (Banerjee, 1997; Hart, Shleifer, and Vishny, 1997). In the agency view, misallocation of resources occurs because managers shirk or divert resources for private use, whereas in the political view, misallocation of resources is the objective rather than the result of perverse incentives because politicians deliberately transfer resources to their supporters (Shleifer, 1998).

Because of a more diffuse objective function, SOEs may be more sympathetic than POEs to employee demands when there are more resources to be distributed. Larger distributions to employees can facilitate the SOE manager's job by reducing turnover and bargaining for higher effort or simply by keeping the peace with employees (Bertrand and Mullainathan, 2003). POE managers, who report to shareholders and are mainly interested in maximizing profits, may face difficulties distributing any additional resources generated by the company to employees. More broadly, SOE managers might care more than POE managers about improving workplace relations and satisfying employee demands.

The different theories of SOEs cannot be disentangled by examining the effects of state or private ownership on the value appropriated by employees. In other words, it is not clear whether SOEs would allow employees to appropriate more value because they maximize broader social objectives, have incentives to do so, or inefficiently cater to politicians' wishes. However, these theories suggest that state ownership may be an important factor affecting employees' bargaining ability and that the effects of SOEs on this bargaining ability should be increasing in the percentage of the firm's state ownership.

Briefly, if state ownership increases the bargaining ability of employees, there should be an increase in the value appropriated by employees when firms have increasing levels of financial resources. This discussion leads to hypothesis 3 (H3).

H3: *The positive effect of an exogenous increase in the price of the firm's product on the value appropriated by employees increases with the extent of state participation in the mine's ownership.*

### 3.2.5 The moderating effect of ownership concentration

Firms can have diluted or concentrated ownership. Large shareholders (usually associated with higher ownership concentration) have the power to monitor managers and influence them, thus reducing managerial malfeasance and shirking (Shleifer and Vishny, 1986). Given that controlling shareholders, or their representatives, often serve as directors and officers, they can influence management decisions directly, facilitating the alignment of managerial and shareholder interests (Wang and Shailer, 2015). In contrast, in corporations with diluted ownership, individual shareholders have little incentive or power to monitor managers (Grossman and Hart, 1980; Stulz, 2005), allowing managers to pursue their own goals, which can lead to more pronounced principal-agent problems (Jensen and Meckling, 1976).[3]

Because low ownership concentration makes monitoring more difficult, a positive relation between ownership concentration and firm performance is expected. Shleifer and Vishny (1986) show that large shareholders play an important role in increasing firm value and that the share price of the firm increases with the proportion of shares held by large shareholders. Consistent with this view, we hypothesize that more concentrated ownership improves the capacity to monitor managers, increasing the alignment of shareholder and manager interests and reducing the possibility of discretionary resource transfers to employees. Less closely monitored managers have more discretionary power to allocate resources when firms have more funds available. More transfers to employees may increase the management's prestige among employees and provide a more pleasant work environment despite possible reductions in firm profits and share prices.

Briefly, the possibility and power to monitor managers is reduced when ownership concentration is low (Grossman and Hart, 1980). This lower monitoring capacity promotes larger transfers of resources to employees when there are more resources available as long as managers obtain benefits from better-paid employees. This discussion leads to hypothesis 4 (H4).

H4: *The positive effect of an exogenous increase in the price of the firm's product on the value captured by employees is larger when the firm's ownership is less concentrated.*

Our theoretical framework and hypotheses are summarized in Figure 3.1.

[3] Regardless of the benefits of improved monitoring and reduced principal-agent problems between shareholders and managers, concentrated ownership may increase conflicts of interest between controlling and minority shareholders, exacerbating the difficulty of assuring that minority shareholders are not expropriated and that inefficient activities will not be performed (La Porta et al., 1998; Morck, Shleifer, and Vishny, 1988).

## 3.3 Empirical setting

Our empirical setting is the copper mining industry. We have detailed information about the performance and characteristics of the main copper mines in the world. To better understand our setting, it is useful to briefly explain the basic phases of the most common copper production process. The first phase involves mining the ore (the rock that contains the metal). This step is performed in either open pit or underground mines. Whereas the extraction method is the same for all open pit mines, different mining methods can be used in underground mines, and the selection of one over another depends on economic and geological factors. The second phase, milling, involves crushing of the ore to obtain a powder that, after being mixed with chemical reagents, is deposited in flotation cells from which copper concentrate is obtained (copper concentrate is approximately 30 percent copper). The next phases are smelting (to remove impurities) and refining (to obtain other valuable metals if they are present at this stage). At the Our empirical setting is the copper mining industry. We have detailed information about the performance and characteristics of the main copper mines in the world. To better understand our setting, it is useful to briefly explain the basic phases of the most common copper production process. The first phase involves mining the ore (the rock that contains the metal). This step is performed in either open pit or underground mines. Whereas the extraction method is the same for all open pit mines, different mining methods can be used in underground mines, and the selection of one over another depends on economic and geological factors. The second phase, milling, involves crushing of the ore to obtain a powder that, after being mixed with chemical reagents, is deposited in flotation cells from which copper concentrate is obtained (copper concentrate is approximately 30 percent copper). The next phases are smelting (to remove impurities) and refining (to obtain other valuable metals if they are present at this stage). At the end of these processes, 99.95–99.99 percent pure copper is obtained.

Metals such as silver, lead and zinc are regularly found in copper ore. These metals are separated during either the concentration (lead and zinc) or the refining and smelting phases (silver) (Ayres, Ayres, and Rade, 2002). The mining and crushing steps are necessary processes common to all metals found in the ore. Table 3.1 shows the main metals present in the ore treated by copper mines in our sample. On average, copper accounts for 65 percent of net revenues.

The copper mining industry provides a good setting in which to study the factors that affect the value appropriated by employees. First, mines are almost certainly price takers.

For instance, copper is a homogeneous product that is traded in the worldwide market, and no producer controls a large share of annual extraction (Ericsson and Massey, 2011). Thus, unilateral market power appears to be absent, and coordinated oligopoly action seems unlikely.[4] Copper producers recognize that the industry is highly competitive and that financial performance depends heavily on the international price of copper (Corporation, 2009).

Second, during the period covered by our data (2000 to 2008), the prices of copper and other metals produced by copper mines changed dramatically. From a stable copper price of approximately one dollar per pound between 2000 and 2003, the price jumped to 2.5 dollars per pound in 2004, remaining at a price higher than three dollars per pound in 2005 and 2006 and declining thereafter. Thus, the period of nine years considered in our sample includes different subperiods. Additionally, our database includes information on 157 mines located in 30 countries, which provides variation in labor regulations and employee opportunity costs in different geographic regions and over time that is worth analyzing.

Finally, our setting is also interesting because labor in the mining industry has been characterized as active in terms of worker organization and disputes with management (U.S. Congress Office of Technology Assessment, 1988). There is evidence that in the U.S., mine employees receive better compensation than employees with comparable skills in other industries (Itkin, 2007). A higher employee compensation creates pressure on management to reduce labor costs and switch to less labor-intensive technologies (U.S. Congress Office of Technology Assessment, 1988).[5] As Table 3.2 shows, after consumables – which include maintenance parts, diesel, and inputs such as chemical products used in the production process – labor is the second-largest cost for mining companies.[6]

[4]Copper mining companies face several constraints that preclude their ability to modify output in the short term. High operational and startup costs make it difficult to decrease or stop operations for short periods, while the need for large investments to expand capacity makes it difficult to increase production in the short run (Mikesell, 2011).

[5]According to the Occupational Wages around the World database (WDR, 2013), industries related to the extraction of natural resources such as coal mining, petroleum, and other types of mining generally pay above the median wage between 2000 and 2003. Starting in 2004, the rank of these industries improved significantly, and for most of the remaining of the period, the average earnings of employees in these industries increased in both absolute and relative terms.

[6]The decrease in the average number of employees per mine over the years is explained by the entry of smaller mines with fewer employees and not by a decreasing number of employees at the mines that were operating at the beginning of our sample period.

## 3.4 Data and descriptive statistics

Our data come from different sources. Brook Hunt, which is now part of Wood Mackenzie, a leading consulting company in the metal and energy industries, provides detailed financial and operational information for the largest copper mining companies in the world. Information for each mine includes, among other things, the year in which the operation started, estimated year of closure, mining methods used, ores mined, metals present in ore, tons of concentrate produced, revenue, cost breakdowns for mining and milling, depreciation of physical assets, energy and fuel consumption, number of employees, average earnings, and total hours worked. Other sources of data are provided in the variable definitions below.

Before the application of any filter, the database comprised the activities of 164 mines operating in 34 countries from 2000 to 2008. The full sample of mines includes 1,086 mine-year observations. Given that mines can start or stop producing at any point during the period under study, we have an unbalanced panel of firms. All monetary variables are reported annually in US dollars (USD). Hereafter, mines and years are denoted by the letters $i$ and $t$, respectively.

### 3.4.1 Dependent variable

**Average Earnings**. Our dependent variable is the average annual earnings per employee at mine i in year t. This is the actual average compensation received by employees after all social security costs and taxes have been deducted. We choose this measure because it does not include the effects of laws that might affect employees' pre-tax income via changes in payroll deductions that are the responsibility of employers. Our variable includes bonus payments made in year t by mine i to its workers, if any. By examining how average earnings change over time, we can better understand the effects of price variations in the firm's product on the actual value captured by employees. In general, an employee appropriates more value when she receives higher compensation for a similar amount of work.

### 3.4.2 Independent variables

**Price**. This variable reflects the exogenous shock that mining companies experienced during the period under study. Instead of considering one price for all mines, we take into account that ores can contain different compositions of valuable minerals. Thus, we compute a price per ton of ore mined, which is specific to each mine-year, since the composition of minerals found in the ores extracted could change from mine to mine and even from year to year. To compute this variable, we multiply the average quantity of each mineral found per ton of rock mined by its international price, as reported by the Chilean Ministry of Mining and the World Bank. Given that almost 100 percent of the mines in our sample produce some copper, in Figure 3.2, we show the evolution of the average of our measure of price and the international price of copper between 2000 and 2008.[7] As can be seen, the patterns of prices are similar.

**Labor regulations**. Our measures of labor regulations come from the Executive Opinion Survey (EOS), whose results are included in the Global Competitiveness Report (GCR) prepared by the World Economic Forum (WEF). The GCR includes hard data, such as economic indicators and demographic information, along with country scores for each of the dimensions of the EOS. Executives around the globe are contacted through partner institutes responsible for administering the survey in each country following precise guidelines with respect to the sampling process. Respondents evaluate each question on a 7-point Likert scale ranging from 1 to 7, where 1 represents the worst scenario. Data from the GCR and EOS have been used in recent academic research on labor regulations (Belenzon and Tsolmon, 2016; Freeman, Kruse, and Blasi, 2008; Sleuwaegen and Boiardi, 2014). The raw data obtained from the surveys are screened to detect issues with the answers (e.g., surveys that are not at least 50 percent complete are discarded) and to ensure that answers are representative of the overall sample from a country (i.e., detection of outliers using a multivariate test and z-scores).

Consistent with our theoretical developments, we examine variables related to wage decentralization, hiring and firing, and pay and productivity. These variables have been considered in other cross-national studies of labor practices (Freeman, Kruse, and Blasi, 2008). Table 3.3 provides the actual wording of each question.

**Wage decentralization**: The wage bargaining process can be centralized (e.g., at the industry level) or conducted independently by each firm. A high value of wage decentralization implies a more decentralized negotiation process.[8] **Flexibility of hiring and firing**: This variable measures the flexibility of companies to hire and to fire workers, where a higher value implies more flexibility. **Pay and productivity**: A higher value of this variable im-

[7]There is only one mine in our sample that does not produce copper between 2000 and 2008.

