Leisure Time instead of a Monetary Bonus?

A computerized real effort task to elicit productivity levels

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Behavioural Economics

C.C. Damen

Student number: 474165cd

Supervisor: J.T.R. Stoop

Second assessor: T.R. Lambregts

Abstract

Recently, many big industries are doubting whether bonus schemes are the most optimal way to reward employees. Also, leisure time has gained high importance amongst the newest generations. This paper examines whether leisure time as an incentive could be as effective as the often used monetary reward. An experiment that consists of a computerized real-effort task is spread amongst (ex-) students. The finding is that the immediate responses on both incentives do not differ significantly. However, after the first few minutes are terminated, the monetary incentive wins it over a reward of leisure time. That is at least in the case of the currently used experimental design and amongst (ex-) students. As soon as either one of the two changes, results are likely to change accordingly. This paper provides valuable insights that further research in a different setting or amongst another audience can build upon.

Keywords: Incentive, Leisure time, monetary reward, productivity, preference

JEL classification: C21, C83, J22, J33, M52

ERASMUS UNIVERSITEIT ROTTERDAM

19-06-2018

"BONUS SCHEMES DOES NOT SEEM LIKE THE BEST OPTION ANYMORE" – PEET VOGELS (ALGEMEEN DAGBLAD, 2018; P. 16-17)

This heading of a Dutch news article shows exactly what this paper aims to examine. Recently, many big industries have dropped the monetary bonus to incentivize their workers. Why? Some academics say that bonus schemes lead to cheating on the workplace (Gill et al., 2013; Cadsby et al., 2010). Others mention the suppression of intrinsic motivation (Deci et al., 1999; Gneezy et al., 2011). This paper will test the productivity level that derives from a monetary bonus. To be more precise, the aim is to find out if a comparable reward of leisure time is able to substitute the monetary bonus.

In economics, a major theme is how to motivate workers to put in the most amount of effort. Until now, money is the most common used incentive. The basic idea is that higher effort is translated into better performance of workers and this, in turn, positively affects their pay (Bonner and Sprinkle, 2002). An early research by Locke et al. (1980) claims to find that employee productivity rose with 30 percent, on average, after individual monetary incentives were implemented. More importantly, it is said that other types of rewards are less effective.

On the other hand, anecdotal and empirical evidence sometimes suggest that monetary incentives have varying effects on the effort level and may even worsen performance (Bonner et al., 2000; Camerer and Hogarth, 1999; Ariely et al., 2009). A more recent study performed by Bradler et al. (2016) has confirmed that non-financial motivators might work as well. In this case, public recognition positively influences performance. Recognition is just one of the many non-financial motivators that are possible. For this reason, Fisher (2015) correctly claims that non-monetary incentives in order to motivate workers are often overlooked.

This paper ensures to shed light on a non-financial incentive, which is leisure time. Time can be seen as a similar scarce resource just like money. Amongst the newest generations, leisure time has become more and more desirable (Twenge, 2010). We all aim to allocate our time amongst various activities (schooling, work, sports, family, friends, and so on). While doing so we are always complaining about a loss of time, and aim to seek for more. Rajagopal and Rha (2009) notice that people are even willing to spend money on tools in order to be more effective in managing time.

In other words, it is evident that people care about time. The question remains *how much* we care about time. This paper performs an experiment whereby people are assigned to the same target but with different incentives. Accordingly, the main question of this paper is: *Does money or leisure time works best to induce the highest productivity level?*

Previous studies have already shown related insights into this topic. A first, most obvious study to address is a field experiment that has been performed in Sweden. It is found that shorter working days for nurses in a retirement home resulted in happier employees and higher quality of care (Bernmar, 2017). Another closely connected paper is the one by Vogelsang (2017). It is discovered that a gift of time is able to extract high effort levels from participants. However, productivity is not explicitly addressed in these researches yet. More importantly, leisure time has not been studied as a reward that is given *contingent* on performance.

Once the necessary background literature is obtained, an experiment is performed among (ex-) students. The experiment consists of a computerized real effort task that is able to extract participants' productivity levels. Respondents are asked to correctly position sliders on their mobile phone, while being informed on a pre-defined target. Once this target is met, they can either gain money or a comparable reward of leisure time.

The immediate responses do not seem to differ significantly. But after the first 2 minutes have terminated, the monetary incentive wins it from leisure time. More interestingly, it seems that those being incentivized by money keep up their productivity throughout the whole task (even after the target is met). In addition, it is concluded that those with the monetary reward at stake are significantly more distracted during the task compared to those that could win leisure time. However, it appeared that the distraction was not too substantial that it also made them less productive. A finding that is worth to mention.

After the first results are obtained, additional analyses have led to more interesting insights. There are two factors of influence that led to the results as they are currently obtained. The first one is the design of the experiment. The second factor is the audience amongst whom the experiment is spread. The results indicate that once people grow older, the preference for time seems to change as well. Further research in different settings and amongst different workers is able to amplify the valuable insights that are currently obtained.

Chapter 2 will first elaborate on some related literature. Afterwards, the experimental design will be introduced in chapter 3, whereas chapter 4 explains the received knowledge on comparable rewards. Chapter 5 directly compares the monetary incentive to the incentive of leisure time. Next, chapter 6 will perform some additional analyses. At last, chapter 7 discusses all limitations of the experiment and chapter 8 concludes.

2. Background Literature and Hypothesis

This chapter will first deepen into the knowledge that has been obtained so far. Afterwards a labour economical model should shed light on what leisure time means to all of us. By combining both kinds of literature, it is possible to close this chapter by a well-founded hypothesis.

A. Previous Related Literature:

Let me start of by a fairly old paper. In 1929, people worked 49 hours per week instead of the nowadays more common 40 hours. Denison (1962) was the first to claim that fewer working hours will be completely offset in the gain that output remains unchanged. In other words, people can create similar output in fewer working hours. By that time, Denison (1962) presumed that a 10 percent decrease in working hours would lead to a 6 percent reduction in output. A conjecture that is worth to investigate.

By now, many more researchers agree on the fact that productivity might decline under long working days, especially due to the fatigue effect. For instance, Collewet and Sauermann (2017) have observed workers in a call centre. They found that when working hours rise, the average handling time per call increases as well. This signals a lower productivity of workers. Also Pencavel (2015) concluded that the relationship between output and working hours is non-linear. That is, above a certain threshold, more working hours will have a decreasing positive impact on output.

A study performed by Noussair and Stoop (2015) shows that if people have to allocate time between them and a counterparty, people act pro-socially. Also Gino and Mogilner (2014) found that time can be used in order to discourage dishonesty (and in particular cheating). In other words, both papers show that time can positively influence individuals' behaviour.

Taken this to the next level, Pearson (2009) states that under certain conditions, people value time even more than money. Likewise, in Germany's railway industry, it has been proven that workers care about leisure time. When given the choice between a wage increase and an increase in leisure time, the majority of the railway workers preferred to take the extra time off (Drost, 2016). As well, Brown et al. (2013) concluded that people value the gift of time more than a comparable gift of money.

On the other hand, Ellingsen and Johannesson (2009) have shown that in an ultimatum game participants were more willing to accept a loss of time compared to a loss of money. Also Vogelsang (2017) has investigated the preference between money and time. When explicitly asked, most people (approximately 75 percent) prefer to receive a monetary gift, compared to a time gift.

Nevertheless, Vogelsang (2017) did find contradictions to this in actual respondents' behaviour. The study by Vogelsang (2017) is highly in accordance with the aim of this paper. It is discovered that a gift of leisure time causes participants to perform better. Even more interestingly, there is no such

difference found between the gift of a wage increase and the baseline. Thereafter, it can be reasoned that the gift of time might induce more effort compared to the gift of money.

