ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS

M.Sc. BEHAVIORAL ECONOMICS

# THE ROLE OF THE FEELING OF COMPETENCE IN MOTIVATION CROWDING OUT: A FRAMED FIELD EXPERIMENT

Supervisor: Dr. Jan P.M. Heufer

Second Reader: Dr. Jan T.R. Stoop

STEFANO DEL PARIGI

STUDENT NO: 435088sp

## Abstract:

An important theoretical contribution to the understanding of motivation crowding out comes from psychological literature with Cognitive evaluation theory. This theory suggests that extrinsic rewards will only undermine intrinsic motivation if perceived as controlling, while they will have a positive effect on intrinsic motivation if they rouse feeling of competence. Building on this insight the current thesis investigates the role of the feeling of competence on the perception of incentives. A framed field experiment involved university students from two different backgrounds (technical and non-technical) facing two tasks pertaining those two backgrounds. For each of the two student backgrounds, a treatment group received (monetary) extrinsic rewards and a control group did not. The hypothesis tested was that, due to the relevance of the feeling of competence, the extrinsic rewards would have been more effective (rising the performance to a higher extent) on students facing tasks matching their field of expertise. Within the empirical analysis, the data obtained from the experimental sessions were analyzed with parametric, non-parametric and regression techniques. No evidence in support of the hypothesis stated above emerged, while evidence of motivation crowding out did.

# Acknowledgments:

I would like to express my gratitude to Dr. Jan Heufer for his support, patience and guidance throughout the entirety of this thesis process. His ability to interpret my too often very raw thoughts and transform them into material for academic research I grew to be proud of, is valuable beyond what I can express in words.
I would like to thank Dr. Jan Stoop: as I already had the opportunity to express to him, reading of his experimental adventures and experimental papers is what made me enthusiast about the experimental side of economic in the first place.
I would like to thank my parents and my sister for the material and emotional support that never lacked since well before this master begun.
Mr. Olaolu Adeboye, our journey started as housemates and continued as a strong friendship: observing your dedication and your wise determination has been an important inspiration. Mr. Vincent Seguin, your out-of-the box approach to the world and visceral sense of friendship contributed to make this learning journey unforgettable.
Thanks to Ami, who patiently witnessed the struggle, and never denied her support. Thank you for your teachings.

# Table of Contents

# 1. Introduction

## 1.1 Research Question and Aim of the Study

Motivation crowding out is a well-established phenomenon in labor economics and cognitive psychology consisting in extrinsic motivators possibly crowding out intrinsic motivation, resulting in an overall worsening of performance level. Different theories within the psychological literature tried to explain this phenomenon. In particular, the current thesis builds on the explanation resulting from the cognitive evaluation theory. Cognitive evaluation theory predicts that extrinsic rewards will crowd out intrinsic motivation when they are perceived as controlling, while they will enhance intrinsic motivation if they are perceived as identity enhancing. The reason lies in the needs of self-determination and competence.

The original research idea involved a broad analysis of all the possible circumstances that could contribute to motivation crowding out, but it soon became clear it was not feasible, provided the constraints a master thesis poses. I then decided to focus on one of the many possible causes of motivation crowding out, considering the insights provided by cognitive evaluation theory. This theory stresses the relevance of self-determination and competence, thus I had to select one of the aspects that could possibly play a major role in these respects. What came to my mind was that an aspect very specific to the labor market is expertise, which is closely linked to competence. I thus wondered how someone's awareness of his or her expertise, and ultimately competence, in a specific task could affect the perception of a reward payed on that task. Using cognitive evaluation theory's terminology, it appeared sensible to expect that extrinsic rewards paid on tasks someone feels belonging to their field of expertise would have been perceived as more identity enhancing, and less controlling. For this reason among the many possible factors playing a role in motivation crowding gout according to cognitive evaluation theory, I indeed chose to investigate this specific aspect. Thus, the research question the current thesis investigates is the following:

*RQ: Are subjects more positively responsive to extrinsic rewards when these are paid on tasks they perceive as belonging to their field of expertise?*

This research question is relevant to both practitioners and the academia. As far as practitioners are concerned, ex ante this thesis could possibly be able to provide Human Resource departments with new evidence suggesting the importance that the relation between a task and the field of expertise of an employee has in enhancing intrinsic motivation, and ultimately productivity. The result would be providing academic evidence allowing payment schemes' designers to minimize the outflows of their

company. On the academia side, this thesis contributes to the process of integration between Cognitive Psychology and Economics, strengthening the role Behavioral Economics plays as support to all others field of Economics, in this case Personnel Economics.

In order to investigate the research question I implement an experimental design involving a framed field experiment with students, and analyze the data using non-parametric, parametric and regression techniques.

This study contributes to previous literature on motivation crowding out by, to the knowledge of the author, being the first to implement a framed field experiment in order to test the relation between a subject's field of expertise and the sense of competence. The evidence obtained does not support the hypothesis of an affirmative answer to the research question, opening the path for further research on the relation between motivation crowding out theory and cognitive evaluation theory. The next section will provide the theoretical framework to the thesis, Section 3 will introduce the experimental framework. In section 4, I present the statistical analysis, and section 5 concludes the thesis with the presentation of key findings, an overview of the limitations, and suggestions for further research.

# 2. Theoretical Framework

## 2.1 Introduction

Richard M. Timmus, in his 1970 essay *The Gift Relationship: From Human Blood to Social Policy*, claimed that monetary compensation could undermine the sense of social duty of individuals. His example regarding the blood donations alleged that monetary incentives for blood donors would indeed decrease their willingness to donate blood. Albeit not providing any substantial evidence, the theory immediately attracted a considerable amount of attention, both positive and negative. Most notably, Solow (1971) and Arrow (1972), two of the most influential economists and later Nobel Price winners, when reviewing the book containing such a bald claim not supported by empirical evidence nor substantial links to any contemporary psychological theory, did not find it plausible that extrinsic rewards could reduce the desired behavior. Since then, two main streams of literature investigated the intrinsic motivation crowding out issue: psychological literature and economic literature, and both empirical evidence and more solid theoretical frameworks emerged.

From a theoretical perspective, within social psychology it is possible to identify the two theories on which all the subsequent argumentations on crowding out build on: self-perception theory – Bem (1967a,b) – and cognitive evaluation theory – Deci and Ryan (1980) and Deci and Ryan (1985).

Self-perception theory is a theory of self-attribution of motives. It builds on the assumption that individuals are not perfectly aware whether intrinsic or extrinsic motivations are driving their behavior when they perform a task, and they infer it from the circumstances. If the external incentives provided for a task are strong, they tend to perceive it as extrinsically motivated, while if those are weak and they still perform the task, they will tend to assume that the hedonic characteristics of that task make it interesting by itself, thus their behavior must be intrinsically motivated. In this framework, the crowding out effect is explained by over-justification. When the extrinsic incentives on a task that is intrinsically enjoyable are strong and salient, the individual will tend to attribute the motive for performing the over-justified task to the extrinsic motivators – as they are easier to recognize compared to the intrinsic motivators.
Cognitive evaluation theory instead, builds on the assumption that people have a psychological need for self-determination and competence. How extrinsic rewards affect intrinsic motivation is determined by their effect on perceived self-determination and perceived competence. If extrinsic rewards are perceived as controlling, they will lower the degree of satisfaction of the self-determination need, hence undermining intrinsic motivation, while if they are perceived as informative of an individual's competence, they will increase intrinsic motivation.
Following, the main contributions from the psychological and economic streams of research to crowding out literature.