[8]This variable is called wage flexibility in the EOS.

plies that wages are more closely correlated with worker productivity.

Since the questions from the EOS are rated on a 1 to 7 scale, we use the raw scores for each question as our measures of labor regulation. Table 3.4 shows the descriptive statistics of the variation in labor regulations between and within countries. For example, the average wage decentralization for the full sample is 5.243, with a minimum value of 2.1 and a maximum value of 6.206. Additionally, the minimum value and maximum value of the average mine are 2.673 and 6.085, respectively. The "within" row shows the minimum and maximum deviation of a mine from its average. For instance, the largest negative deviation of a mine from its own average is -0.944, whereas the largest positive deviation is 0.903. The standard deviation of this variable (the "within" standard deviation) is 0.205, whereas the standard deviation of the average value per mine (the "between" standard deviation) is 0.725.[9]

**State ownership**. To measure state-ownership, we use a continuous variable ranging from 0 to 1, where a higher value indicates more state ownership (a value of zero implies no state ownership – i.e., the mine is completely private owned – while a value of one indicates that the state is the sole owner of the mine). To determine the percentage of state ownership, we reviewed the name and percent ownership of each shareholder for each year in our database. To provide a more complete measure of state ownership, we computed the total participation of a state in a mine by adding its direct and indirect participation. By indirect participation, we mean the percentage of state ownership through another firm. For example, let us consider the Chambishi mine in 2008. Chambishi is located in Zambia, and it had two owners at that time: China Non-Ferrous Metals (85 percent) and ZCCM (15 percent). In 2008, the Zambian government owned 77.7 percent of ZCCM. Thus, state participation in Chambishi in 2008 was recorded as 11.7 percent ($0.85 \times 0.77$).

We do not incorporate the state-ownership variable as a stand-alone regressor because of its almost nonexistent variance within mines. Including the main effect of the state-ownership variable would capture specific mine characteristics for which there is variation in the percentage of state ownership, making its interpretation specific to those mines rather than providing a generalizable estimate of the effect would be in the population. Thus, we incorporate the state-ownership variable as an interaction with the price to measure how the value appropriated by employees is affected by changes in the price of metals. Table 3.5

[9]In all cases, the between standard deviation is greater than the within standard deviation, meaning that there is more variability across mines than within mines. Because we estimate our models using mine fixed effects, we only exploit the within variation to identify our parameters.

presents some descriptive statistics relative to our state-ownership variable.

**Ownership concentration**. To compute this variable we analyzed the complete ownership structure of each mine in each year and computed a Herfindahl-Hirschman index (HHI) of ownership concentration, which equals the sum of the squares of the individual ownership percentages. We consider a continuous variable in which a higher HHI value implies higher ownership concentration. Table 3.6 presents more detailed descriptive statistics for this variable.

**Controls**. To capture the potential effects of extended work hours (overtime) on average annual earnings per employee, we control for the average number of hours worked per employee. Since mines can also change the number of employees hired in response to variations in price, we include the total number of employees per mine. Additionally, we control for the presence of subcontractors in mine i's operations using a dummy variable to catch any effect on average wages.

We also control for the average amount of ore treated per employee each year to better isolate the effect of price variations on employee earnings from that of productivity changes (Berchicci, Dowell, and King, 2012). Mine age, computed as the number of years since the mine started its most current operations, is included as a control given the observed relationships between firm age and performance, innovation, and business life cycle (Loderer and Waelchli, 2010; Phene, Fladmoe-Lindquist, and Marsh, 2006; Shan, Fu, and Zheng, 2016). Since mines differ greatly in the amount of capital they have available for operations, we also control for the total depreciation cost per worker (Koch and McGrath, 1996).[10]

Because all values reported in USD and employee earnings are usually expressed in local currency, part of the observed change in our dependent variable might reflect exchange rate variation. Additionally, in some countries, earnings are tied to an inflation index to preserve purchasing power in presence of widespread price increases. To control for exchange and inflation rate variation, we include the natural logarithm of the exchange rate (in terms of the quantity of local currency needed to buy one USD) and the natural logarithm of the

[10]The depreciation value includes depreciation costs of equipment related to the mining and milling stages as well as depreciation of the smelting and refining phases if the mine is vertically integrated (these are the next two steps in the process of obtaining refined copper). Unfortunately, this is the best proxy we could find for the capital intensity of the mine. Since our models also consider mine fixed effects to control for time-invariant characteristics of the mines, we believe that characteristics such the vertical integration of refined copper production will be captured by these fixed effects, which partially alleviates our concerns about this control variable.

inflation rate (measured as a percentage).[11]

To control for the evolution of a country's gross domestic product (GDP) of the natural resources sector, we include GDP per capita in current prices multiplied by the percentage of rents explained by natural resource activities in each country as reported by the World Bank. Although this measure will not control for the actual GDP per capita of the natural resources sector (since we do not have the aggregate number of workers employed in each country), we expect this variable to help us capture – to some extent – the evolution of the true GDP per capita of the natural resources sector in the countries included in our sample.

As actual measures of opportunity costs are hard to find, to control for workers' opportunity costs, we use the unemployment rate in year t for the country in which mine i is located as a proxy for variation in employee opportunity costs. A higher unemployment rate is understood as a lower opportunity cost for a hired worker, decreasing the attractiveness of leaving the mining company (Burdett and Mortensen, 1998; Shapiro and Stiglitz, 1984). A lower unemployment rate implies an increase in the employees' opportunity cost because it is less difficult to find a job elsewhere. The data on unemployment rates come from the World Bank.

Year fixed effects, although common in econometric models with panel data, are not included in our models. Although our price variable is not a perfect linear combination of year fixed effects, much of the overall trend in mineral prices is captured by year fixed effects if they are included in our models. Therefore, the inclusion of year fixed effects would hinder us from analyzing the impact of price variation on the value captured by employees, which is a major goal of our study.

Even though we have included a reasonable number of mine-level controls, differences in available resources, quality of management, and employee human capital are examples of controls not included in our models that are time-invariant characteristics of the mines. To control for these potentially important variables, we use mine fixed effects in all of our specifications. The inclusion of mine fixed effects prevents us from introducing other time-invariant variable into our models, especially dichotomous variables that represent characteristics that do not vary over time, such as country fixed effects. Table 3.7 summarizes the variables considered in our study and their sources.

[11] Given that a logarithmic transformation cannot take a negative argument and to avoid losing observations in countries that experienced negative inflation rates, we add a base value of two to the reported inflation rates before computing the logarithmic transformation. In our database, five observations have reported inflation rates of less than zero.

Table 3.8 presents summary descriptive statistics for the observations included in the final sample. For an observation to be included, we require complete information for all the variables in the regression model. As a result, our final sample includes 983 mine-year observations. Table **??** shows the pairwise correlation matrix of the variables considered in this study.

The composition of mines and countries considered in our final sample is shown in Table 3.9. Approximately 33 percent of the countries have only one mine operating during the study period (when we consider countries with one or two mines, this percentage increases to 50 percent). This is a reflection of the geographical distribution of copper producers rather than negative or biased coverage of our data (*Copper's top 10 Countries and Companies*).

## 3.5 Econometric model

We consider two equations to test our hypotheses about the relationships between the value appropriated per employee and our selected explanatory variables.

$$\begin{aligned} Average\,Earnings_{it} &= \alpha + \beta Rock\,Price_{it} + \lambda Concentration_{it} \\ &\quad + \sum_{l=1}^{3} \omega_l LR_{lit} + \sum_{j=1}^{10} \theta_j Control_{jit} + u_i + \epsilon_{it}. \end{aligned} \tag{3.1}$$

In Equation 3.1, $\beta$ represents the average impact of the price on earnings per employee. The $\lambda$ term summarizes the effect of our ownership concentration variable on average earnings. The coefficients $\omega_l$ capture the impact of the variables that measure labor regulation characteristics ($LR_{lit}$). The effects on our dependent variable of our group of control variables are captured by the $\theta_j$ parameters. As mentioned in the previous section, mine fixed effects ($u_i$) are included in our model to control for time-invariant characteristics that might affect the value appropriated by employees. The idiosyncratic error term in our model is denoted by $\epsilon_{it}$.

$$
\begin{aligned}
Average\,Earnings_{it} \quad = \quad & \alpha + \beta_1 Rock\,Price_{it} + \lambda_1 Concentration_{it} \\
& + \beta_2 (Rock\,Price_{it} \times State\,Participation_{it}) \\
& + \lambda_2 (Rock\,Price_{it} \times Concentration_{it}) \\
& + \sum_{l=1}^{3} \omega_l LR_{lit} + \sum_{l=1}^{3} \rho_l (Rock\,Price_{it} \times LR_{lit}) \\
& + \sum_{j=1}^{10} \theta_j Control_{jit} + u_i + \epsilon_{it} \qquad (3.2)
\end{aligned}
$$

Equation 3.2 expands 3.1 by incorporating interaction terms between price and our variables that measure ownership concentration ($\lambda_2$), state ownership ($\beta_2$), and labor regulations ($\rho_l$).

## 3.6 Results

To make our estimates easier to interpret, we mean center all of our variables of interest by taking the average and subtracting it from the original value. This re-parameterization does not affect the estimates of the interaction effects, but the coefficients and interpretation of the main effects of the variables that interact with our price variable do change, since they now reflect the effect of each variable when the price takes on its mean value (Jaccard, Turrisi, and Wan, 1990). This centering occurred after the variables were log-transformed, with exception of the presence of contractors, variables related to labor regulations, and our state-ownership regressor, which are included in our models without log-transformation.

### 3.6.1 Equation 1

We can appreciate the positive impact of price on average earnings per employee. Since all of our variables of interest —with the exception of those related to labor conditions— have been log-transformed, we can interpret the estimated coefficients as elasticities. In the case of price, a 100-percent increase in the average price is associated with an increment in average earnings of approximately 15.5 percent (column (1)).[12] This result is highly significant, confirming hypothesis 1. We also found that, in general, more flexibility (in terms of hiring and firing policies and the sensitivity of employee earnings to productivity) is, on average, positively related to employee earnings in the mining industry and that ownership concentration does not affect employee compensation.

With respect to our control variables, an important result is associated with the opportunity cost of labor. We found that a 100-percent increase in the unemployment rate is associated, on average, with a reduction in average earnings of approximately 19 percent.

### 3.6.2 Equation 2

In the full model reported in column (2) of Table 3.10, the interactions between labor regulations and price are statistically significant. The negative interaction between wage decentralization and price (p-value less than 0.001) indicates that companies that negotiate directly with their employees (instead of relying on a centralized bargaining process) are able to reduce the positive impact of an increase in the price on employee earnings. On the other hand, the positive and significant coefficients of the interactions of (i) price and pay and productivity and (ii) price and hiring and firing (p-values of 0.03 and 0.026, respectively) show that the more closely wages are tied to productivity and the more flexible are the regulations associated with hiring and firing, the higher the positive impact of changes in price on employee earnings. These results provide support for hypotheses 2a and 2c but not for hypothesis 2b.

Figure 3.3 presents three graphical depictions of the marginal effect of price on yearly earnings as a function of the actual observed values of wage decentralization, hiring and firing, and pay and productivity.[13] For example, keeping everything else at its mean value, when wage decentralization takes on its lowest value (i.e., centralized negotiations), a 100-percent increase in price is, on average, associated with an increase of almost USD 7,500 in employee earnings, whereas when individual mines are able to negotiate with their employees directly, the same change in price raises earnings by approximately USD 800. When regulations regarding the hiring and firing of employees take on the highest value observed in our sample (i.e., more flexibility), a 100-percent increase in price is, on average, associated with an

[12] This 15.5 percent comes from the following expression: $exp0.144 - 1$.