Even though this is highly in accordance with the research question of this study, there is a significant difference. As Vogelsang mentions himself, he studied the "gift" of leisure time, and a gift is "something given voluntarily without any expectations" (Vogelsang, 2017; p. 2). My study will set targets on people and reward them contingent on their performance.

The most prominent difference and advantage that my study obtains, is that it will not solely provide the effort levels but also an assessment of productivity. As said, I will study the difference in productivity between the incentive of money and the incentive of leisure time.

B. The Neoclassical Model of Labour-Leisure Choice:

A framework that is often used by economists is the "Neoclassical Model of Labour-Leisure Choice" (Borjas and Van Ours, 2010). This model is used to investigate labour supply behaviour. What is said, is that a person gains satisfaction from both the consumption of goods, as well as the consumption of leisure time. This can be summarized in the following utility function:

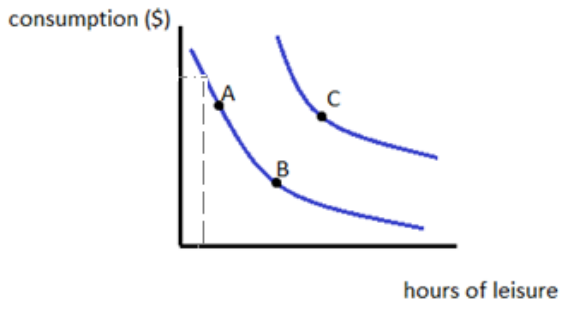
U= f(C, L)

Whereas C is the total monetary value of all goods that the person consumes in a certain period. Likewise, L represents the hours of leisure that this person has consumed during the same defined period of time.

Within this model it deems logical that different kinds of combinations between the consumption of goods and the consumption of leisure time can create a similar utility level. All these combinations together lie on the same indifference curve.

Figure 1: Indifference Curves.

Adapted from *Labor Economics* (p. 28), by Borjas and Van Ours (2010).



In figure 1, point A and B yield the same utility, whereas point C is linked to a higher utility level. However, what is more interesting in light of my research is the slope of the indifference curve. The

slope equals what is called the 'Marginal Rate of Substitution'. This MRS tells us how many additional dollars are demanded in order to compensate for a particular loss in leisure time. Due to the convexity of the curve, we can say that when someone is left with little leisure time, a lot of money is necessary to compensate for this loss.

In mathematical terms, the MRS is equal to: $\frac{\Delta C}{\Delta L} = -\frac{MU_L}{MU_C}$

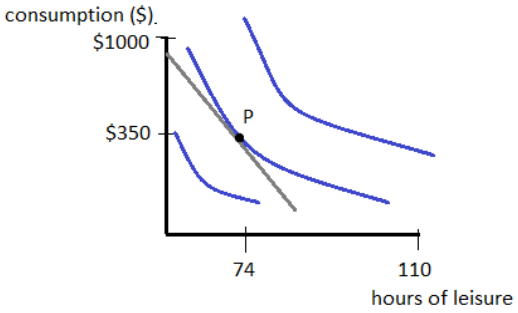
Of course, figure 1 only presents the preferences for one particular worker. In reality, every single worker is dealing with different preferences, i.e. everyone has a different trade-off between consumption and leisure time.

In addition, the consumption possibilities of goods and leisure time are constrained by an individual's time and income. This is what is called the budget constraint. The most obvious feature from this constraint is that when more time is devoted to leisure, less time is left to work. If less time is devoted to work, the income that one obtains from the hourly wage rate also suppresses. In other words, we arrive at the trade-off between income and leisure time. Borjas and Van Ours (2010) state it as follows: "Each hour of leisure consumed has a price, and the price is given by the wage rate" (p. 32).

Once we assume that people aim to maximize their utility, we can combine the budget constraint with the indifference curves and we arrive at the optimal decision for an individual.

Figure 2: The Labour-Leisure Decision.

Adapted from *Labor Economics* (p. 34), by Borjas and Van Ours (2010).



The highest possible indifference curve that is touching the indifference curve provides us with point P. At this optimal solution, the individual consumes 74 hours of leisure, works 36 hours, and consumes $350 on goods on a weekly basis. Still, this optimal point differs per person. Every person values money and time in a different manner. Within this paper, the aim is to investigate the trade-off between income and leisure time. To be more precise: Could leisure time be as strong as an incentive as money?

C. Hypothesis:

Now that the research question and theoretical framework is clear, I will introduce my hypothesis. It is shown that people face a trade-off between time and money. When given the exact choice between both factors, people respond differently.

Some studies have said that people value time more than money (Pearson, 2009; Drost, 2016; Brown et al., 2013). Other studies claim the opposite (Ellingsen and Johannesson, 2009; Vogelsang, 2017).

Since previous literature has proven two opposing preferences, I choose to test the so-called null hypothesis.

- *H_0: Comparable incentives of money and time induce a similar level of productivity.*

3. Experimental Design

A. The Task:

The aim of this paper is to test if leisure time works as affective as the often used monetary incentive. In order to test this, I first need to obtain a measure of effort. When effort is closely linked to output, it is easily possible to move towards productivity.

Several papers have introduced ways to keep track of effort in an online experimental setting (Gächter et al.,2016; Berger and Pope, 2011; Erkal et al., 2011; Gneezy et al., 2003). These ways involve a ball-catching task, pressing buttons on the keyboard in succession, encrypting sentences, and solving mazes respectively. However, within this experiment the 'slider task' is chosen to be performed. This real effort task is developed and first used by Gill and Prowse (2011).

The slider task is a computerized real effort task. In comparison to the previously mentioned tasks to measure effort, this task has some real advantages. First of all, the slider task remains identical throughout every repetition. Therefore this task is relatively easy to explain and understand. In addition, there is no need to test for pre-existing knowledge. Also, according to Gill and Prowse (2011) the task is involved with little noise. This is highly important because it ensures that the amount of correctly positioned sliders (or output) is highly similar to the effort level that is provided by the participant. At last, the designers of the task correctly mention that there is no room for guessing, which makes it again more likely that we are measuring the true productivity level.

Now that we know the advantages of the task, I would like to shortly elaborate on the exact design. Simply said, the task is to position the slider on the demanded position. Each correctly positioned slider is worth one point to the participant.

Figure 3: An Illustration of the Slider Task.

Every slider can be repositioned an unlimited amount of times and the actual position is shown on the right hand side of the slider. The trick is to correctly position it as soon as possible because the total amount of time for the task is limited. The more sliders are correctly positioned, the higher the participant's score. This score then represents his/her effort level. Even more importantly, the amount of sliders per minute is interpreted to be the productivity level.

The task has some specific features:

- The required positions differ per question, such that all sliders are equally challenging to perform correctly.
- However, the order of the questions remains similar for every single participant. This is done in order to ease the comparison between subjects.
- At last, everyone needed to perform the slider task on his/her mobile phone. In this way most respondents are gathered and comparison remains possible.

B. The Baseline and Treatment Groups:

The experiment deals with two issues that need to be solved before participants could be placed into treatments. The first issue concerns how much people value time. This since we are in need of comparable incentives of money and leisure time. The second issue is how to determine the amount of sliders that is reachable within a certain time period. This one is required in order to set reasonable targets.

To solve the latter issue, Gill and Prowse (2011) provide some first insights. They led people perform 48 sliders within 120 seconds, whereby it was found that 41 sliders was the maximum. By assuming this maximum effort level, we can conclude that one correctly positioned slider takes approximately 3 seconds. Vogelsang (2017) shows that, on average, people were able to correctly position approximately 50 sliders in 10 minutes. This amount of correctly positioned sliders could enlarge to 62 in 10 minutes as soon as participants are *given* 25 minutes of leisure time. Again assuming the largest effort level, one correctly positioned slider takes almost 10 seconds.