## 2.2 Psychological Literature

In the psychological stream, by 1970, extensive literature had confirmed the significant role incentives can have on behaviors – e.g. Kruglansky et al (1971). In that period, the possible drawbacks of incentives were started to be investigated as well: in particular Deci (1971) and Lepper, Greene and Nisbett (1973) investigated the possibility that rewards could negatively affect the performance of a task enjoyed by the subject of the rewards. There, the hypothesis that the extrinsic reward could undermine the "intrinsic motivation" was introduced, and tested by comparing behavior levels of a treatment group that received a tangible reward and a group that received no such reward, after the reward had been removed. In the first study the intrinsic motivation was operationalized by the choice of tasks already being performed consistently by participants (puzzle solving and writing students'

newspaper headlines), while in the second by the selection of participant (nursery schoolchildren) who were interested in drawing.

A further important contribution on the matter came by Cognitive Evaluation Theory, CET (Decy and Ryan, 1985), predicting that extrinsic rewards will only undermine intrinsic motivation if perceived as controlling, why they will have a positive effect on intrinsic motivation if they rouse feeling of competence. The selection of the tasks and the choice of framing the tasks in the experiment of the current thesis, build on this very insight (more about this in section 3). In general, the psychological literature, most notably CET, finds crowding out evidence only for behaviors where intrinsic motivation was high prior to the treatment.

The psychological stream of literature also investigated the possibility of crowding out for incentives and disincentives over health-related behaviors. As mentioned above, the very milestone of the literature on this topic, Timmus (1970), regarded a health-related context. Promberger and Marteau (2013) provides a rich overview. The results are that, consistently with the general tendency, psychological literature does not find strong evidence of crowding out here, as the pre-reward high intrinsic motivation does not appear to be there for health related issues – if smokers were intrinsically motivated not to smoke, they would not be smokers in the first place –.

## 2.3 Economic Literature

The economic literature contributed with the introduction of laboratory experiments involving more complex tasks and field experiments. Frey and Oberholzeer-Gee (1996) presents a study of motivation crowding out based on a survey regarding the so called “Not in my backyard” problem, which refers to the socially desirable but locally non-wanted projects. Contingent valuation (CT) questions were asked in referendum format, on whether or not the respondents would have accepted the building of a nuclear waste repository in their area, in two locations in the central Switzerland. First the question was asked alone, then it was asked adding an upfront payment by the government to compensate the citizens, ranging from more than 2,000$ per year per individual to more than 6,000$. The introduction of the monetary incentive corresponded to a drop in the acceptance rate by half, disregarding the amount offered. Gneezy and Rustichini (2000a) is worth a mention for its outstanding spread among the scientific community. The study involved a field experiment where a fine was introduced for parents picking their children late form a daycare in Israel, which according to classical economic

theory should have induced the parents to reduce the misbehavior. The treatment, after a temporary adaptation phase, resulted instead in an increased number of parents being late, both during and after the fine was in place, suggesting that the intrinsic motivation was crowd out for good. The fine amount was modest, which induced the authors to open to the possibility that the reason for the result was the incentive not being strong enough. The other explanation suggested, which gave the title to the paper (i.e. "*A Fine is a Price"*), was that once they introduced a fine, they priced the time of the day care operators, transforming a non-economical transaction into an economic one. Another paper written by the same authors that year, Gneezy and Rustichini (2000b) included a double set of experiment: different groups of students from University of Haifa were asked to perform tests similar to the IQ ones, and collect money for donations. For each task, different treatment group would receive different amount of money per right answer or amount of donation collected. What emerged is that, for both tasks, students receiving very low amount of money (around 10 cents of the local currency) were performing worse than students not receiving or being mentioned any compensation. Again, the authors stressed that the result was driven by the shift in the relation, perceived as an economic one by the students receiving money for their performance, but not enough money to be worth their effort. This results raised the argument that the reason why low compensation may negatively affect performance, might be that they give a signal of 'low social value' of the tasks it is paid on.