[13] Marginal effects were computed by fixing all other variables at their full-sample mean values.

increase in employee earnings of almost USD 4,500, whereas the effect of the same increase in price on employee earnings is approximately USD 700 for low values of hiring and firing flexibility. When employee wages depend greatly on productivity, a 100-percent increase in the price raises yearly wages by approximately USD 4,000. However, for low values of pay and productivity, i.e., when wages are mostly fixed, the impact of price on earnings is statistically indistinguishable from zero at the 95 percent confidence level. The graphs in Figure 3.3 were computed assuming that all other variables are fixed at their sample means. However, extreme values of wage decentralization, low flexibility of hiring and firing, and a low correlation between pay and productivity are related to cases in which workers capture little to none of the extra value generated by firms when the price of the products they sell are affected by a positive exogenous shock.

The estimated effect of the interaction between price and state ownership is positive and statistically significant (p-value less than 0.001) in our full model (column (2)). Using this last estimate, when the government is the only owner of a mine, the impact of a 100-percent increase in the price of minerals extracted by that mine on employee earnings increases by approximately 10 percentage points compared to the case of no government ownership. This result is consistent with different CEO motivations and shows that employee earnings increase more in state-owned companies than in fully privately owned companies following an improvement in the firm's financial prospects, providing support for hypothesis 3. We did not find any statistical effect of ownership concentration on employee compensation; thus, we do not find support for hypothesis 4.

In columns (3) to (6), we present the estimates of the models in Equations 3.1 and 3.2 using the estimated average earnings for production employees (those engaged in the mining and milling steps of the production process) and general and administrative (G&A) or clerical employees.[14] In general, the results are consistent for both types of workers. However, one interesting difference can be appreciated in terms of the main effect of price in columns (4) and (6) of Table 3.10. For production workers, the main effect of price when the interacted variables take on their mean values is positive and highly significant (p-value of 0.018), whereas this coefficient is indistinguishable from zero in statistical terms when the earnings of G&A workers are used instead. While productive workers enjoy, on average, an increase of 20

[14]The earnings for these two types of workers are not reported directly in our database, so we had to estimate them using information on labor costs per function, productivity, and hours worked. For nine observations, we could not estimate the earnings, so we opted to exclude those observations from these analyses.

percent in earnings if they live in a country with "average" levels of ownership concentration, state ownership, and labor regulations, G&A workers only experience pass-through of 3.2 percent. This result reflects that different types of workers are affected differently by the same shock to the firm's product prices.

### 3.6.3 Robustness checks

To ensure that mines that started or stopped operations at some point between 2000 and 2008 are not driving our results, we divide our sample into two main groups: mines that operated over the whole period (nine years) and those that did not. This latter group includes the following mines: (i) operating before 2000 and stopped before 2008; (ii) started operations after 2000; or (iii) stopped operations after 2000 but resumed then before 2008. Our results show that the main conclusions drawn from the full database do not change when the subsets of mines with fewer than nine years of data are excluded (see column (7) of Table 3.10).[15] Additionally, we check whether countries with one or two mines could be driving our results. We re-estimate the full model considering only countries with three or more mines operating between 2000 and 2008. The results obtained for this subsample —presented in column (8) of Table 3.10— are in line with those presented in column (2).

To check the robustness of our results to the relative presence of copper, we re-estimate our models considering only mines for which copper represent over 50 percent of revenues (column (9)). In this case, we use the price of copper as an independent variable and control for the total percentage of copper of mine revenues. The results remain quite similar to the estimates in column (2).

Up to this point, we have implicitly assumed that wages are determined in the same year that institutional changes occur. However, it is possible that variations in labor regulations do not have contemporaneous effects on earnings but lagged effects (i.e., the full impact of these changes are observed in future rather than in current average earnings ). Column (10) of Table 3.10 presents the results of our full model (Equation 3.2) considering lagged values of our labor regulation variables. The estimates obtained from these specifications are in line

[15]The total number of observations is not a multiple of nine since an observation included in our sample was required to have complete information for all the variables in our models. Therefore, although one mine could have operated during the nine years covered by our database, if in one of the years has missing information for at least one of the variables we consider in our regressions, that observation is automatically discarded.

with those presented in column (2) in terms of both the point estimates and their statistical significance.

In column (11), we re-estimate the full model (Equation 3.2) using the average cost per worker as the dependent variable, with the percentage of on-costs of the total cost as an additional control.[16] The main results and statistical significance remain similar to those of the main specification.[17]

We also analyze whether factors other than those studied in this article might be affecting location decisions, such as natural and geological characteristics. For example, some deposits might be so convenient that their exploitation can be undertaken without other major considerations relative to – for example – labor regulations and/or ownership structure. To check for this possibility, we took the top 50 percent of mines in terms of the value of ore extracted at the start of the study period and re-estimated our main specification using this subsample. The results obtained are in line with those presented in Table 3.10.[18]

## 3.7 Discussion

The division of value between economic actors is a key issue in the analysis of heterogeneity in firm performance. Economic actors usually seek to appropriate higher shares of the value generated in an exchange; for example, employees want higher wages, suppliers try to obtain higher prices for their inputs, and shareholders seek higher stock prices or profits. Although the strategic management field has been concerned with the appropriation of value in organizations since its emergence, only recently have scholars started considering stakeholders other than shareholders in their analyses (**Greenan2014**; Bennett, 2013). This consideration has led to a need for broader empirical analysis of who appropriates the value generated

[16] This variable includes the average earnings and bonuses received by employees plus on-costs, such as insurance contributions and other social security expenses (that are the responsibility of the employer).

[17] In our baseline scenario, we use a continuous variable ranging from 0 to 1 to measure state ownership. However, and consistent with the body of literature (Inoue, Lazzarini, and Musacchio, 2013; Musacchio and Lazzarini, 2014; Tian and Estrin, 2008) that argues that the participation of the state in managerial decisions matters only when the state has at least some minimum percentage of ownership, we re-estimate our models considering that state ownership only when the participation of the state is greater or equal to 10 percent and 20 percent, respectively. The results obtained by using these different thresholds remain the same. These estimates are available from the authors upon request.

[18] These results are available from the authors upon request.

by a firm and the determinants of that value.

In this article, we focus on the value appropriated by employees and empirically demonstrate that institutional variation impacts the value captured when there is an exogenous change in the price of the firm's product. We focus on the impact of price on average compensation instead of other options, such as the labor percentage of total costs, because a dependent variable in the form of average salary allows us to analyze how an exogenous shock in prices affects the value captured by employees independently of other costs and of which stakeholders benefited (or suffered) from the change in price. In industries in which employees are important stakeholders (such as mining), understanding the determinants of worker compensation is a necessity for both managers and shareholders.

Our results show that a centralized bargaining process increases the bargaining power of employees and that this increase in power becomes more important when the size of the pie to be distributed increases. The positive effect of more flexibility of hiring and firing on the value appropriated by employees when there is a positive shock to revenues was not consistent with our predictions and shows that with more flexibility, firms have to pay more to retain their employees and to hire from outside when firm prospects improve. Our findings also suggest that when the price of a firm's product increases, workers are able to appropriate more of the value generated by the firm.

Consistent with different CEO motivations in SOEs and POEs, and thus contributing to our understanding of the effects of ownership on value appropriation, we find that the value appropriated by employees after an increase in the exogenous price of the firm's product is higher in companies with state ownership. However, our results show that concentration of ownership does not influence the change in value appropriated by employees, possibly indicating that more monitoring from large shareholders does not change employee compensation or that the benefits obtained by managers from better paid employees do not differ from those obtained by shareholders. Overall, we conclude that depending on regulations and ownership structures, and although they may lack formal residual-claimant rights (Coff, 1999; Fama and Jensen, 1983), employees can still capture a relevant share of the additional value obtained by firms when companies experience positive shocks to the prices of the products they sell.

Our findings, which are robust to different measures of variables and econometric specifications, suggest that employee compensation is not the only answer to changes in value generated by firms. Another response to increasing output prices is the opening of new mines. This is coherent with the conjecture that there is more entry when prices increase. Higher prices support the entry of mines for which there are higher extraction costs and the extraction of minerals from higher cost deposits in current mines, affecting the marginal product of labor. Other firm responses to higher output prices include adding capital and extracting larger quantities from existing veins. Our econometric modeling controls for all of these factors.

For managers and investors, our work highlights the importance of institutional and ownership variables in the determination of the distribution of value between a firm and its employees. As firms move across countries and have different ownership arrangements and partners, we need to consider how the institutional characteristics of countries might affect value appropriation by different stakeholders. Choices such as the locations and types of companies in which to invest should consider how the allocation of value among stakeholders changes following shocks that affect the value captured by a firm as a whole. For instance, shareholders of firms with positive future expectations about the prices of their products may prefer to enter markets in which salary negotiations are not centralized or where partnership with the local government is not obligatory.

Several avenues for further research stem from this study. First, we only analyze the effect of an exogenous shock to the price of the firm's product on the value appropriated by employees. It would be interesting to analyze the effects of similar shocks to the value captured by other stakeholders. Second, a more detailed analysis of the main institutional determinants of value generation and appropriation is needed. Because more complete sets of indicators of the quality of labor regulations are being developed, researchers should soon be able to provide more comprehensive explanations of the ways institutions shape bargaining power and influence the value captured by each party. Finally, the analysis of the effect of ownership type on the value appropriated by different stakeholders deserves further attention. Ownership differs not only in terms of concentration and in private versus state ownership but also in whether firms belong to a business group, for example.

Our work has several limitations. First, we analyze a specific industry. It would be interesting to compare our results with those of other industries with exogenously driven variation in prices or other relevant variables. As mentioned previously, given that employees in the mining industry have been characterized as active in terms of disputes with management (U.S. Congress Office of Technology Assessment, 1988), we should be cautious when generalizing our results to employees in other industries. Second, our dataset covers a relatively short period (2000 to 2008). Although an important structural change in copper prices occurred during that period, it would be informative to analyze value appropriation by different stakeholders over a longer period. Third, it would be interesting to determine whether the conclusions of our analysis can be extrapolated to other exogenous shocks that might affect the firm's value appropriation process, such as shifts in the prices of inputs driven by technological change. Fourth, despite controlling for a vast set of variables, it is not possible to completely disregard the possibility that the results can be explained by other factors that change contemporaneously with prices.

In sum, we hope that our study provides relevant results and motivates the development of a more comprehensive theory of value appropriation, encouraging future investigations into the sources of variation in bargaining power among firms and suppliers. This endeavor is especially relevant given that the distribution of economic value is a fundamental concern in the history of economic thought (Asher, Mahoney, and Mahoney, 2005) and that empirical work in this field is still in its infancy (Gans and Ryall, 2017).

Table 3.1: Metals in ore.

| **Metal** | **Occurrence** |
|---|---|
| Copper (Cu) | 99% |
| Silver (Ag) | 80% |
| Gold (Au) | 70% |
| Zinc (Zn) | 33% |
| Lead (Pb) | 17% |
| Molybdenum (Mo) | 14% |
| Cobalt (Co) | 11% |
| Nickel (Ni) | 6% |
| N = 983. Mines = 157. | |

Table 3.2: Average percentage of each cost component of the total cost to concentrate per ton of ore.

| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|---|---|---|---|---|---|---|---|---|---|
| Labor | 31 | 30 | 30 | 30 | 29 | 29 | 28 | 27 | 25 |
| Energy | 17 | 17 | 17 | 17 | 16 | 16 | 16 | 15 | 16 |
| Consumables | 33 | 33 | 33 | 33 | 33 | 33 | 34 | 34 | 34 |
| Others | 19 | 19 | 19 | 19 | 20 | 20 | 21 | 23 | 24 |

Percentages are rounded to the next whole number; thus, the total for each year might not add up to 100.