And even more studies have used the slider task to elicit effort. For example, Faravelli et al. (2017) quotes that "using a macOS operating system, a mouse and a high definition screen (27-inch widescreen monitor with a 5120 by 2880 pixel resolution), each set of 3 slider tasks takes approximately 15 seconds" (p. 8). This finding is equal to 5 seconds per correctly positioned slider. What can be concluded from these three diverging findings is that, based on the design, a single slider task could take many differing time periods.

The first issue becomes even harder to address. Only very few studies have addressed the monetary value of time. More importantly, this value is not expected to move linearly. Therefore it is not possible to infer anything from previous studies. To get the most reliable answers for my current experiment, I took advantage of the baseline group to set realistic targets and comparable rewards. A number of 31 participants are asked to perform the slider task (in the way I designed it) before the so-called "treatment surveys" were spread. This group has formed the baseline.

- *Baseline* (N=31): They first get informed on the slider task and no target is set for this particular group. Rather the task is performed for 8 minutes and afterwards questions are asked to extract additional information.

The baseline group provided information on the average productivity level of participants. This shows how many sliders are possible within a certain time period and is thereafter used to set a realistic target on both treatment groups. Additionally, the baseline group is used to extract the average monetary value that is subscribed to a certain amount of leisure time. Appendix A shows the layout of the survey that is used for the baseline group.

The experiment is performed in Qualtrics, an online survey platform whereby participants have to perform the slider task. From April 26 till May 14, the baseline group is collected. Afterwards, from the May 15 till May 28, the Qualtrics survey randomly selected participants into one of the two treatment groups (see appendix B for layouts). All participants are either students or recently graduated (within the past three years).

- *Treatment Money* (N= 42): The *Money* group performs the exact same task as the baseline group, however a pre-defined target is set. That is, participants get informed on the fact that if they are able to collect 27 points, a 20% chance exist to win €20[1]. After the task has been performed for 8 minutes, again a small amount of additional questions need to be answered.

- *Treatment Time* (N= 31): Likewise, the *Time* group performs the exact same task while being informed on the pre-defined target of 27 points. However, in this case, participants that reach the target are able to move directly to the few closing questions. This makes that they gain

[1] These numbers are retrieved from the baseline of this experiment (see chapter 4).

some leisure time. The duration of the task drops from 8 minutes towards any time within the target is reached.

C. Control Variables:

As shortly mentioned, the experiment ends with some additional questions (see Appendix A for details). Most of these questions have the purpose to control for individual specific characteristics. These characteristics consist of:

- Gender
- Age
- Highest obtained educational level
- Distraction / Leisure time outside of the slider task
- The kind of mobile phone

The "distraction" question is here since the aim is to find out if people in the time treatment are less lured into distraction by the environment compared to the money treatment. As well, it might be wise to control for the kind of mobile phone since every device has a different screen size and this might affect the ease of the slider task.

A last, participants are asked to show their preferences between money and leisure time. The design of this question is a choice list and participants are asked to pick that row where they first want to move to the right option.

4. Intermediate Results

The baseline data is meant to solve two appointed issues, namely (a) setting targets and (b) eliciting comparable incentives. The baseline consists of 31 respondents with the largest percentage being female (70.97 percent versus 29.03 percent). The average respondent is 23 years old with a standard deviation of 1.64. The majority is university educated (83.87 percent). The rest either obtains high school education (6.45 percent) or is higher professionally educated according to the Dutch educational system (9.68 percent).

In addition, there is one more variable worth to address: the distraction. Most respondents claim not to be distracted at all throughout the experiment (48.39 percent). The second largest group admit to be distracted once (35.48 percent), and only one respondent was distracted more than 5 times while performing the slider task (3.23 percent).

A. Setting Targets:

The maximum score that has been achieved by one respondent equals 101 and the minimum score is 28. Both extremes are close to other responses and not interpreted as outliers, therefore I continue to work with mean scores. The total mean score equals 69.00 per 8 minutes. In other words, the productivity of an average respondent is 8.63 sliders per minute. Since I was not yet rewarding contingent on performance, I assume this productivity to rise once incentives are in place. For this reason, the average of 9 sliders per minute is used to set targets on both treatment groups. The time treatment is defined to be rewarded by a leisure time of 5 minutes on average. With a time period of 3 minutes, assuming a productivity of 9 sliders per minute, it is decided that 27 sliders are possible for an average respondent. Therefore, the target is set: 27 sliders.

In order to determine the actual target, it should be ensured that no significant learning is in place throughout the experiment. The slider task needs to be performed for 8 minutes. However, these 8 minutes are subdivided into 2 minutes per page such that four samples are collected per individual. A within subject test should be conclusive in whether or not learning is in place.

The target is designed to be accomplished after 3 minutes and therefore most likely to be reached within the second stage. For this reason, a Wilcoxon signed-rank test is performed that compares the second stage against the first stage of 2 minutes. A p-value of 0.1841 concludes that there is no significant immediate learning of individuals at 10 percent significance level. In conclusion, a target of 27 sliders seems most reasonable and is set on both treatment groups.

B. Comparable Incentives:

The second issue that the baseline aimed to solve is to find comparable incentives for both treatment groups. After investigating the preferences of respondents, some outliers have been observed. There are 4 respondents that claim to prefer a zero percent chance to win €20 over 5 minutes of leisure time. On the other hand, there are also 3 respondents that prefer the 5 minutes of spare time over the certain amount of €20.

These two diverging findings once more show that each person has different preferences and attach different values to time and money. I choose for these outliers to have only little impact and thus decided to make use of the median preference. This median represents an indifference between a 20 percent chance of winning €20 and leisure time of 5 minutes. For this reason, those placed into the money treatment obtain a 20 percent chance of winning €20 as soon as they have correctly positioned 27 sliders.

5. Comparing Two Kinds of Incentives

Table 1 shows the descriptive statistics of both treatment groups, plus baseline. Treatment money consists of 42 respondents, while treatment time is slightly smaller with 31 respondents. This is in order to enlarge the significance of tests because treatment money is dealing with a larger (expected) variance in responses (Faul et al., 2007). Both groups consist for the largest part of women and university graduates (just like the baseline).

Since we are in possession of quite similar baseline and treatment groups, it is now possible to do some comparison analyses.

Table 1: Descriptive Statistics.

(Mean scores are represented)

	(1) **Baseline**	(2) **Time**	(3) **Money**
N	31	31	42
Percentage female	70.97%	67.74%	69.05%
Age (st dev)	23 (1.64)	23 (2.26)	23 (1.96)
University graduates	83.87%	64.56%	66.67%
Working time on task (min)	8	4.24	8
Total points	69.00	38.45	83.50
Productivity level (sliders per min)	8.63	9.11	10.44
Mistakes	2.77	0.97	2.50

A. The Productivity Levels:

Let me start off with the main research question of this paper. Which kind of incentive provokes the highest productivity level?

First of all, it should be noted that the money treatment includes one person able to position all possible sliders to the demanded position. This is problematic in the sense that we are uncertain about his/her true productivity level. However, since this has occurred only once out of 104 observations, I assume the limited finding of this individual not to be much of an influence.

Table 2 already reveals the main solution. That is, both incentives of money and leisure time do *not* induce a comparable level of productivity. A non-parametric Mann-Whitney U test provides a p-value of 0.0470. This means that there is enough evidence to reject the null hypothesis at 5 percent

significance level. But which incentive is most effective? Those being incentivised with money are more productive compared to those that gain leisure time. To be more precise, the probability of an individual in treatment money having a higher true value than an individual in treatment time is 63.7 percent (Conroy, 2012).