Frey and Jengen (2000), containing an extensive review of the evidence supplied by literature in laboratory and field experiments, more formally defined the Motivation Crowding Theory, building on CET. The study, which insisted on the centrality of the perception of the extrinsic motivation factor, also formalized the possibility for crowding in intrinsic motivation. More specifically, while both intrinsic and extrinsic motivation have a positive direct effect on performance, extrinsic motivation has also an effect over intrinsic motivation, a crowding effect, which can be either positive and reinforce the intrinsic motivation – crowding in – , or negative and weaken it – crowding out – , and will indirectly affect the performance. The direction of the crowding effect depends on the perception the agents have of the extrinsic motivation factor. Should they perceive it as controlling, intrinsic motivation is crowded out, while if it is perceived as supportive, the crowding in will prevail. Heyman and Ariely (2004), building on Fiske (1992)'s relational theory, contributed by underlying another feature to be considered relevant for the crowding out effect extrinsic compensations can have on the performance on a task. The authors assumed that it is possible to distinguish monetary, social and mixed markets, where monetary markets are highly sensitive to the magnitude of compensation while social markets are not, and mixed markets tend to resemble the former more closely. They test these hypotheses in three different experiments with students of University of California, Berkley and Massachusetts

Institute of Technology. In the first one they put the subjects in front of a hypothetical situation by asking them the likelihood that other students – the question was not about themselves so to avoid any desirability bias – would have helped loading a sofa into a van for no money or different amounts of money or candies (to proxy for social markets). In the second one, students were asked to perform an actual, repetitive – so to be voided of any intrinsic motivation –, task under the same payment conditions, while in the third experiment they repeated this latter scenario with an added element, the statement of the money value of the candy, which should proxy for a mixed market situation. In all of the three experiments, empirical evidence fully supported the suggested theory.

The subset of the economic literature on crowding out this paper builds more closely on, is the one that applied the theory in the workplace. The reasons why this "context" aspect deserves particular attention are many. It is in the first place reasonable to expect that the role, nature and strength of the intrinsic motivation might indeed differ quite substantially in a job framework. Kreps (1997) suggests that intrinsic motivation on the workplace could be hard to properly detect in the first place, as it might actually be the result of a hidden'*worker's response to fuzzy motivators as fear of discharge, censure by fellow employees, or even the desire for coworkers' esteem*'. Building on this aspect the author claims that pride in an employee's own job my increase the optimal level of effort he exerts, for the pride of the job itself. Fehr and Falk (2002) explains the psychological foundation of incentives by pointing out three main non-pecuniary motives can be shown to shape human behavior: the desire to reciprocate, the desire to gain social approval, and the intrinsic enjoyment arising from working on interesting tasks. The authors insists that, being extrinsic rewards and their variation overtime are expected on the workplace, it is plausible to assume that their effect on intrinsic motivation is weakened, if not nulled. Furthermore, economic rewards are intrinsically part of labor contracts, whether flat or performance-related. Considering that a consistent amount of literature, as explained above, suggests that monetary incentives can crowd motivation by signaling the value of a task, it is worth investigating how individuals interpret this signal in a workplace framework.

Huffman and Bognano (2015) provides further empirical evidence on the subject. The study applies the experimental design typically used by psychology on this issue, ABA for the treatment group, where period B is the one where treatment is submitted, and AAA for the control group, but with several contributions to the previous literature. In the first place, the authors claim this to be the first study where this "three stage" approach is applied to an experiment with actually paid workers in a real world setting for a crowding out investigation. This allowed to monitor how the effect of the introduction of the new compensation scheme affected performance on a larger spam of time. Second,