Table 3.3: Labor regulation variables.

| Variable | Definition in GCR 2008-2009 |
|---|---|
| Wage flexibility | *"In your country, wages are (1 = set by a centralized bargaining process, 7 = up to each company)"* |
| Hiring and firing | *"The hiring and firing of workers is (1 = impeded by regulations, 7 = flexibly determined by employers)"* |
| Pay and productivity | *"In your country, pay is (1 = not related to worker productivity, 7 = strongly related to worker productivity)"* |

Table 3.4: Labor regulation variables: descriptive statistics.

| **Variable** | | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|
| (LR) Wage Flexibility | Overall | 5.243 | 0.800 | 2.100 | 6.206 |
| | Between | | 0.725 | 2.673 | 6.085 |
| | Within | | 0.205 | -0.944 | 0.903 |
| (LR) Hiring Firing | Overall | 3.886 | 0.872 | 1.900 | 5.897 |
| | Between | | 0.823 | 2.250 | 5.454 |
| | Within | | 0.369 | -0.917 | 1.722 |
| (LR) Pay and Productivity | Overall | 4.387 | 0.643 | 2.103 | 5.900 |
| | Between | | 0.626 | 2.809 | 5.447 |
| | Within | | 0.251 | -1.085 | 1.330 |
| N = 983. Mines = 157. | | | | | |

Table 3.5: State ownership variable: descriptive statistics.

| **Variable** | | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|
| State Ownership | Overall | 0.100 | 0.262 | 0.000 | 1.000 |
| | Between | | 0.236 | 0.000 | 1.000 |
| | Within | | 0.037 | -0.622 | 1.000 |
| N = 983. Mines = 157. | | | | | |

Table 3.6: Ownership concentration: descriptive statistics.

| **Variable** | | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|
| Herfindahl-Hirschman Index | Overall | 8,124 | 2,515 | 2.100 | 10,000 |
| | Between | | 2,228 | 2.673 | 10,000 |
| | Within | | 1,161 | -4,444 | 5,884 |
| N = 983. Mines = 157. | | | | | |

Table 3.7: Variables and sources.

| Variable | Summary | Source |
|---|---|---|
| Average Earnings | Average annual earnings per worker (including bonus payments) | Brook Hunt |
| Price | International prices of minerals times presence of minerals per ton of ore mined | Cochilco (Chile's Ministry of Mining) / World Bank – Brook Hunt |
| Unemployment | Country's unemployment rate | World Bank |
| Labor Regulations | Wage decentralization (wage flexibility), flexibility of hiring and firing, and pay and productivity | Global Competitiveness Report (World Economic Forum) |
| State Ownership | Sum of direct and indirect participation of a state in a mine's ownership | Brook Hunt / Own computations based on financial statements and public sources |
| Ownership Concentration | Herfindhal-Hirschman index of concentration | Own computations (see above) |
| Controls | Number of hours worked, ore treated, and depreciation cost per worker, number of employees, presence of contractors, on-costs (for models that consider total cost per employee), and mine age | Brook Hunt |
| Controls | Inflation rate; average exchange rate, GDP per capita (current prices); rents of natural resources sector as percentage of GDP | World Bank |

Table 3.8: Summary descriptive statistics.

| Variable | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Average Earnings | 983 | 27401.5 | 21940.99 | 246 | 120300.05 |
| Average Hours Worked | 983 | 1994.23 | 386.08 | 60 | 2864 |
| Number of Employees | 983 | 1885.97 | 4562.45 | 18 | 48750 |
| Ore Treated per Worker | 983 | 10098.61 | 12867.16 | 59.91 | 63832.84 |
| Depreciation per Worker | 983 | 37818.61 | 46557.2 | 190.24 | 437021.91 |
| GDP (Nat. Res.) | 983 | 1140.31 | 1338.85 | 6.75 | 6383.66 |
| Mine's Age | 983 | 30.98 | 30.54 | 0 | 116 |
| Exchange Rate | 983 | 6887.47 | 92582.32 | 0.68 | 1507226 |
| Inflation | 983 | 5.1 | 5.19 | -1.11 | 54.87 |
| Contractors | 983 | 0.14 | 0.35 | 0 | 1 |
| On-costs | 983 | 40.82 | 12.6 | 8 | 133 |
| Unemployment | 983 | 8.14 | 5.03 | 2.3 | 37.6 |
| Rock Price | 983 | 121.11 | 146.62 | 6.59 | 1654.68 |
| Herfindahl Index | 983 | 8123.51 | 2516.28 | 2096.43 | 10000 |
| State Participation | 983 | 0.1 | 0.26 | 0 | 1 |
| LR: Wage Flexibility (GCI) | 983 | 5.24 | 0.8 | 2.1 | 6.21 |
| LR: Hiring and Firing (GCI) | 983 | 3.89 | 0.87 | 1.9 | 5.9 |
| LR: Pay and Productivity (GCI) | 983 | 4.39 | 0.64 | 2.1 | 5.9 |

Table 3.9: Countries and mines.

| Country | Mines | Obs. | Country | Mines | Obs. |
|---|---|---|---|---|---|
| Argentina | 1 | 9 | Mongolia | 1 | 4 |
| Australia | 22 | 147 | Morocco | 1 | 7 |
| Botswana | 2 | 14 | Namibia | 2 | 14 |
| Brazil | 3 | 16 | Peru | 10 | 67 |
| Bulgaria | 2 | 18 | Philippines | 5 | 19 |
| Canada | 26 | 159 | Poland | 1 | 9 |
| Chile | 16 | 123 | Portugal | 1 | 9 |
| China | 4 | 29 | Russia | 4 | 31 |
| Congo D.R. | 7 | 13 | South Africa | 4 | 27 |
| Finland | 1 | 9 | Spain | 2 | 2 |
| India | 1 | 9 | Sweden | 4 | 34 |
| Indonesia | 2 | 18 | Turkey | 1 | 9 |
| Kazakhstan | 7 | 23 | USA | 12 | 83 |
| Mauritania | 1 | 3 | Vietnam | 1 | 3 |
| Mexico | 5 | 35 | Zambia | 8 | 40 |

Table 3.10: Results

| Variables | (1) Base | (2) Base | (3) Prod | (4) Prod | (5) G&A | (6) G&A | (7) 9 years | (8) 3 mines | (9) Cu 50% | (10) Lagged LR | (11) Total cost |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unemployment | -0.212 | -0.205 | -0.189 | -0.189 | -0.296 | -0.290 | -0.193 | -0.153 | -0.157 | -0.245 | -0.208 |
| | (-3.07) | (-3.41) | (-2.61) | (-2.85) | (-2.79) | (-3.16) | (-3.01) | (-1.75) | (-2.07) | (-4.69) | (-3.45) |
| Price | 0.144 | 0.172 | 0.150 | 0.196 | 0.129 | 0.032 | 0.120 | 0.015 | 0.248 | 0.103 | 0.174 |
| | (6.51) | (2.18) | (6.56) | (2.39) | (4.07) | (0.25) | (1.16) | (0.18) | (2.46) | (4.26) | (2.19) |
| Herfindhal-Hirschman index | -0.001 | 0.019 | 0.015 | 0.034 | -0.032 | -0.010 | 0.012 | -0.036 | 0.009 | 0.05 | 0.017 |
| | (-0.02) | (0.35) | (0.23) | (0.58) | (-0.55) | (-0.20) | (0.19) | (-0.67) | (0.16) | (1.05) | (0.33) |
| HHI x price | | -0.012 | | -0.013 | | 0.016 | -0.003 | -0.003 | -0.031 | -0.016 | -0.013 |
| | | (-0.39) | | (-0.43) | | (0.38) | (-0.08) | (-0.09) | (-0.97) | (-0.53) | (-0.43) |
| State participation x price | | 0.100 | | 0.092 | | 0.098 | 0.097 | 0.089 | 0.111 | 0.132 | 0.099 |
| | | (4.89) | | (3.86) | | (4.04) | (4.48) | (5.07) | (2.00) | (5.87) | (4.82) |
| Wage decentralization | -0.027 | -0.04 | -0.028 | -0.043 | -0.014 | -0.012 | -0.054 | -0.074 | -0.067 | -0.031 | -0.038 |
| | (-0.77) | (-1.14) | (-0.77) | (-1.12) | (-0.38) | (-0.37) | (-1.24) | (-1.87) | (-1.23) | (-0.88) | (-1.06) |
| Wage decentralization x price | | -0.076 | | -0.080 | | -0.066 | -0.078 | -0.050 | -0.115 | -0.050 | -0.077 |
| | | (-4.39) | | (-4.08) | | (-3.18) | (-4.01) | (-2.44) | (-3.69) | (-3.63) | (-4.37) |
| Hiring and firing | 0.060 | 0.077 | 0.054 | 0.072 | 0.051 | 0.053 | 0.096 | 0.055 | 0.103 | 0.120 | 0.077 |
| | (3.68) | (3.74) | (2.72) | (2.85) | (2.60) | (2.3) | (3.93) | (2.99) | (3.31) | (5.97) | (3.75) |
| Hiring and firing x price | | 0.039 | | 0.043 | | 0.018 | 0.055 | 0.035 | 0.052 | 0.054 | 0.04 |
| | | (2.25) | | (2.18) | | (0.92) | (2.65) | (1.87) | (2.09) | (3.21) | (2.29) |
| Pay and productivity | 0.101 | 0.103 | 0.104 | 0.105 | 0.051 | 0.052 | 0.126 | 0.112 | 0.134 | 0.104 | 0.103 |
| | (3.31) | (3.31) | (2.74) | (2.67) | (1.13) | (1.26) | (3.08) | (4.27) | (4.20) | (3.77) | (3.30) |
| Pay and productivity x price | | 0.042 | | 0.039 | | 0.075 | 0.044 | 0.039 | 0.052 | 0.021 | 0.042 |
| | | (2.19) | | (2.06) | | (3.30) | (1.65) | (2.13) | (2.39) | (1.27) | (2.19) |
| Constant | 1.590 | 1.518 | 1.856 | 1.478 | 2.059 | 1.712 | 2.320 | 3.118 | 0.952 | 1.510 | 0.509 |
| | (4.88) | (4.93) | (3.55) | (3.00) | (3.04) | (2.71) | (2.19) | (8.43) | (2.26) | (3.37) | (1.03) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mine Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 983 | 983 | 974 | 974 | 974 | 974 | 703 | 846 | 680 | 821 | 983 |
| R-squared | 0.884 | 0.894 | 0.862 | 0.871 | 0.822 | 0.834 | 0.796 | 0.921 | 0.904 | 0.832 | 0.895 |
| Number of mines | 157 | 157 | 156 | 156 | 156 | 156 | 83 | 137 | 118 | 148 | 157 |

Robust t-statistics in parentheses.

Figure 3.1: Summary of theoretical framework and hypotheses.