Two similar Mann-Whitney U tests that compare our treatment groups against the baseline provide p-values of 0.0000 for both treatments. This means that both incentives of money and leisure time lead to significant different productivity levels as compared to the situation with no incentive in place. Logically, either incentive induces higher productivity as compared to the baseline. It is also possible to compare our three groups at once. A Kruskal-Wallis test with a p-value of 0.0332 rejects that the populations are the same at 5 percent significance level.

A final note should be made on the descriptive statistics found in table 1. As one can see, the reward of leisure time is, on average, equal to (8 - 4.24 =) 3.76 minutes. What is said, is that a reward of 20 percent chance on €20 is similar to a reward of 5 minutes of leisure time. However, the results claim that 21 out of 31 respondents in treatment time dealt with a reward less than 5 minutes of leisure time. How come? Simply said, the reward that people in treatment time obtain, is the one that they make on their own. According to the baseline analysis, the target should be reachable in 3 minutes, on average. Moreover, the fact that 10 out of 31 respondents did manage to reach the target within 3 minutes shows that it is possible. If I continue to build on that, I can assume that those being placed into the time treatment are not always willing to work hard enough (be more productive) to reach their target in time. In other words, the significant differences in productivity as found in this section *are* actually present.

Table 2: Productivity.

Productivity level (sliders per min)	(1) **Baseline**	(2) **Time**	(3) **Money**
Min:	3.50	5.12	3.50
Max:	12.63	15.17	20.00
Mean (st dev):	8.63 (2.57)	9.11 (2.48)	10.44 (3.39)
Median:	8	8.98	10.38

B. Leisure Time Outside of the Task:

During the task, participants could be distracted and even decide to take some time off in between questions. The survey led people fill in how many times they got distracted during the slider task. This

(categorical) variable tells us how much (additional) leisure time people took while performing the task.

According to the study by Vogelsang (2017), people that gain the gift of leisure time show a larger reduction of "on-the-job leisure consumption", compared to those with a gift of money. More importantly, it is claimed that this reduction has led to higher performance. However, what happens if different kinds of incentives are put *contingent* on performance?

My experiment has found some significant differences between treatments as well. A Mann-Whitney U test with a p-value of 0.0927 suggests that both incentives do *not* lead to similar amounts of distraction at a 10 percent significance level. After adding the porder option, it is found that the probability of an individual in treatment money having a higher true value of distraction than an individual in treatment time is 59.8 percent (Conroy, 2012). In addition, a Kruskal-Wallis test claims that there is too little evidence to reject that there are significant differences between both treatments and the baseline (p-value of 0.1244).

Most importantly, those being incentivized by money are more often distracted, compared to those with a reward of leisure time. Even though distraction is negatively correlated with productivity (-0.0923, p-value: 0.4375), they are still significantly more productive. This is a finding that contradicts the research of Vogelsang (2017).

C. The Number of Mistakes:

While performing the slider task, people can easily drag the slider to a wrong position. Mistakes are made in both treatment groups and in the baseline, but are there any significant differences? A Mann-Whitney U test between our treatment groups claims that there is too little evidence to reject the null-hypothesis (p-value of 0.1159). In other words, both incentives cause participants to make similar amount of mistakes.

By comparing treatment money to the baseline, a p-value of 0.5495 shows that giving money or no incentive does *not* lead to a differing amount of mistakes. However, a similar Mann-Whitney U test shows a significant difference in the amount of mistakes between those that gain leisure time and the baseline. To be precise, once the reward of time comes at stake, people are less likely to make mistakes, compared to the situation without incentive (p-value of 0.0388). Again it is also possible to compare all three groups simultaneously. A Kruskal-Wallis test is unable to reject the null-hypothesis that populations are similar (p-value of 0.1055).

D. True Preferences:

The survey contained two questions to gather something on individual preferences between time and money. In some instances, the preference as derived from the choice list was highly contradicting to the direct scaling preference of the same individual (question 6 and 7 from appendix A+B). For example, several have picked the first row of the choice list but also prefer the money option over leisure time (21 out of 73 observations). This means that these individuals prefer 5 minutes over the certain amount of €20, while simultaneously preferring the monetary lottery over 5 minutes of leisure time. This is inconsistent and therefore I presume that some have misunderstood the choice list question[2]. This paper continues to analyse the direct scaling question as this one can be seen as more reliable.

Table 3 shows that the majority of 63.01 percent prefers to have a 20 percent chance to win €20. On the other hand, 36.99 percent directly prefers 5 minutes of leisure time. This relative small inequality is an interesting finding amongst (ex-) students. Especially since many researchers have recognized that students are generally poorer and therefore more in need of money (Fehr and List, 2004; Druckman and Kam, 2011).

Table 3: True Preference Count.

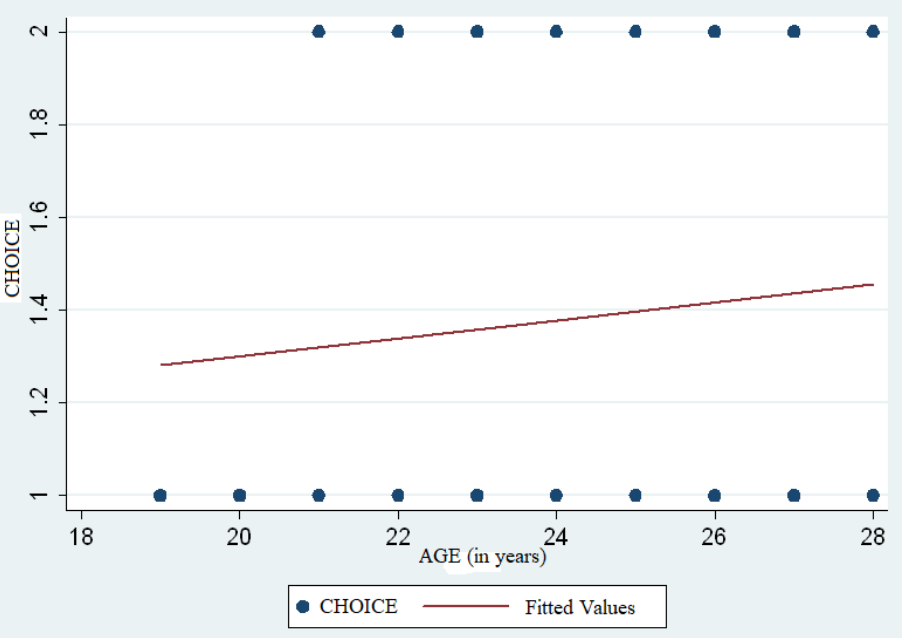
CHOICE	Frequency	Percentage
1 (The money option)	46	63.01%
2 (The time option)	27	36.99%
Total	73	100%

As someone grows older, I assume this person to become wealthier. Those who are older are more likely graduated and working on their first full-time job. For this reason, the relation between AGE and CHOICE is sketched, whereby CHOICE is the variable that equals 1 if a person prefers the monetary lottery and 2 if a person rather has spare time. Figure 4 shows that there is a slightly positive relationship between the two. So if someone gets older, he/she is more likely to prefer spare time of 5 minutes over the lottery of €20.

[2] This finding confirms that picking the median preference in analysing the baseline was the most optimal thing to do.

Figure 4: The Relationship between Age and Preference.