it investigates how the heterogeneity of psychological features and social preferences among the participants affect the reaction to performance pay through questionnaires submitted to the participants on their personality. Furthermore, it involves strong-enough monetary incentives so to avoid the possibility that the incentives are perceived as too small, as shown above, and, beyond the scope of this thesis, it investigates how different forms of compensation, more than the level of compensation itself, affect the performance. The experiment itself regarded 39 workers hired to convince the attendants to a street festival to get registered for a company database, with the sign-ups being registered minute-by-minute. The participants randomly assigned to the control group received a flat amount of 18$ per hour, while those in the treatment group, on top of that money were also entitled for additional 5$ per sign-up during the second out of the five total hours of that working day. Three main findings emerged. Workers in the treatment group, while achieving a considerably higher result during the treatment hour compared to the control group, performed way worse in the subsequent hours. Furthermore, self-reported intrinsic motivation dropped in the treatment group after the incentive was removed, as the task was considered 'less fun' by the majority of the subjects. Lastly, many personality traits and social preferences as extroversion and positive reciprocity were found relevant for understanding how the incentive experience affected the workers.

## 2.4 Conclusions

The literature provides quite substantial evidence of extrinsic rewards possibly crowding out intrinsic motivation from both field and laboratory experiments. Nevertheless it less clear under which circumstances this can happen. The factors that seemingly contribute the most to the presence of a crowding effect when extrinsic rewards are applied are the level of previous intrinsic motivation, the perception of the rewards as either controlling or rewarding, the perception of the context, and individual characteristics of the subjects in terms of social preferences and psychologic characteristics.

Building particularly on the cognitive evaluation theory, this thesis will contribute to the existing literature as follows. It will investigate the presence of crowding out effects for tasks explicitly related to a certain field of academic expertise, first paper to do so at the author's knowledge. The link between the task and the field of expertise will be stressed by the framing of the task, as explained below in the experimental design section. The choice for this specific experimental structure is driven by the hypothesis that, according to cognitive evaluation theory, rewarding subjects on a task that is perceived as being pertinent to their expertise, should narrow down the controlling perception of incentives while reinforcing the self-determination need.

The hypothesis tested within this thesis, stemming from the literature examined in this section is the following:

*$H_1$: Subjects are more positively responsive to extrinsic rewards when these are paid on tasks they perceive as belonging to their field of expertise.*

# 3. Methodology

## 3.1 A Framed Field Experiment

The debate about external validity of laboratory experiments in social sciences has quite a long history. In particular, the debate throughout the years focused on the experiments regarding social preferences. Cross (1980) provides an effective summary of the main argument against the external validity in this field: "*It seems to be extraordinarily optimistic to assume that behavior in an artificially constructed 'market' game would provide direct insight into actual market behavior*". More specifically, five aspects of laboratory experiments on social sciences that are some of the main sources of criticism follow[1]. The first one consists of the nature and extent of obtrusiveness, which refers to the fact that a subject who knows she is being observed will tend to bias her behaviors towards what she considers socially desirable or the experimenter will. The second one is the context of the game. It is claimed that, if on the one hand the neutral framework used in laboratory settings is a signal for a too tight control exerted by the experimenters, on the other hand it is not possible for the experimenters to achieve any control at all on the environment, due to the personal subjects' interpretation of the context and instructions. The third criticized aspect regards the stakes of the game, and the relative argument points out that is usually hard for researchers to access funding big enough to provoke subjects less desirable behaviors. The fourth aspects r

egards the possibility for self-selection of participants, meaning that the most pro-social people might self-select themselves in the experiment, thus providing the mere illusion of a representative sample. Fifth, the artificial outcome space of choice is criticized, meaning that the span of actions and interactions possible in the real life are almost impossible to replicate in a stylized context such as the laboratory. For brevity (external validity of lab experiment on social preferences is a largely debated

[1] These come from the slides on the lecture pertaining External Validity from the course of Experimental Economics held by Dr. Jan Stoop at the Erasmus School of Economics in academic year 2016-2017.