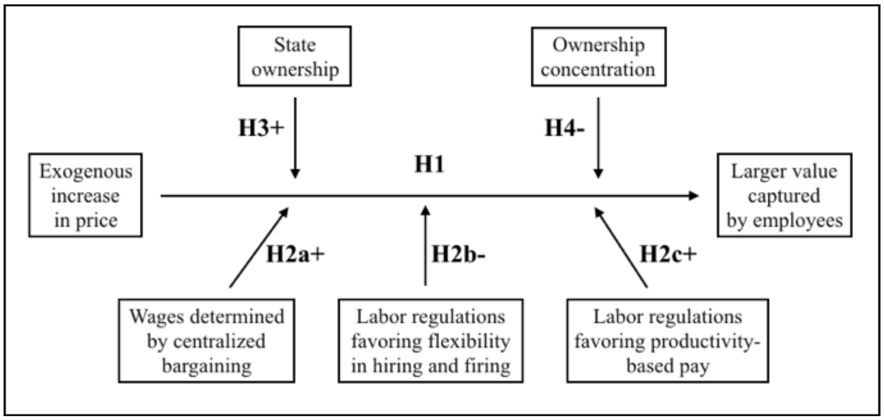



Figure 3.2: Price variable and international price of copper between 2000 and 2008.

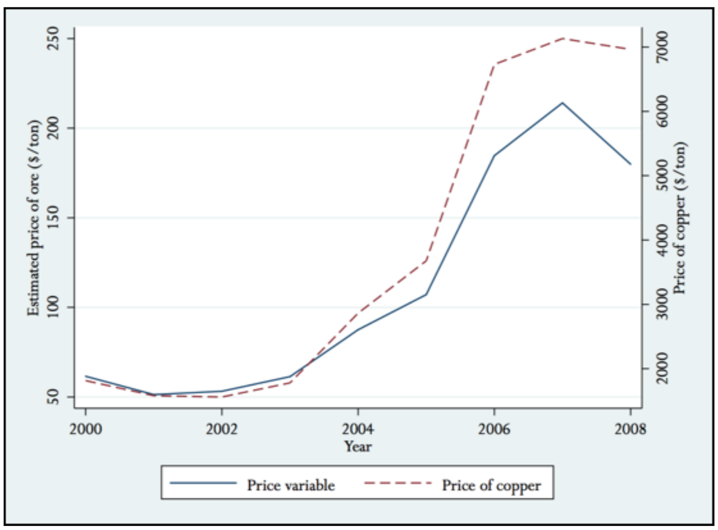
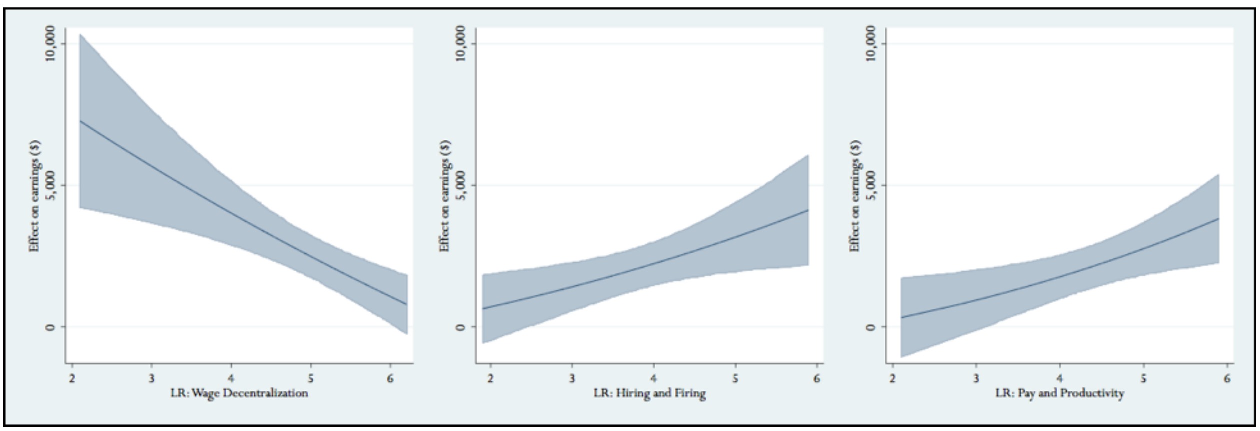


Figure 3.3: Marginal effects of price on earnings (95 percent confidence intervals).



# Bibliography


Abramovich, Samuel, Christian Schunn, and Ross Mitsuo Higashi (2013). “Are badges useful in education?: It depends upon the type of badge and expertise of learner”. In: *Educational Technology Research and Development* 61.2, pp. 217–232. ISSN: 10421629. DOI: 10.1007/s11423-013-9289-2.

Amit, Raphael and Christoph Zott (2001). “Value creation in e-business”. In: *Strategic Management Journal* 22, pp. 493–520. ISSN: 09647058. DOI: 10.1002/smj.187.

Anderson, Ashton et al. (2013). “Steering user behavior with badges”. In: *WWW’13 Proceedings of the 22nd international conference on World Wide Web*, pp. 95–105. ISBN: 9781450320351.

Angrist, Joshua D. and J. S. Pischke (2008). *Mostly harmless econometrics: An empiricist’s companion*. Princeton, NJ: Princeton University Press.

Antin, Judd and Elizabeth F. Churchill (2011). “Badges in social media : A social psychological perspective”. In: *CHI 2011 Gamification Workshop Proceedings*. Vancouver, BC, Canada, pp. 1–4. ISBN: 9781450302685.

Aouad, Marion, Timothy Brown, and Christopher Whaley (2017). *Reference pricing: The case of screening colonoscopies*.

Asher, Cheryl Carleton, James M. Mahoney, and Joseph T. Mahoney (2005). “Towards a property rights foundation for a stakeholder theory of the firm”. In: *Journal of Management and Governance* 9.1, pp. 5–32. ISSN: 13853457. DOI: 10.1007/s10997-005-1570-2.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee (2014). “Awards unbundled: Evidence from a natural field experiment”. In: *Journal of Economic Behavior & Organization* 100, pp. 44–63. ISSN: 01672681. DOI: 10.1016/j.jebo.2014.01.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0167268114000079.

Atkinson, A and Joseph E Stiglitz (1980). *Lectures on Public Economics*. New York, NY: McGrow-Hill.

Ayres, R. U., L. W. Ayres, and I. Rade (2002). “The life cycle of copper, its co-products and byproducts”.

Azmat, Ghazala and Nagore Iriberri (2012). *The provision of relative performance feedback information : An experimental analysis of performance and happiness*.

Baker, George P. and Thomas N. Hubbard (2004). “Contractibility and asset ownership: On-board computers and governance in U.S. trucking”. In: *The Quarterly Journal of Economics* 119.4, pp. 1443–1479.

Banerjee, Abhijit V. (1997). "A theory of misgovernance". In: *The Quarterly Journal of Economics* 112.4, pp. 1289–1332.

Belenzon, Sharon and Ulya Tsolmon (2016). "Market frictions and the competitive advantage of internal labor markets". In: *Strategic Management Journal* 37, pp. 1280–1303. ISSN: 00014273.

Benabou, Roland and Jean Tirole (2002). "Self-confidence and personal motivation". In: *Quarterly Journal of Economics* 117.3, pp. 871–915.

Bénabou, Ronald and Jean Tirole (2003). "Intrinsic and extrinsic motivation". In: *Review of Economic Studies* 70, pp. 489–520. ISSN: 0098-6283. DOI: 10.1177/0098628312437704.

Bennett, Victor Manuel (2013). "Organization and Bargaining: Sales Process Choice at Auto Dealerships". In: *Management Science* 59.January 2015, pp. 2003–2018. ISSN: 0025-1909. DOI: 10.1287/mnsc.1120.1691. URL: http://search.ebscohost.com/login.aspx?direct=true{\&}db=bth{\&}AN=90132167{\&}site=ehost-live{\&}scope=site.

Berchicci, Luca, Glen Dowell, and Andrew A. King (2012). "Environmental capabilities and corporate strategy: Exploring acquisitions among US manufacturing firms". In: *Strategic Management Journal* 33, pp. 1053–1071.

Berinsky, Adam J., Gregory a. Huber, and Gabriel S. Lenz (2012). "Evaluating online labor markets for experimental research: Amazon.com's mechanical turk". In: *Political Analysis* 20.3, pp. 351–368. ISSN: 10471987. DOI: 10.1093/pan/mpr057.

Bertrand, Marianne and Sendhil Mullainathan (2003). "Enjoying the quiet life? Corporate governance and managerial preferences". In: *Journal of Political Economy* 111.5, pp. 1043–1075.

Betcherman, Gordon (2012). "Labor market institutions: A review of the literature". Washington, DC.

Blanes i Vidal, Jordi and Mareike Nossol (2011). "Tournaments without prizes: Evidence from personnel records". In: *Management Science* 57.11, pp. 1721–1736.

Blau, Francine D and Lawrence M. Kahn (1999). "Institutions and laws in the labor market". In: *Handbook of Labour Economics* 3, pp. 1399–1461. ISSN: 1573-4463. DOI: 10.1016/S1573-4463(99)03006-0. URL: http://dx.doi.org/10.1016/S1573-4463(99)03006-0.

Borjas, George J and Kirk B Doran (2013). "Prizes and productivity: How winning the Fields medal affects scientific output".

Botero, Juan C et al. (2004). "The regulation of labor". In: *The Quarterly Journal of Economics* 119.4, pp. 1339–1382.

Bradler, Christiane et al. (2014). "Employee recognition and performance: A field experiment".

Brandenburger, Adam and Harborne Stuart (2007). "Biform games". In: *Management Science* 53.4, pp. 537–549.

Brandenburger, Adam M and Harborne W Jr Stuart (1996). "Value Based Business Strategy". In: *Journal of Economics & Management Strategy* 5.1, pp. 5–24. ISSN: 10586407. DOI: 10.1111/j.1430-9134.1996.00005.x.

BTS (2016). *Transportation Statistics Annual Report 2016*. Tech. rep. Washington, DC: U.S. Department of Transportation, Bureau of Transportaton Statistics.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011). "Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data?" In: *Perspectives on Psychological Science* 6.1, pp. 3–5. ISSN: 1745-6916. DOI: 10.1177/1745691610393980.

Burbano, Vanessa C. (2016). "Social responsibility messages and worker wage requirements: Field experimental evidence from online labor marketplaces". In: *Organization Science* 27.4, pp. 1010–1028. ISSN: 1047-7039. DOI: 10.1287/orsc.2016.1066. URL: http://pubsonline.informs.org/doi/10.1287/orsc.2016.1066.

Burdett, Kenneth and Dale T Mortensen (1998). "Wage differentials, employer size, and unemployment". In: *International Economic Review* 39.2, pp. 257–273.

Card, David, Francis Kramarz, and Thomas Lemieux (1999). "Changes in the relative structure of wages and employment: A comparison of the United States, Canada and France". In: *Canadian Journal of Economics* 32.4, pp. 843–877.

Charness, Gary, David Masclet, and Marie Claire Villeval (2014). "The dark side of competition for status". In: *Management Science* 60.1, pp. 38–55. ISSN: 0025-1909. DOI: 10.1287/mnsc.2013.1747. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2013.1747.

Checchi, Daniele et al. (2008). "Labour market institutions and income inequality". In: *Economic Policy* 23.56, pp. 601–649.

Coff, Russell W. (1999). "When competitive advantage doesn't lead to performance: The resource-based view and stakeholder bargaining power". In: *Organization Science* 10.2, pp. 119–133. ISSN: 1047-7039. DOI: 10.1287/orsc.10.2.119. URL: http://pubsonline.informs.org/doi/abs/10.1287/orsc.10.2.119.

Corporation, Southern Copper (2009). *10-K Report*. Tech. rep.