(Adapted from STATA)



E. Checking For Robustness and Power:

In order to close the first analyses, I would like to debate on the power and robustness of the previously found results. Firstly, an OLS regression could mention more about the robustness of results. While performing this regression, it is possible to control for personal characteristics, screen size, distraction, the choice of preference, and even the number of mistakes that are made.

$$PROD_i = \beta_0 + \beta_1 * Treatment_i + \beta_2 * Gender_i + \beta_3 * AGE_i + \beta_4 * Educ_i + \beta_5 * Screensize_i + \beta_6 * Distraction_i + \beta_7 * CHOICE_i + \beta_8 * Mistakes_i + \varepsilon_i$$

Table 4 shows the corresponding results. On average, being incentivized with money increases the productivity level by 1.80 sliders per minutes, as compared to being incentivized with leisure time, ceteris paribus. This finding is statistically significant at 5 percent significance level and this strengthens the results as they are found previously. Moreover, gender, screen sizes and even mistakes seem to have significant influence on productivity as well.

Table 4: OLS Regression

Independent Variables	PROD
Treatment	1.80* (0.77)
Gender	-2.01* (0.80)
AGE	-0.27 (0.17)
Educ	0.10 (0.25)
Screensize	-0.29** (0.10)
Distraction	-0.12 (0.43)
CHOICE	-0.25 (0.79)
Mistakes	-0.21* (0.08)
Number of obs	73
R-squared	0.29
Prod > F	0.0045**
**, **, represent statistical significance levels*	*at 5 percent and 1 percent, respectively*

Secondly, power analyses are able to provide actual power and determine the necessary group sizes. Appendix C elaborates on the calculations. It is found that the current power of my experiment equals 0.43. This means that there was a 43 percent chance to found the significant results as they are retrieved now. Since this is not close to 100 percent, one could reason that the results might not be very likely to correspond with "real" findings. Still, it makes sense to find that (ex-) students respond more strongly to a monetary reward, as they do to a reward of time. As is said, (ex-) students are generally poorer and obtain a greater desire for money.

6. Additional Analyses

It is found that productivity levels are higher once a reward of money is at stake, compared to the reward of leisure time. However, did the participants get the reward of their preference? How did

productivity progresses over the different stages of the task? This chapter will provide answers to these questions. As well, the design of the task will be discussed.

A. Getting What You Prefer:

There are participants being in treatment time while they would rather obtain the reward of money, and vice versa. The discrepancy between preference and the actual reward might cause people to be less productive than they could be. The presumption here is that people are most productive when they get the reward of their choice.

This additional analysis selects participants of whom their reward is in accordance with their preference. We end up with 30 observations in treatment money and 15 in treatment time. A non-parametric Mann-Whitney U test once more declares that there are significant differences in productivity levels amongst both treatments. More interestingly, the finding is even more significant as before (p-value of 0.0311 versus 0.0470). Those that prefer to receive the monetary reward and *are* getting this lottery are more productive, compared to those who get their preferred option of leisure time.

What does this mean? The significant finding of differing productivity levels between treatments is not due to the fact that people in treatment time are being rewarded with the "wrong" incentive. After controlling for true preferences, it is still found that people who are incentivized with money are more productive, compared to those being incentivised with spare time.

B. The Effectiveness of Incentives Over Time:

Another question one could ask is: Which incentive leads to the highest instant productivity?[3] A Mann-Whitney U test shows that there are *no* significant differences in first-stage productivity amongst our two treatments (p-value of 0.9643). This could either mean that there is simply too little evidence to reject the null-hypothesis or that both incentives induce similar instant productivity.

A similar test on the second stage shows that there *are* significant productivity differences between treatments (p-value of 0.0170). In order to perform this test, 4 responses of treatment time were deleted since they had already reached their target in the first period. A Mann-Whitney U test, plus porder option, shows that the probability of an individual in treatment money having a higher true value than an individual in treatment time is 67.1 percent for the second period (Conroy, 2012). This shows that as soon as we start looking at a longer time horizon, the monetary incentive seems to work best. For this reason, it deems logical to investigate all stages of the slider task.

[3] Recall that we obtain four samples per individual, since the task is subdivided into four stages of 2 minutes.

Figure 5: The Productivity-Patterns.

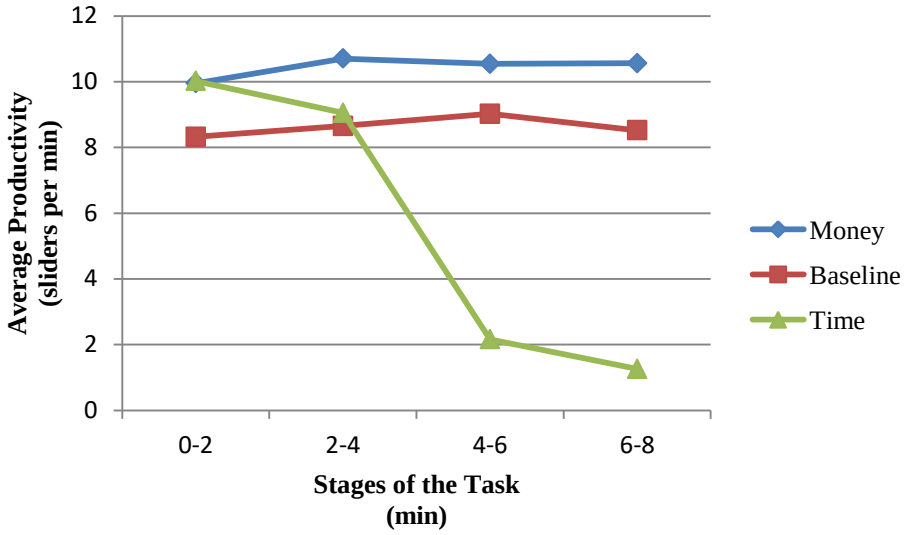


Figure 5 presents the patterns of productivity over all four stages for each treatment and baseline group. The figure clearly shows that there are (indeed) no instant differences between the productivity levels of our two treatment groups. However, after the first stage has terminated, differences become visible and significant. After the second stage, most respondents of treatment time have reached their target and therefore a large drop in productivity is visible.

It does not make sense to compare both incentives in the last two stages. Nevertheless, it can be investigated if there is a (significant) pattern across treatment money. A Jonckheere-Terpstra test with a p-value of 0.7686 suggests there is not[4]. Still, a Wilcoxon signed-rank test amongst first and last stage of the money treatment can create more insights. A p-value of 0.0313 concludes that there is enough evidence to reject the null hypothesis at 5 percent significance level, i.e. both stages differ in distribution. But which stage deals with the highest productivity?

By looking at the average productivity levels, a valuable insight became apparent. It seems that those being incentivised by money do not lose much in productivity over all four periods. More importantly, the productivity levels of the final stage outperform those of the first ($\mu 4=10.56 > \mu 1=9.95$).

C. Zero Productivity Takes Time Too:

While setting a target, I was short-sighted. The participants placed into treatment time where rewarded with leisure time. As soon as they have reached their target, they were allowed to skip all following slider questions and move towards the final few personal questions of the survey. However, the time

[4] One should note that a Jonckheere-Terpstra test is meant for between-subject designs so this p-value is the highest possible value.

that it takes to scroll down the sliders and click continue was not recognized. On average, it took participants 8.12 seconds to do so.

After taking these individual seconds into account, a new analysis has been performed. This modified Mann-Whitney U test has led to an interesting finding. There are no longer significant differences in productivity between our two treatment groups (p-value of 0.4152). This suggests that with a different design of the task, both incentives of money and leisure time could induce similar productivity levels. Taking it differently, this data might be too small to reject the null hypothesis of this newly composed issue.

There are two different manners to interpret this new finding. Firstly, if you decide to exclude those minutes since it is accompanied with zero production, you could conclude that leisure time is as effective as money. Or at least, there is too little evidence to reject that it is not.