topic in more proper forums[2]) , I will reassume the counterarguments by mentioning the five precepts from Smith (1982) that an economic experiment should have in order to ensure control, and thus address all of the mentioned critiques. The first one is Nonsatiation: "*Given a costless choice between two alternatives, identical except that the first yields more of a reward than the second, the first will always be chosen over the second, by an individual. Hence, utility U(V) is a monotone increasing function of the monetary reward, U'(V)>0 , where V is dollars of currency*". Namely, as long as subjects value more over less, control can be achieved by introducing monetary consequences to actions. The second one is Saliency: "*Individuals are guaranteed the right to claim a reward which is increasing (decreasing) in the good (bad) outcomes, of an experiment. Individual property rights in messages, and how messages are to be translated into outcomes, are defined by the institution of the environment*". This precept means that subjects must bear the consequences of their actions, and they should not be deceived. The third one is Dominance: "*The reward structure dominates any subjective costs associated with participation in the activities of an experiment*". This means that the stakes must be high enough to make subjects willing to exert effort on the tasks presented in the experiment. The fourth one is Privacy: "*Each subject in an experiment is given information only on his/her own payoff alternatives*". Privacy addresses the potential issue of subjects attaching weight to the outcome of others. The fifth one is Parallelism: "*Propositions about the behavior of individuals and the performance of institutions that have been tested in laboratory micro-economies also apply to non-laboratory micro-economies where similar ceteris paribus conditions hold*". It is worth noticing that this precept does not mean that behaviors observed in the laboratory are completely predictive of real behaviors, as there might still be characteristics that makes subjects in the laboratory differ from those in the real life, but if those differences could be kept constant, the behaviors would be the same.

I summarized the main critiques to external validity of laboratory experiments, and presented the five requirements an experiment needs in order to overcome these potential weakness. Nonetheless, there are two more specific reasons beyond these general remarks that suggest me that an experiment with a standard lab setting might not be the best fitting one for this research question. The first one lays in the importance that a proper framing has in this thesis. Indeed, at the core of the investigation there is the relation between intrinsic motivation, extrinsic rewards and identity, specifically the set of competences that determines one's field of expertise. Thus, ensuring that all of the features of the experiment enhance one's own professional/didactical identity perception cannot but bring benefits to the design of the experiment. In other words, the more a subject feels to be in her professional/didactical environment during the experiment, the more likely she will perceive one of the tasks as pertaining her field of expertise, as will become clearer below. The second reason is as

[2] For example Benz and Meier(2008) or Cleave et al. (2013).

simple as practicality. Once again, as the objective was to obtain two groups of subjects homogenous within each group in terms of academic backgrounds, the logistically more convenient way was to bring the experiment to the subjects, instead of the other way around.

For these reasons, I decided to opt for what in the field of experimental economics is defined as a framed field experiment, meaning that this experiments features non-standard pools of subjects in their own academic environment, facing tasks related to their fields of expertise.

## 3.2 Experimental Design

This paper aims at testing whether or not extrinsic rewards crowd out to a greater or lesser extent intrinsic motivation when applied to tasks belonging to one's perceived field of competence, compared to when applied to other tasks. In an ideal scenario, the best possible experiment design for this goal would consequently include a treatment group composed by subjects receiving some form of extrinsic reward on their actual job, and a control group of their colleagues (for sample homogeneity sake) undergoing the same kind of treatment for tasks considered to be out of their professional competences. In this way the only variable changing would be the job-related vs. non-job-related nature of the incentivized task (arguably, provided all the care due to exclude selection biases and confounding factors), allowing us to find the causal effect of the nature of the incentivized task. Unfortunately, setting up this experiment would require having access to the HR management department of a company, in order to get in the position to affect the payment schemes design, and this was not possible. The challenge then became how to replicate as closely as possible the set up briefly described above without involving actual firms.
Not running the experiment with actual employees implies that in the research design I had to recreate the "on-the-job" feeling in the subjects without them actually being on the job. I tried to rationalize which aspects of being on the job could be the crucial ones when it comes to changing the receptiveness of workers toward extrinsic rewards. The process led me to recognize two specific factors: first, both explicit and implicit contracts that regulate labor involve reciprocation of efforts in the workplace, making people expect external (monetary) consequences to their efforts, which is exactly what extrinsic motivation consists of in the labor market. In other words, when in their workplace, people are aware they are executing an economic activity, and expect their effort to be reciprocated by monetary rewards. A second factor is what can be called the professionality display. On the job people are supposed to be paid for their competences, thus not responding to an extrinsic

reward offer might be more costly than the value of the extrinsic reward itself, as it might suggest low motivation, or lack of specific competence.