Costello, Bob and Rod Suarez (2015). *Truck driver shortage analysis 2015*. Tech. rep. American Trucking Associations. URL: http://www.trucking.org/ATADocs/NewsandInformation/ReportsTrendsandStatistics/10615ATAsDriverShortageReport2015.pdf.

Croson, Rachel and Uri Gneezy (2009). "Gender differences in preferences". In: *Journal of Economic Literature* 47.2, pp. 448–474. ISSN: 0022-0515. DOI: 10.1257/jel.47.2.448.

Csikszentmilhalyi, M and J Lefevre (1989). "Optimal experience in work and leisure". In: *Journal of Personality and Social Psychology* 56.5, pp. 815–822.

Deci, Edward, Wayne Cascio, and Judy Krusell (1973). "Sex differences, positive feedback and intrinsic motivation". In: *Eastern Psychological Association Convention*. URL: http://eric.ed.gov/?id=ED076904.

Deci, Edward L. (1971). "Effects of externally mediated rewards on intrinsic motivation". In: *Journal of Personality and Social Psychology* 18.1, pp. 105–115.

— (1972a). "Intrinsic motivation, extrinsic reinforcement, and inequity". In: *Journal of Personality and Social Psychology* 22.1, pp. 113–120. ISSN: 0022-3514. DOI: 10.1037/h0032355. URL: http://content.apa.org/journals/psp/22/1/113.

— (1972b). "The effects of contingent and noncontingent rewards and controls on intrinsic motivation". In: *Organizational Behavior and Human Performance* 8, pp. 217–229.

Deci, Edward L., Wayne F. Cascio, and Judith Krusell (1975). *Cognitive evaluation theory and some comments on the Calder and Staw critique*. DOI: 10.1037/h0076168. URL: http://content.apa.org/journals/psp/31/1/81.

Deci, Edward L., James P. Connell, and Richard M. Ryan (1989). "Self-determination in a work organization". In: *Journal of Applied Psychology* 74.4, pp. 580–590. ISSN: 1939-1854. DOI: 10.1037/0021-9010.74.4.580. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0021-9010.74.4.580.

Deci, Edward L., Richard Koestner, and Richard M. Ryan (1999). "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation". In: *Psychological Bulletin* 125.6, pp. 627–668.

Deci, Edward L., Anja H. Olafsen, and Richard M. Ryan (2017). "Self-determination theory in work organizations: The state of a science". In: *Annual Review of Organizational Psychology and Organizational Behavior* 4, pp. 19–43. ISSN: 2327-0608. DOI: 10.1146/annurev-orgpsych-032516-113108. URL: http://www.annualreviews.org/doi/10.1146/annurev-orgpsych-032516-113108.

Deci, Edward L. and Richard M. Ryan (1980). “The empirical exploration of intrinsic motivational proceses”. In: *Advances in Environmental Social Psychology, Vol 13*, pp. 39–80.

— (2000). “The ”what” and ”why” of goal pursuits: Human needs and the self-determination of behavior”. In: *Psychological Inquiry* 11.4, pp. 227–268. ISSN: 0956-2478. DOI: 10.1207/S15327965PLI1104. arXiv: 9809069v1 [arXiv:gr-qc].

Denny, Paul (2013). “The effect of virtual achievements on student engagement”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. New York, New York, USA: ACM Press, pp. 763–772. ISBN: 9781450318990. DOI: 10.1145/2470654.2470763. URL: http://dl.acm.org/citation.cfm?doid=2470654.2470763.

Deterding, Sebastian et al. (2011). “From game design elements to gamefulness: defining ”gamification””. In: *Proceedings of the 15th international Academic MindTrek Conference: Envisioning Future Media Environments*. Tampere, Finland, pp. 9–15. ISBN: 9781450308168.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996). “Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach”. In: *Econometrica* 64.5, pp. 1001–1044. ISSN: 00129682. DOI: 10.2307/2171954. URL: http://www.jstor.org/stable/2171954?origin=crossref.

DOE (2017). *Fuel economy in cold weather.* Tech. rep. U.S. Department of Energy. URL: https://www.fueleconomy.gov/feg/coldweather.shtml.

Domínguez, Adrián et al. (2013). “Gamifying learning experiences: Practical implications and outcomes”. In: *Computers and Education* 63, pp. 380–392. ISSN: 03601315. DOI: 10.1016/j.compedu.2012.12.020. URL: http://dx.doi.org/10.1016/j.compedu.2012.12.020.

Douglas, Genevieve (2013). *Survey finds many employers not satisfied with their pay-for-performance programs*. URL: http://www.bna.com/survey-finds-employers-n17179881003/ (visited on 09/05/2015).

Dur, Robert (2009). “Gift exchange in the workplace: Money or attention?” In: *Journal of the European Economic Association* 7.2-3, pp. 550–560. ISSN: 1542-4766. DOI: 10.1162/JEEA.2009.7.2-3.550. URL: http://doi.wiley.com/10.1162/JEEA.2009.7.2-3.550.

Easley, David and Arpita Ghosh (2013). “Incentives, gamification, and game theory: an economic approach to badge design”. In: *Proceedings of the fourteenth Electronic commerce*, pp. 359–376. doi: 10.1145/2482540.2482571. url: http://dl.acm.org/citation.cfm?id=2492002.2482571.

EPA (1976). *Factors affecting automotive fuel economy*. Tech. rep. U.S. Environmental Protection Agency.

Ericsson, M. and F. L Massey (2011). "Overview of state ownership in the global minerals industry". Washington, DC.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval (2009). "Feedback and incentives: Experimental evidence". In: *Labour Economics* 16.6, pp. 679–688. issn: 09275371. doi: 10.1016/j.labeco.2009.08.006. url: http://dx.doi.org/10.1016/j.labeco.2009.08.006.

Fama, Eugene F. and Michael C. Jensen (1983). "Separation of ownership and control". In: *The Journal of Law & Economics* 26.2, pp. 301–325.

Farzan, Rosta et al. (2008). "Results from deploying a participation incentive mechanism within the enterprise". In: *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, p. 563. doi: 10.1145/1357054.1357145. url: http://dl.acm.org/citation.cfm?id=1357054.1357145.

Ferrell, Jared Z. et al. (2016). "Gamification of human resource processes". In: *Emerging Research and Trends in Gamification*. Ed. by Harsha Gangadharbatla and Donna Z. Davis. Hershey, PA.

Fizek, Sofia (2014). "Why fun matters: In search of emergent playful experiences". In: *Rethinking gamification*. Ed. by Mathias Fuchs et al. Lüneburg, Germany: Meson Press, pp. 273–287.

Freeman, Richard B., Douglas Kruse, and Joseph Blasi (2008). "The same yet different: Worker reports on labour practices and outcomes in a single firm across countries". In: *Labour Economics* 15, pp. 334–355. issn: 09275371. doi: 10.1016/j.labeco.2007.06.011.

Frey, Bruno S (2006). "Giving and receiving awards". In: 1.4, pp. 377–388. doi: 10.1111/j.1745-6916.2006.00022.x.

Frey, Bruno S. (2007). "Awards as compensation". In: *European Management Review* 4.1, pp. 6–14. issn: 17404754. doi: 10.1057/palgrave.emr.1500068. url: http://doi.wiley.com/10.1057/palgrave.emr.1500068.

Frey, Bruno S and Jana Gallus (2014). "Awards are a special kind of signal".

Frey, Bruno S. and Reto Jegen (2001). "Motivation crowding theory". In: *Journal of Economic Surveys* 15.5, pp. 589–611. issn: 0950-0804. doi: 10.1111/1467-6419.00150. url:

http://www.blackwell-synergy.com/links/doi/10.1111/1467-6419.00150{\%}5Cnhttp://onlinelibrary.wiley.com/doi/10.1111/1467-6419.00150/pdf.

Frey, Bruno S. and Susanne Neckermann (2008). “Awards - A view from psychological economics”. In: *Zeitschrift für Psychologie / Journal of Psychology* 216.4, pp. 198–208. ISSN: 0044-3409. DOI: 10.1027/0044-3409.216.4.198.

Frey, Bruno S. and Alois Stutzer (2002). *Happiness and economics: How the economy and institutions affect human well-being*. Princeton, NJ: Princeton University Press.

Gallus, Jana (2016). “Fostering public good contributions with symbolic awards: A large-scale natural field experiment at wikipedia”. In: *Management Science* Articles in Advance. ISSN: 0025-1909. DOI: 10.1287/mnsc.2016.2540. URL: http://pubsonline.informs.org/doi/10.1287/mnsc.2016.2540.

Gallus, Jana and Bruno S. Frey (2016). “Awards: A strategic management perspective”. In: *Strategic Management Journal* 37, pp. 1699–1714. ISSN: 00014273. DOI: 10.1002/smj. arXiv: 1.

Gans, J. S. (2005). *Core Economics for Managers*. South Melbourne: Thomson.

Gans, Joshua and Michael D. Ryall (2017). “Value capture theory: A strategic management review”. In: *Strategic Management Journal* 38, pp. 17–41. ISSN: 00014273.

Garcia-Castro, Roberto and Ruth V. Aguilera (2015). “Incremental value creation and appropriation in a world with multiple stakeholders”. In: *Strategic Management Journal* 36, pp. 137–147.

Gartner Inc. (2012a). *Gartner predicts over 70 percent of Global 2000 organisations will have at least one gamified application by 2014*. URL: http://www.gartner.com/newsroom/id/1844115 (visited on 02/17/2016).

— (2012b). *Gartner says by 2014, 80 percent of current gamified applications will fail to meet business objectives primarily due to poor design*. URL: http://www.gartner.com/newsroom/id/2251015 (visited on 02/17/2016).

Gibson, David et al. (2013). “Digital badges in education”. In: *Education and Information Technologies* 20.2, pp. 403–410. ISSN: 13602357. DOI: 10.1007/s10639-013-9291-7.

Grandey, Alicia A, Anita P Tam, and Analea L Brauburger (2002). “Affective states and traits in the workplace: Diary and survey data from young workers”. In: *Motivation and Emotion* 26.1, pp. 31–55. ISSN: 01467239. DOI: 10.1023/A:1015142124306.

Grant, Adam M and Francesca Gino (2010). “A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior.” In: *Journal of personality and social psychology* 98.6, pp. 946–955. ISSN: 0022-3514. DOI: 10.1037/a0017935.

Grant, Scott and Buddy Betts (2013). “Encouraging user behaviour with achievements: An empirical study”. In: *IEEE International Working Conference on Mining Software Repositories*, pp. 65–68. ISSN: 21601852. DOI: 10.1109/MSR.2013.6624007.

Greckhamer, Thomas (2016). “CEO compensation in relation to worker compensation across countries: The configuration of country-level institutions”. In: *Strategic Management Journal* 37, pp. 793–815. DOI: 10.1002/smj.2370. arXiv: 1.

Grennan, Matthew (2014). “Bargaining Ability and Competitive Advantage: Empirical Evidence from Medical Devices”. In: *Management Science* 60.12, pp. 3011–3025. ISSN: 0025-1909. DOI: 10.1287/mnsc.2014.2006. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2014.2006.

Grossman, S. J. and O. D. Hart (1980). “Disclosure laws and takeover bids”. In: *The Journal of Finance* 35.2, pp. 323–334.

Gubler, Timothy, Ian Larkin, and Lamar Pierce (2016). “Motivational spillovers from awards: Crowding out in a multitasking environment”. In: *Organization Science* 27.2, pp. 286–303. ISSN: 1047-7039. DOI: 10.1287/orsc.2016.1047. URL: http://pubsonline.informs.org/doi/10.1287/orsc.2016.1047.