Secondly, one could say that the time to continue is part of the job (which it is at the moment). Therefore, it can be claimed that the significant results as found previously are still true. In other words, amongst (ex-) students it seems that the monetary incentive wins it over an incentive of leisure time. An argument that is in accordance with this last opinion is that the minimum time one participant took to continue is 2.81 seconds. Since it is still part of the job, you can also click continue in the fastest way possible. If everyone would have done so, it can be said that both treatments do not differ significantly in distributions (p-value of 0.2457). This means that even in the current design and with the current data, it could have been found that two different incentives elicit similar productivity levels. However, it is the case that leisure time elicits a smaller productivity level and this is also reflected into that time that people take to close the task.

Of course, this kind of reasoning can be interpreted as speculating. There are arguments in favour of both opinions. The biggest conclusion to draw here is that the design of the experiment is not flawless. This and more limitations will be discussed next.

7. Discussion

It seems like the design of the task is the biggest limitation so far, as recognized in the previous section. However, there is more to mention about the set-up of the experiment, and this chapter will do so.

First, a limitation of my experiment is that I kind of pushed the baseline people into a direction while asking for their preference. The choice list that is used immediately sets the amount of 5 minutes fixed against a lottery of €20. Another lottery with a different amount of money might affect preferences. In addition, it is well-known that people are often biased to pick the middle option (Beauchamp et al.,

2015; Andersen et al., 2006). On the other hand, asking participants a direct question makes it way more difficult for them to come up with an answer. I needed a starting point and picked this one.

Second, it should be noted that this experiment is mostly spread among students. Students are not the best representatives. One could imagine that students' preferences might deviate from adults with a secured income level. Figure 4 already shows that when people get older, the time option is likely to become more preferable. The question remains whether or not this will be reflected into the productivity levels of people. A similar research performed among people aged 30-50 could be highly valuable.

Third, more observations are always better. The between-subject comparison makes it hard to gather a sufficient amount of observations. Still, the amount is reached whereby significant results are found. This is certainly valuable but power calculations provide us with the insight what amount of observations should lead to an ensured 5 percent significance level. This amount equals 83 participants placed into treatment time and 89 into treatment money (see appendix C for details). For now, this has not been possible and can therefore be interpret as a limitation.

Fourth, there could be a possible misunderstanding of the task. This possibility exists for both treatment groups. In treatment time, it could be that participants wrongly concluded they had to reach their target every 2 minute period in sequence. In treatment money, the same mistake could have been made. In addition, people might misunderstood the fact that collecting points beyond target does not enlarge their chances to win the lottery. The survey did include an audit question on whether participants understood the instructions and if not, they were unable to continue. For this reason, I assume that at least the majority of the participants understood their job.

Fifth, some participants have failed to meet their target. As a matter in fact 7 respondents out of 104 did not meet the target, of whom 4 did reach 26 points (which is one point below target). Noticeably, all who failed their target, where in time treatment and it probably had to do with a wrong collection of numbers. In other words, they might believed that their target was met but some mistakes did not let that happen. Since these 7 respondents were still operating with the right target in mind, I assume their responses still to be valuable.

Sixth, the risk attitude of participants matters. Unfortunately, a certain amount of money as reward would be too expensive. Nevertheless, both rewards are stated as comparable amongst my participants and that is why they are still highly useful. This is a big advantage in comparison to the paper by Vogelsang (2017). To me, it seems like he never declared why his gifts of money and time could be seen as similarly valuable to his participants. As long as this is not clear, it might be that people over- (or under-) appreciate the gift of spare time in comparison with the monetary gift. This could affect results.

Seventh, my experiment as it is composed now is rather small. A time frame of 5 minutes is so small that it is hard for people to form a preference. As soon as one enlarges this experiment to comparing 1 hour of leisure time with a 20% chance to win €50 for example, results might become quite different. Still, the purpose was and is to provide some insights that further research can build upon.

At last, throughout the paper bonus schemes are mentioned. This experiment does not provide a standard (show-up or hourly) wage and this might affect the results substantially as well. Once someone already has money in its pocket, what kind of additional incentive works best? This is a research question that future work can build upon.

8. Conclusions

This paper has shed light on a non-monetary incentive: Leisure time. The main question was whether or not leisure time as a reward would lead to similar productivity levels as the often used monetary reward. In other words, is the heading of the Dutch news article right? Could there be a substitute for monetary bonus schemes?

Using time as a bonus is a loss in the sense that people leave work earlier and thus no production occurs during that time. On the other hand, fewer working hours also lead to less wage compensation for firms. The real profitability therefore depends on the productivity level that is present in the shortened working period. This paper suggests that, at least amongst (ex-) students, the monetary incentive is most effective in eliciting high productivity.

The immediate responses do not seem to differ significantly. But after the first 2 minutes have terminated, the monetary incentive leads to significant higher productivity levels as compared to the reward of leisure time. The null hypothesis is rejected. More importantly, the productivity levels in treatment money remain at similar level throughout the whole task. You can say that once workers get closer to their target, motivation drops for those being incentivised by leisure time. On the other hand, when a monetary reward is put on hold, it seems that people keep up their productivity.

What should be noted is that these conclusions solely hold in the currently used experimental design and amongst the current audience of (ex-) students. This paper has already proven that once people get older, the preference for time seems to be rising as well. This can be attributed to the fact that adults most often have a secured income and are less dependent on money. Also it is shown that as soon as the design changes, the results are likely to change as well.

Another finding of this paper is that those who have a monetary reward at stake are more often distracted, compared to those that could win leisure time. In contrary to the statement of Vogelsang (2017) this does not lead to lower performance. Even though they are more lured into distraction, they

still are significantly more productive. A final discovery is that rewarding based on workers' preference seems to make results even more significant. This is something that firms could take into consideration. Rewarding people with money is most effective if they want to be rewarded with it.

Besides, the considerations that firms are dealing with differ per company. If the target is all that you need and nothing more, rewarding with leisure time could be considered. What is found within this study is that instant productivity levels between both incentives do not differ significantly. Logically, this has yet to become apparent in a real working situation (over a longer period of time). However, since rewarding with time saves a lot of money, it might be wise for companies to consider the option. On the other hand, if more production is always preferred over less, and most of the workers are in their early 20s, the monetary reward is determined to work best.

In conclusion, much research is left to be done. Each particular setting might demand a different incentive. However, this paper has provided the first insights to what will happen if leisure time is set as a reward contingent on performance. Moreover, this paper expands the knowledge on how people value time in comparison to money. This is an insight that is useful throughout many more domains, e.g. the value of time in speed tickets, the value of waiting time for a health treatment (Ashenfelter and Greenstone, 2004; Ryan et al., 2001).

References

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383-405.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, *76*(2), 451-469.

Ashenfelter, O., & Greenstone, M. (2004). Using mandated speed limits to measure the value of a statistical life. Journal of political Economy, 112(S1), S226-S267.

Beauchamp, J. P., Benjamin, D. J., Chabris, C. F., & Laibson, D. I. (2015). *Controlling for the Compromise Effect Debiases Estimates of Risk Preference Parameters* (No. w21792). National Bureau of Economic Research.

Berger, J., & Pope, D. (2011). Can losing lead to winning? *Management Science*, 57(5), 817827.

Bernmar, D. (2017, January 6). Ignore the headlines: a six-hour working day is the way forward. *The Guardian*. Accessed on March 24, 2018 via https://www.theguardian.com/commentisfree/2017/jan/06/ignore-headlines-six-hour-working-day-swedish

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, *12*(1), 19-64.

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4-5), 303-345.

Borjas, G. J., & Van Ours, J. C. (2010). *Labor economics* (pp. 346-382). Boston: McGraw-Hill/Irwin.