In the absence of an actual firm, the contractual aspect is hard to be replicated, thus I chose to focus on the second one. In order to substitute workers with other subjects who could still be consider representative of their respective fields in terms of expertise, I chose graduate and undergraduate students, because it is relatively easy to categorize them in fields of expertise based on the faculty of their study. Furthermore, by interacting with them in a didactic environment, more precisely during lectures, I tried to interact with them when their identity as students of those specific backgrounds was the most vivid.

The scheme below illustrates the structure of the experimental design:

**Figure 1**. Experimental Design

| | Task related to technical field (Sudoku) | | Task related to non-technical field (word puzzle) | |
|---|---|---|---|---|
| Subjects in technical field (I.T. Students) | treatment (extrinsic reward) | control (no extrinsic reward) | treatment (extrinsic reward) | control (no extrinsic reward) |
| Subjects in non-technical field (Law Students) | treatment (extrinsic reward) | control (no extrinsic reward) | treatment (extrinsic reward) | control (no extrinsic reward) |

Two tasks were selected, regarding two easily distinguishable fields, technical (Sudoku) and non-technical (Word Puzzle). Also two groups of subjects, respectively involved in studies in technical (I.T. students) and non-technical field (I.T. students), were selected. In each of the four categories obtained by the interaction of subjects and tasks, two subgroups were defined: treatment and control, respectively receiving and not receiving extrinsic rewards. In this way, eight subgroups were obtained such that subjects in both field could be assigned both types of task, and that for each of the four kind-of-task-to-kind-of-people interaction, there was a subgroup receiving extrinsic rewards, and a subgroup supposed to be performing the task only based on intrinsic motivation. What the hypothesis

suggests is that the difference in performance between people receiving extrinsic rewards and their counterparts not receiving extrinsic rewards (keeping constant field of expertise and task), will be greater in the groups where field of expertise and task type are matched. This would imply that in those cases, the identity-enhancing aspect of the extrinsic rewards has overcome the controlling aspect in the perception of the subjects, thus proving the hypothesis tested right.

## 3.3 Experimental Material

Appendix 1 contains the different versions of the cards handed out to the subjects. Each card had the technical task on one side and the non-technical task on the other side. Since the courses were taught in Italian, the cards were distributed in Italian, and consisted of a translation of the ones displayed in the appendix. The cards distributed to the technical students (I.T. Students), both incentivized and non-incentivized, contained a sentence at the top of the page containing the instructions for the technical task that pointed out how they were likely to have acquired throughout their studies specific competences to address this kind of tasks. The same sentence was on the non-technical task side of the cards handed out to non-technical students (Law students), both incentivized and non-incentivized. I find this feature to be of great relevance for the experiment, as it was meant to contribute to the framing of the tasks in such a way that they could be perceived as part of the students' academic background. The rest of the information contained at the top of the page regarded instructions on the rules of the tasks, on the timing (five minutes for each side), an example. The cards of the incentivized subjects (the treatment groups) of both the technical and non-technical students groups, also contained information on the payment schemes of the tasks. Below this informational part, on each side it was asked age and how frequently the subjects performed those tasks. This latest question was meant to both allow to control for experience and to provide a measure of intrinsic motivation. More about this in section 3.5 where data are presented. Furthermore, I prepared two different version of the cards for every student's background-treatment match, in which the order of page 1 and 2 assigned to the technical and non-technical task was inverted. This was intended to avoid that all of the subjects systematically performed better on the first task (say always Sudoku) only because it was the first one and they were less tired. More on this in the data section.