Haaranen, Lassi et al. (2014). “How (not) to introduce badges to online exercises”. In: *Proceedings of the 45th ACM technical symposium on Computer science education - SIGCSE '14*, pp. 33–38. DOI: 10.1145/2538862.2538921. URL: http://dl.acm.org/citation.cfm?id=2538862.2538921{\%}5Cnhttp://www.scopus.com/inward/record.url?eid=2-s2.0-84899752672{\&}partnerID=tZOtx3y1.

Hackman, J. Richard and Greg R. Oldham (1976). “Motivation through the design of work: Test of a theory”. In: *Organizational Behavior and Human Performance* 16, pp. 250–279. ISSN: 00305073. DOI: 10.1016/0030-5073(76)90016-7. arXiv: arXiv:1011.1669v3.

Hakulinen, Lasse, Tapio Auvinen, and Ari Korhonen (2013). “Empirical study on the effect of achievement badges in TRAKLA2 online learning environment”. In: *Proceedings - 2013 Learning and Teaching in Computing and Engineering, LaTiCE 2013*, pp. 47–54. ISSN: 18630383. DOI: 10.1109/LaTiCE.2013.34.

Hamari, Juho (2011). “Framework for designing and evaluating game achievements”. In: *Proceedings of DiGRA 2011 Conference: Think Design Play*, p. 20. arXiv: 11307.59151. URL: http://www.mendeley.com/catalog/framework-designing-evaluating-game-achievements/.

Hamari, Juho, Kai Huotari, and Juha Tolvanen (2015). “Gamification and economics”. In: *The gameful world: Approaches, issues, applications*. Ed. by Steffen P. Walz and Sebastian Deterding. Cambride, Massachusetts: The MIT Press.

Hamari, Juho and Jonna Koivisto (2013). “Social motivations to use gamification : An empirical study of gamifying exercise”. In: *Proceedings of the 21st European Conference on Information Systems*. Utrecht, The Netherlands.

Hamari, Juho, Jonna Koivisto, and Harri Sarsa (2014). “Does gamification work? - A literature Review of Empirical Studies on Gamification”. In: *2014 47th Hawaii International Conference on System Sciences*, pp. 3025–3034. DOI: 10.1109/HICSS.2014.377. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6758978.

Hart, Oliver, Andrei Shleifer, and Robert W. Vishny (1997). “The proper scope of government: Theory and application to prisons”. In: *The Quarterly Journal of Economics* 112.4, pp. 1127–1161.

Harvey, Joan, Neil Thorpe, and Richard Fairchild (2013). “Attitudes towards and perceptions of eco-driving and the role of feedback systems”. In: *Ergonomics* 56.3, pp. 507–521. ISSN: 0014-0139. DOI: 10.1080/00140139.2012.751460. URL: http://www.tandfonline.com/doi/abs/10.1080/00140139.2012.751460{\%}5Cnhttp://www.tandfonline.com/doi/pdf/10.1080/00140139.2012.751460.

Horton, John and Lydia Chilton (2010). “The labor economics of paid crowdsourcing”. In: *Proceedings of the 11th ACM conference on Electronic Commerce*. 1, pp. 209–218. ISBN: 9781605588223. DOI: 10.1145/1807342.1807376. arXiv: 1001.0627. URL: http://arxiv.org/abs/1001.0627.

Hubbard, Thomas N (2000). “The demand for monitoring technologies: The case of trucking”. In: *The Quarterly Journal of Economics* 115.2, pp. 533–560.

Huberman, B. a., C. H. Loch, and a. ONculer (2004). “Status As a Valued Resource”. In: *Social Psychology Quarterly* 67.1, pp. 103–114. ISSN: 0190-2725. DOI: 10.1177/019027250406700109.

Huotari, Kai (2012). “Defining gamification - A service marketing perspective”. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pp. 17–22. DOI: 10.1145/2393132.2393137.

Inoue, Carlos F K V, Sergio G. Lazzarini, and Aldo Musacchio (2013). “Leviathan as a minority shareholder: Firm-level implications of state equity purchases”. In: *Academy of Management Journal* 56.6, pp. 1775–1801. ISSN: 00014273. DOI: 10.5465/amj.2012.0406.

Ipeirotis, Panagiotis G. (2010). "Analyzing the amazon mechanical turk marketplace". In: *XRDS: Crossroads, The ACM Magazine for Students* 17.2, p. 16. ISSN: 15284972. DOI: 10.1145/1869086.1869094.

Itkin, Dina (2007). "Wage and Employment Patterns in the Mining Sector". In: *Occupational Employment and Wages*. Bureau of Labor Statistics. Bulletin 2585.

Jaccard, J., R. Turrisi, and C. Wan (1990). *Interaction Effects in Multiple Regression*. Thousand Oaks, CA: Sage.

Jensen, Michael C. and William H. Meckling (1976). "Theory of the firm: Managerial behavior, agency costs and ownership structure". In: *Journal of Financial Economics* 3, pp. 305–360. ISSN: 0304405X. DOI: 10.1016/0304-405X(76)90026-X. arXiv: 1.

Jones, Julie, Nathan Altadonna, and W Lindsey (2012). "We don't need no stinkin' badges : Examining the social role of badges in the Huffington Post". In: *Proceedings of the CSCW*, pp. 249–252. DOI: 10.1145/2145204.2145244.

Kanfer, Ruth and Phillip L. Ackerman (1989). "Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition". In: *Journal of Applied Psychology* 74.4, pp. 657–690. DOI: 10.1037/0021-9010.74.4.657.

Kast, Audrey and Kathleen Connor (1988). "Sex and age differnces in response to informational and controlling feedback". In: *Personality and Social Psychological Bulletin* 14.3, pp. 514–523.

Kittur, a, E H Chi, and B Suh (2008). "Crowdsourcing user studies with Mechanical Turk". In: *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 453–456. ISSN: 01651781. DOI: 10.1145/1357054.1357127. URL: http://dl.acm.org/citation.cfm?id=1357127.

Koch, Marianne J. and Rita G. McGrath (1996). "Improving labor productivity: Human resource management policies do matter". In: *Strategic Management Journal* 17.5, pp. 335–354.

Koeniger, Winfried, Marco Leonardi, and Luca Nunziata (2007). "Labor market institutions and wage inequality". In: *Industrial and Labor Relations Review* 60.3, pp. 340–356.

Koivisto, Jonna and Juho Hamari (2014). "Demographic differences in perceived benefits from gamification". In: *Computers in Human Behavior* 35, pp. 179–188. ISSN: 07475632. DOI: 10.1016/j.chb.2014.03.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S0747563214001289.

Kosfeld, Michael and Susanne Neckermann (2011). "Getting more work for nothing? Symbolic awards and worker performance". In: *American Economic Journal: Microeconomics* 3.August, pp. 86–99.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe (2012). "The currency of reciprocity: gift exchange in the workplace". In: *The American Economic Review* 102.4, pp. 1644–1662.

Kuhnen, Camelia M and Agnieszka Tymula (2012). "Feedback, self-Esteem, and performance in organizations". In: *Management Science* 58.1, pp. 94–113.

La Porta, Rafael et al. (1998). "Law and finance". In: *Journal of Political Economy* 106.6, pp. 11131–55. ISSN: 00223808. DOI: `10.1086/250042`.

La Tercera (2015). *Codelco estima en US 27 millones el costo diario de movilización de contratistas*.

Lacetera, Nicola and Mario Macis (2010). "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme". In: *Journal of Economic Behavior and Organization* 76.2, pp. 225–237. ISSN: 01672681. DOI: `10.1016/j.jebo.2010.08.007`. URL: `http://dx.doi.org/10.1016/j.jebo.2010.08.007`.

Landers, Richard N and Rachel C Callan (2011). "Casual social games as serious games: The psychology of gamification in undergraduate education and employee training". In: *Serious Games and Edutainment Applications*. Ed. by Minhua Ma, Andreas Oikonomou, and Lakhmi C. Jain. London: Springer. Chap. 20, pp. 399–423. ISBN: 978-1-4471-2160-2. DOI: `10.1007/978-1-4471-2161-9`. URL: `http://link.springer.com/10.1007/978-1-4471-2161-9`.

Larkin, Ian, Lamar Pierce, and Francesca Gino (2012). "The psychological costs of pay-for-performance: Implications for the strategic compensation of employees". In: *Strategic Management Journal* 33, pp. 1194–1214. DOI: `10.1002/smj`.

Larrick, Richard P. and Jack B. Soll (2008). "The MPG illusion". In: *Science* 320.5883, pp. 1593–1594. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1154983`. URL: `http://www.sciencemag.org/content/320/5883/1593{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/18566271{\%}5Cnhttp://www.waccobb.net/forums/showthread.php?38094-The-MPG-Illusion`.

Latham, Gary P. (2012). *Work motivation: History, theory, research, and pratice*. Second. Sage Publications.

Latham, Gary P. and Edwin A. Locke (2007). "New developments in and directions for goal-setting research". In: *European Psychologist* 12.4, pp. 290–300. ISSN: 10169040. DOI: 10.1027/1016-9040.12.4.290.

Lazear, Edward P. (2000). "Performance pay and productivity". In: *The American Economic Review* 90.5, pp. 1346–1361.

Locke, Edwin A. and Gary P. Latham (2002). "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey." In: *American Psychologist* 57.9, pp. 705–717. ISSN: 1935-990X. DOI: 10.1037/0003-066X.57.9.705. arXiv: 1603.07846. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.57.9.705.

Loderer, Claudio and Urs Waelchli (2010). "Firm age and performance".

Loewenstein, George F. and S Issacharoff (1994). "Source dependence in the valuation of objects". In: *Journal of Behavioral Decision Making* 7, pp. 157–168. ISSN: 0894-3257. DOI: 10.1002/bdm.3960070302. URL: http://onlinelibrary.wiley.com/doi/10.1002/bdm.3960070302/abstract.

MacDonald, Glenn and Michael D. Ryall (2004). "How Do Value Creation and Competition Determine Whether a Firm Appropriates Value?" In: *Management Science* 50, pp. 1319–1333. ISSN: 0025-1909. DOI: 10.1287/mnsc.1030.0152.

Malmendier, Ulrike and Tate Geoffrey (2009). "SUPERSTAR CEOs". In: *The Quarterly Journal of Economics* 124.4, pp. 1593–1638.

Mas, Alexandre and Moretti Enrico (2009). "Peers at work". In: *The American Economic Review* 99.1, pp. 112–145.

Mason, J. (2015). *Strike like it's 2011? Low copper prices loom large over wage talks.* URL: Strikelikeit's2011?Lowcopperpricesloomlargeoverwagetalks.http://www.reuters.com/article/2015/04/20/chile-copper-labor-idUSL2N0X900V20150420{\#}eiieOqfwuqBUB2ma.97) (visited on 04/16/2016).

Mason, Winter and Siddharth Suri (2012). "Conducting behavioral research on Amazon's Mechanical Turk". In: *Behavior Research Methods* 44.1, pp. 1–23. ISSN: 1554-3528. DOI: 10.3758/s13428-011-0124-6. arXiv: /ssrn.com/abstract=1691163 [http:].

Mathieu, John E. and Scott B. Button (1992). "An examination of the relative impact of normative information and self-efficacy on personal goals and performance over time". In: *Journal of Applied Psychology* 22.22, pp. 1758–1775.

Mcgonigal, Jane (2011). *Reality is broken: Why games make us better and how they can change the world*. New York, NY: Penguin Books, p. 416. ISBN: 9780143120612.