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, *62*(11), 3085-3099.

Brown, A. L., Meer, J., & Williams, J. F. (2013). *Why do people volunteer? An experimental analysis of preferences for time donations* (No. w19066). National Bureau of Economic Research.

Cadsby, C. B., Song, F., & Tapon, F. (2010). Are you paying your employees to cheat? An experimental investigation. *The BE Journal of Economic Analysis & Policy*, *10*(1).

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, *19*(1-3), 7-42.

Collewet, M., & Sauermann, J. (2017). Working hours and productivity. *Labour Economics*, *47*, 96-106.

Conroy, R. M. (2012). What hypotheses do" nonparametric" two-group tests actually test?. *Stata Journal*, *12*(2), 182.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, *125*(6), 627.

Denison, E. F. (1962). United States economic growth. *The Journal of Business*, *35*(2), 109-121.

Drost, F. (2016, December 14). Raises vs Time Off in New Rail Worker Labor Deal. *Handelsblatt Global.* Accessed on March 24, 2018 via https://global.handelsblatt.com/companies/rail-deal-signals-move-toward-work-life-balance-660959

Druckman, J. N., & Kam, C. D. (2011). Students as experimental participants. *Cambridge handbook of experimental political science*, *1*, 41-57.

Ellingsen, T., & Johannesson, M. (2009). Time is not money. *Journal of Economic Behavior & Organization*, *72*(1), 96-102.

Erkal, N., Gangadharan, L., & Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, *101*(7), 3330-48. .

Faravelli, M., Kalayci, K., & Pimienta, C. (2017). Costly Voting: A Large-scale Real Effort Experiment.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fehr, E., & List, J. A. (2004). The hidden costs and returns of incentives—trust and trustworthiness among CEOs. *Journal of the European Economic Association*, *2*(5), 743-771.

Fisher, J. (2015). *Strategic reward and recognition : Improving employee performance through non-monetary incentives*. Philadelphia: Kogan Page.

Gächter, S., Huang, L., & Sefton, M. (2016). Combining "real effort" with induced effort costs: The ball-catching task. *Experimental Economics : A Journal of the Economic Science Association*, *19*(4), 687-712. doi:10.1007/s10683-015-9465-9

Gill, D., & Prowse, V. (2011). A novel computerized real effort task based on sliders

Gill, D., Prowse, V., & Vlassopoulos, M. (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior & Organization*, *96*, 120-134.

Gino, F., & Mogilner, C. (2014). Time, money, and morality. *Psychological Science*, 25(2), 414-421.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049-1074.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 25(4), 191-209.

List, J. A., Sadoff, S., & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4), 439.

Locke, E. A., Feren, D. B., McCaleb, V. M., Shaw, K. N., & Denny, A. T. (1980). The relative effectiveness of four methods of motivating employee performance. *Changes in working life*, *363*(1), 388.

Noussair, C. N., & Stoop, J. (2015). Time as a medium of reward in three social preference experiments. *Experimental Economics*, 18(3), 442-456.

Pearson, J. L. (2009). Time Versus Money: What Are They Worth?. *Unpublished Manuscript*.

Pencavel, J. (2015). The productivity of working hours. *The Economic Journal*, *125*(589), 2052-2076.

Rajagopal, P., & Rha, J. Y. (2009). The mental accounting of time. *Journal of Economic Psychology*, *30*(5), 772-781.

Ryan, M., Bate, A., Eastmond, C. J., & Ludbrook, A. (2001). Use of discrete choice experiments to elicit preferences. *BMJ Quality & Safety*, 10(suppl 1), i55-i60.

Twenge, J. M. (2010). A review of the empirical evidence on generational differences in work attitudes. *Journal of Business and Psychology*, *25*(2), 201-210.

Vogelsang, T. (2017). Time is (Not) Money-Incentive Effects of Granting Leisure Time.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health economics*, 17(12), 1329-1344.

Appendix A – Introduction for the Baseline Group

This appendix shows the introduction that is used for the baseline group. As well, the additional questions that each group received are listed below.

- **Baseline:**

Dear participant,

Thanks a lot for participating in my (short) experiment. The experiment is divided into two blocks.

The first block is running for 8 minutes. You need to perform a so-called "Slider Task". This means that you need to simply drag the slider to the desired position. For each correctly positioned slider you gain 1 point. The goal is to collect as many points as possible within the 8 minutes time period.

Once this task is performed, you move towards the second block whereby you are left to answer some short questions about yourself. This will take 1 minute at most.

You can switch between Whatsapp, Facebook, etc in between, but please finish the survey ☺

Thanks and good luck!

The first page runs for 2 minutes and afterwards you have 3 more pages like that..

- *The Additional Questions*:

1. Please state your gender

2. Please state your age in years

3. What is your highest obtained educational level?

A. VMBO B. HAVO C. VWO D. MBO E. HBO
F. WO G. Other:

4: How often were you distracted while performing the slider task? You can think of answering a Whatsapp message, switching to Facebook, having a small conversation, etc.

A. Zero times B. Only 1 time C. 2-5 times D. More often

5: What kind of mobile phone do you have (while participating in this short experiment)?
(Multiple choice list of 15 most popular mobile phones + the open option)

6: Below you have two choices per row. Pick that row where you would move from left to right. That is the one where you decide to choose the right option for the first time.

100% to win 20€	5 minutes of leisure time to spend however you like
90% to win 20€	5 minutes of leisure time to spend however you like
80% to win 20€	5 minutes of leisure time to spend however you like
70% to win 20€	5 minutes of leisure time to spend however you like
60% to win 20€	5 minutes of leisure time to spend however you like
50% to win 20€	5 minutes of leisure time to spend however you like
40% to win 20€	5 minutes of leisure time to spend however you like
30% to win 20€	5 minutes of leisure time to spend however you like
20% to win 20€	5 minutes of leisure time to spend however you like
10% to win 20€	5 minutes of leisure time to spend however you like
0% to win 20€	5 minutes of leisure time to spend however you like

Appendix B – Introductions for the two Treatment Groups

This appendix shows the introductions that both the Money treatment and Time treatment retrieve. In the second part one additional question is added to infer direct preference.

- **Treatment Money**:

Dear participant,

Thanks a lot for participating in my (short) experiment. The experiment is divided into two blocks.

The first block is running for 8 minutes. You need to perform a so-called "Slider Task". This means that you need to simply drag the slider to the desired position.

1) For each correctly positioned slider you gain 1 point.
2) You are allowed to keep track of your own score, but I will add it up afterwards as well.
3) **Note:** If you are able to collect 27 points, you gain a 20% chance to win €20. (You are asked to fill in your email address so that I can inform you afterwards.)
4) If your score grows beyond 27 points, the chance to win does not grow alike.

A final note: the task is divided into 4 pages of 2 minutes, and the pages switch automatically after 2 minutes. If you would like to keep track of your score, you should remember the score that you got on the previous page and continue to count from that same number in the next page.

Once the slider task is performed, you move towards the second block whereby you are left to answer some short questions about yourself. This will take 1 minute at most.

You can switch between Whatsapp, Facebook, etc in between, but please finish the survey :)

Thanks and good luck!

- *Additional Questions*:

- Question 1-6 as mentioned in Appendix A.
- Additional question 7: When given the choice, which option would you rather have? Select your preference.
  A. a 20% to win €20 B. 5 minutes of leisure time
- Additional question 8: Please state your email address.

- **Treatment Time**:

Dear participant,

Thanks a lot for participating in my (short) experiment. The experiment is divided into two blocks.

In the worst case, the first block is running for 8 minutes. You need to perform a so-called “Slider Task”. This means that you need to simply drag the slider to the desired position.

1) For each correctly positioned slider you gain 1 point.
2) Please **note**: you have to keep track of your own score because once you reached the target of **27 points**, you are allowed to skip all following sliders.
3) You can count your score by adding one point for each correctly positioned slider.
4) Once you've collected 27 points, you can skip all sliders and **click next** on the bottom of the page until you reach some personal questions (gender, age, etc).

A final note: the task is divided into 4 pages of 2 minutes, and the pages switch automatically after 2 minutes. So keep in mind the amount of sliders of the previous page and continue to count from that same number in the next page.

The final few personal questions will take you 1 minute at most.

Please keep track of the score in an **honest** way, otherwise your participation is useless to me.

You can switch between Whatsapp, Facebook, etc in between, but please finish the survey :)

Thanks and good luck!

- *Additional questions***:**

- Question 1-6 as mentioned in Appendix A.
- Additional question 7: When given the choice, which option would you rather have? Select your preference.
  A. a 20% to win €20 B. 5 minutes of leisure time

Appendix C - Power Calculations

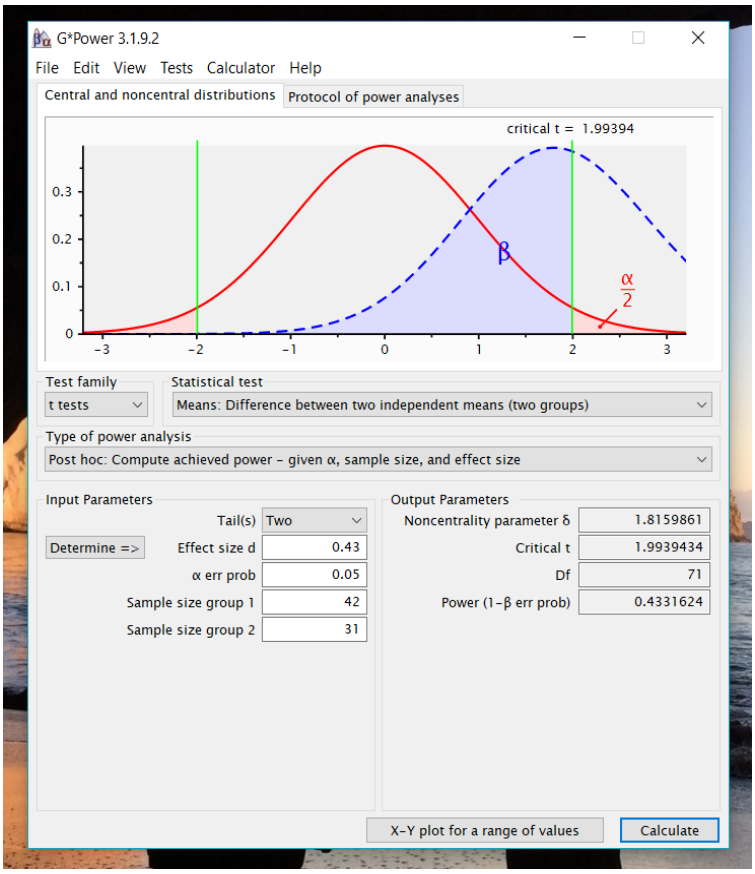
By use of G*Power it is possible to determine the 'post hoc achieved power'. In order to do so, the group sizes as they are collected should be filled in. In addition, the effect size needs to be defined. The effect size equals "the minimum degree of violation of H0" (Faul et al., 2007; p. 176). This is then calculated by use of the mean productivity levels and the common standard deviation:

$$\frac{|\mu 1-\mu 2|}{\sigma} = \frac{|9.11-10.44|}{3.09} = 0.43.$$

Once everything is filled in, G*Power has concluded the power to be equal to 0.43. Since this number is not very close to 1, it cannot be said that the results are highly certain.

Figure C1: Post Hoc Power.

Performed by G*Power.



In addition, the G*Power program has the possibility to establish the required sample size. So how many observations are actually demanded in order to get the secured 5 percent significance level? This is an 'a priori calculation'. In order to do so, the allocation ratio needs to be defined. List et al. (2011) provided the following formula in order to calculate the allocation ratio: $\frac{N2}{N1} = \sqrt{\frac{c1}{c2}} * \frac{\sigma 2}{\sigma 1}$.

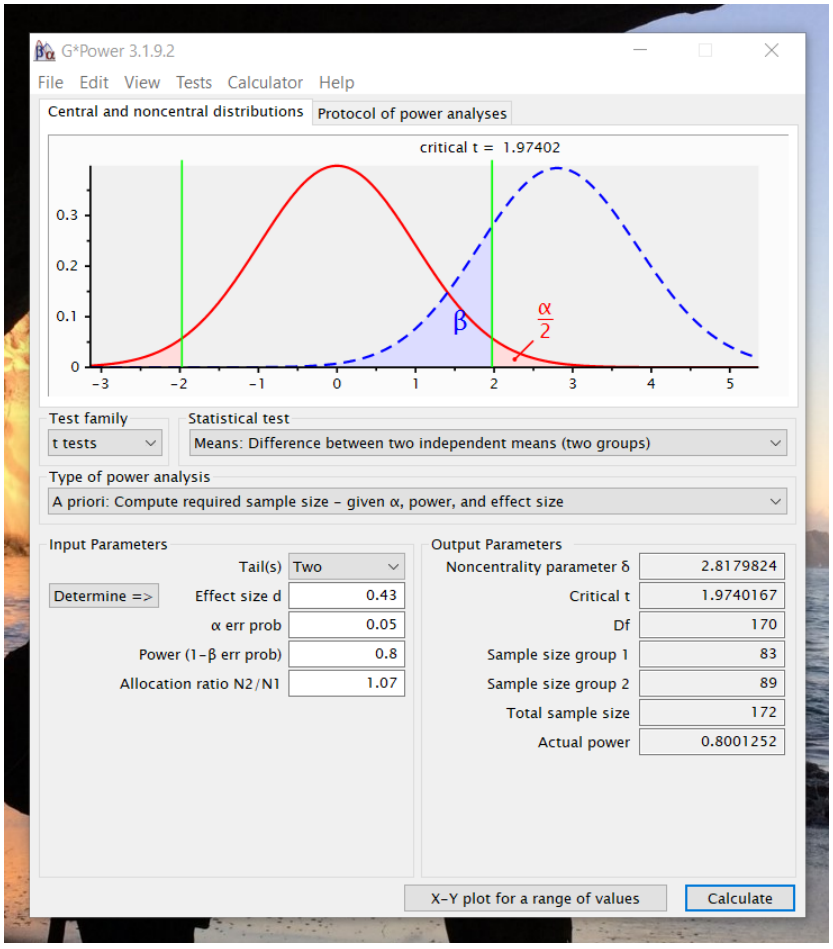
The ratio depends on the treatment variances, as well as the costs per treatment. The treatment variances are given since the data is already collected (σ2=11.46 and σ1=6.16). In my experiment,

treatment time is involved with zero costs, however this will not work in the formula. For this reason, I calculated the ratio as if participants received a fixed show-up fee to participate in either one of the two treatments.

For instance, everybody who participates gets €2 but those in treatment money could win an additional €4 if they reach their target[5]. On the other hand, those with a reward of time are cheaper since time is "costless" to provide. After filling in the numbers, the ratio is set equal to $\frac{N2}{N1} = 1.07$ and the group sizes should consist of 89 and 83 respondents, for treatment money and time respectively. At last, one can see that enlarging the group sizes leads to an increased actual power of 0.80.

Figure C2: Required Sample Sizes.

Performed by G*Power.



[5] Assuming expected utility theory, a 20 percent chance to win €20 equals €4 (Wakker, 2008).