Mikesell, R. (2011). *The World Copper Industry: Structure and Economic Analysis*. Washington, DC: Taylor & Francis.

Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007). "Contests for status". In: *Journal of Political Economy* 115.2, pp. 338–363. ISSN: 0022-3808. DOI: 10.1086/518010. URL: http://www.jstor.org/stable/10.1086/518010.

Mollick, Ethan R. and Nancy Rothbard (2014). "Mandatory fun: Gamification and the impact of games at work". URL: http://www.ssrn.com/abstract=2277103.

Molloy, Janice C. and Jay Bryan Barney (2015). "Who captures the value created with human capital? A market-based view". In: *Academy of Management Perspectives* 29.3, pp. 309–325. DOI: 10.5465/amp.2014.0152.

Montola, Markus et al. (2009). "Applying game achievement systems to enhance user experience in a photo sharing service". In: *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, pp. 94–97. DOI: 10.1145/1621841.1621859. URL: http://portal.acm.org/citation.cfm?id=1621841.1621859{\&}coll=GUIDE{\&}dl=GUIDE{\&}CFID=103096974{\&}CFTOKEN=12484716.

Morck, Randall, Andrei Shleifer, and Robert W Vishny (1988). "Management ownership and market valuation". In: *Journal of Financial Economics* 20, pp. 293–315.

Musacchio, Aldo and Sergio G. Lazzarini (2014). *Reinventing State Capitalism*. Cambride, Massachusetts: Harvard University Press.

Neckermann, Susanne, Reto Cueni, and Bruno S. Frey (2014). "Awards at work". In: *Labour Economics* 31, pp. 205–217. ISSN: 09275371. DOI: 10.1016/j.labeco.2014.04.002. URL: http://dx.doi.org/10.1016/j.labeco.2014.04.002.

Neckermann, Susanne and Bruno S. Frey (2013). "And the winner is. . .? The motivating power of employee awards". In: *Journal of Socio-Economics* 46, pp. 66–77. ISSN: 10535357. DOI: 10.1016/j.socec.2013.06.006. URL: http://dx.doi.org/10.1016/j.socec.2013.06.006.

Nickerson, Jack A and Todd R Zenger (2008). "Envy, comparison costs, and the economic theory of the firm". In: *Strategic Management Journal* 1449.May, pp. 1429–1449. DOI: 10.1002/smj.

Niederle, Muriel and Lise Vesterlund (2007). "Do women shy away from competition? Do men compete too much?" In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101.

Obloj, Tomasz and Metin Sengul (2012). "Incentive life-cycles: Learning and the division of value in firms". In: *Administrative Science Quarterly* 57.2, pp. 305–347. ISSN: 0001-8392.

DOI: 10.1177/0001839212453833. URL: http://journals.sagepub.com/doi/10.1177/0001839212453833.

OECD (2011). *Divided We Stand: Why Inequality Keeps Rising*. OECD Publishing.

Omnitracs (2012). *What is Idling and Why Does It Matter? (Part 1 of 2)*. URL: http://www.omnitracs.com/blog/what-idling-and-why-does-it-matter-part-1-2 (visited on 09/16/2016).

Paolacci, Gabriele, Jesse Chandler, and Pg Ipeirotis (2010). "Running experiments on amazon mechanical turk". In: *Judgment and Decision making* 5.5, pp. 411–419. ISSN: 19302975. DOI: 10.2139/ssrn.1626226. URL: http://repub.eur.nl/pub/31983.

Pekrun, R. and M. Frese (1992). "Emotions in work and achievement". In: *International Review of Industrial and Organizational Psychology* 7, pp. 153–200.

Phene, Anupama, Karin Fladmoe-Lindquist, and Laurence Marsh (2006). "Breakthrough innovations in the U.S. biotechnology industry: The effects of technological space and geographic origin". In: *Strategic Management Journal* 27, pp. 369–388. ISSN: 01432095. DOI: 10.1002/smj.522.

Prendergast, Canice (1999). "The provision of incentives in firms". In: *Journal of Economic Literature* 37.1, pp. 7–63. ISSN: 0022-0515. DOI: 10.1257/jel.37.1.7. arXiv: z0022.

Przybylski, Andrew K., C. Scott Rigby, and Richard M. Ryan (2010). "A motivational model of video game engagement." In: *Review of General Psychology* 14.2, p. 154. ISSN: 1089-2680. DOI: 10.1037/a0019440. URL: http://psycnet.apa.org/journals/gpr/14/2/154/.

Restivo, Michael and Arnout van de Rijt (2012). "Experimental study of informal rewards in peer production". In: *PLoS ONE* 7.3, pp. 30–32. ISSN: 19326203. DOI: 10.1371/journal.pone.0034358.

Robertson, Margaret (2010). *Can't play, won't play*. URL: http://kotaku.com/5686393/cant-play-wont-play (visited on 03/09/2017).

Rodrik, Dani (1999). "Democracies pay higher wages". In: *The Quarterly Journal of Economics* 114.3, pp. 707–738.

Rosaz, Julie and Marie Claire Villeval (2012). "Lies and biased evaluation: A real-effort experiment". In: *Journal of Economic Behavior and Organization* 84.2, pp. 537–549. ISSN: 01672681. DOI: 10.1016/j.jebo.2012.09.002. URL: http://dx.doi.org/10.1016/j.jebo.2012.09.002.

Ryan, Richard M. (1982). “Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory”. In: *Journal of Personality and Social Psychology* 43.3, pp. 450–461.

Ryan, Richard M. and James P. Connell (1989). “Perceived locus of causality and internalization: Examining reasons for acting in two domains”. In: *Journal of Personality and Social Psychology* 57.5, pp. 749–761. ISSN: 1939-1315. DOI: 10.1037/0022-3514.57.5.749. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.57.5.749.

Ryan, Richard M. and Edward L. Deci (2000). “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions.” In: *Contemporary educational psychology* 25.1, pp. 54–67. ISSN: 0361-476X. DOI: 10.1006/ceps.1999.1020. URL: http://www.ncbi.nlm.nih.gov/pubmed/10620381.

Sapienza, Paola (2004). “The effects of government ownership on bank lending”. In: *Journal of Financial Economics* 72, pp. 357–384. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2002.10.002.

Seaborn, Katie and Deborah I. Fels (2015). “Gamification in theory and action: A survey”. In: *International Journal of Human Computer Studies* 74, pp. 14–31. ISSN: 10959300. DOI: 10.1016/j.ijhcs.2014.09.006.

Shan, Liwei, Shihe Fu, and Lu Zheng (2016). “Corporate sexual equality and firm performance”. In: *Strategic Management Journal*. DOI: 10.1002/smj.2624.

Shapiro, Carl and Joseph E. Stiglitz (1984). “Equilibrium unemployment as worker discipline device”. In: *The American Economic Review* 74.3, pp. 433–444.

Shearer, Bruce (2004). “Piece rates, fixed wages and incentives: Evidence from a field experiment”. In: *Review of Economic Studies* 71.2, pp. 513–534. ISSN: 00346527. DOI: 10.1111/0034-6527.00294.

Shleifer, A. and R. W. Vishny (1994). “Politicians and Firms”. In: *The Quarterly Journal of Economics* 109.4, pp. 995–1025. ISSN: 0033-5533. DOI: 10.2307/2118354. URL: https://academic.oup.com/qje/article-lookup/doi/10.2307/2118354.

Shleifer, Andrei (1998). “State versus private ownership”. In: *The Journal of Economic Perspectives* 12.4, pp. 133–150.

Shleifer, Andrei and Robert W Vishny (1986). “Large shareholders and corporate control”. In: *Journal of Political Economy* 94.3, pp. 461–488. DOI: 10.1002/jae.l225.

Siero, Sjef et al. (1989). “Modification of driving behavior in a large transport organization: A field experiment.” In: *Journal of Applied Psychology* 74.3, pp. 417–423. ISSN: 0021-9010. DOI: 10.1037/0021-9010.74.3.417.

Sleuwaegen, Leo and Priscilla Boiardi (2014). "Creativity and regional innovation: Evidence from EU regions". In: *Research Policy* 43, pp. 1508–1522. ISSN: 00487333. DOI: 10.1016/j.respol.2014.03.014. URL: http://dx.doi.org/10.1016/j.respol.2014.03.014.

Smith, Ross (2012). "Productivity games to increase organizational trust: How gaming and fun in the workplace can develop a culture of trust and innovation". In: *2012 IEEE International Games Innovation Conference*, pp. 1–6. DOI: 10.1109/IGIC.2012.6329831. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6329831.

Stajkovic, Alexander D. and Fred Luthans (1998). "Self-Efficacy and Work-Related Performance: A Meta-Analysis". In: *Psychological Bulletin* 124.2, pp. 240–261.

Stulz, René M. (2005). "The limits of financial globalization". In: *The Journal of Finance* 17.2, pp. 1595–1638.

Suazo, Mark M., Patricia G. Martínez, and Rudy Sandoval (2009). "Creating psychological and legal contracts through human resource practices: A signaling theory perspective". In: *Human Resource Management Review* 19.2, pp. 154–166. ISSN: 10534822. DOI: 10.1016/j.hrmr.2008.11.002. URL: http://dx.doi.org/10.1016/j.hrmr.2008.11.002.

Tian, Lihui and Saul Estrin (2008). "Retained state shareholding in Chinese PLCs: Does government ownership always reduce corporate value?" In: *Journal of Comparative Economics* 36, pp. 74–89. ISSN: 01475967. DOI: 10.1016/j.jce.2007.10.003.

Tulusan, Johannes, Thorsten Staake, and Elgar Fleisch (2012). "Providing eco-driving feedback to corporate car drivers: what impact does a smartphone application have on their fuel efficiency?" In: *Proceedings of the 2012 ACM conference on ubiquitous computing*, pp. 212–215. ISBN: 9781450312240. DOI: 10.1145/2370216.2370250. URL: http://dl.acm.org/citation.cfm?id=2370250{\%}5Cnhttp://cocoa.ethz.ch/downloads/2013/08/1203{_}ubicomp{_}tulusan{_}staake{_}fleisch{_}final{_}1.pdf.

U.S. Congress Office of Technology Assessment (1988). *Copper: Technology and competitiveness*. Tech. rep. Washington, DC.

"Value creation, competition, and performance in buyer-supplier relationships" (2010). In: *Strategic Management Journal* 32, pp. 76–102. ISSN: 00014273.

Wang, Kun and Greg Shailer (2015). "Ownership concentration and firm performance in emerging markets: A meta-analysis". In: *Journal of Economic Surveys* 29.2, pp. 199–229. ISSN: 14676419. DOI: 10.1111/joes.12048.

WDR (2013). *Occupational Wages around the World, World Bank Group.*

Weiss, Howard M. and R. Cropanzano (1996). "Affective events theory: A theoretical discussion of the structure, causes, and consequences of affective experiences at work". In: *Research in Organizational Behavior* 18, pp. 1–74.

Williams, L. *Copper's top 10 Countries and Companies*. URL: http://www.mineweb.com/news/base-metals-and-minerals/coppers-top-10-countries-and-companies/ (visited on 02/03/2016).

Wooldridge, Jeffrey M. (2010). *Econometric analysis of cross section and panel data*. Second. MIT Press, p. 1096.

Woolley, Kaitlin and Ayelet Fishbach (2015). "The experience matters more than you think: People value intrinsic incentives more inside than outside an activity". In: *Journal of Personality and Social Psychology* 109.6, pp. 968–982. ISSN: 1939-1315. DOI: 10.1037/pspa0000035. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/pspa0000035.