The technical task consisted of ten 4x4 Sudoku schemes, the non-technical task of a single scheme of word puzzle, where subjects had to find as many words as possible with the sixteen available letters. The tasks were chosen based on the fact that their solution were quick and easy to verify (so that it was possible to pay the subjects right after the session), available in different degrees of difficulty, and

easy to be framed in terms of technical/numerical expertise vs. non-technical/verbal expertise. For both tasks the payment scheme was based on pay-for performance: every Sudoku scheme correctly solved corresponded to 50 eurocents, while for word puzzle every found word of a length of three letters corresponded to 5 eurocents, four letters 10 eurocents, and so on up to 50 cents for ten-letters-long words.

## 3.4 Experimental Setup

The experiment took place in three different sessions in two different days at the end of April 2017 at Università Statale di Milano (University of Milan). The initial idea was to have two sessions for each academic field group. Nonetheless, since I had a limited time to stay in the Country and in those weeks many lectures were canceled because of public holidays, I decided to merge the sessions for the Law students, as their lecture rooms were big enough to do so in an effective way.

The first session involved the non-technical students and occurred with a class of Law students at the beginning of their lecture of "Philosophy of Law", a mandatory course for all 2nd year bachelor Law students. The room had a capacity of about 200 people, and about 80 students were attending. The structure of the room was very convenient: the seats were organized in four big blocks by two perpendicular corridors large two meters that separated the right and left halves, and the front and the rear halves. I had decided to have only 20 subjects in the incentivized group and 20 in the non-incentivized group (for budget reasons, and in order to have control over sample sizes), and originally to conduct four different sessions: one for incentivized technical students, one for non-incentivized technical students, one for incentivized non-technical students, and one for non-incentivized non-technical students. Nevertheless, due to university's schedule issues explained above, I decided to take advantage of the structure of the room and I ran both the non-incentivized and incentivized sessions at the same time in the same room, while still minimizing the information spillovers between the two groups. I explained to the professor hosting me during his lecture that I wanted the subjects involved in the experiment to be in the front so that I could more closely make sure they were not copying from each other, but I also made him clear my concerns about involving in the experiment only the twenty people sitting in the front. It could indeed be the case that the most hard-working students, the most motivated ones, or those who arrived the earliest, took those seats, which could imply a non-representative sample and ultimately a selection bias. He thus proceeded explaining to the students that that morning they would have partaken to an experiment (as I asked him he did not mention the topic, which I did myself at the end of the lecture), and made everyone stand up and change their seats

around the class. I made sure that only 20 people were in the front right block, and 20 in the front left block. Those would have been the subjects of the treatment and control group. He told to the rest of the students not involved in the experiment to be patient and revise until the end of the experiment. In order to avoid for non-incentivized people to know that other subjects were doing the same tasks of their own while receiving money, I only mentioned the money rewards on the incentivized subjects cards. The reason for this is that fairness concerns of the subjects could have impaired the results otherwise.

After this phase I made my explanatory pitch, over these seven points:

"Good morning,

1. Today you will help me in a master thesis experiment.
2. You have two tasks on the cards, one for each side of the sheet I gave you.
3. The cards will be anonymous (here the professor stated that they would not count in any way for their exam grade of that course).
4. Please read the instruction carefully, and in silence. No questions are allowed after you receive the cards.
5. Please start from page 1, the one you will see when we hand you the cards.
6. As written in the instructions, you will have 5 minutes for each side of the task. I will call the time out loud, and will make sure you all start at the same time with page two. At the end we will collect all the cards.
7. At the end of the tasks, please keep the post-it on the left corner of the card, the reason will be explained later."

When handing-out the card, I had the pile for the incentivized students, while my assistant had the ones for the non-incentivized students. I had organized both piles of cards in such a way that after a "version 1" card (numerical task on page 1) there was a "version 2" card (verbal task on page 1). This should avoid any order effect (in case a non-random subsample or the entire sample started from one of the tasks), and avoid that all people in the front or all people in the back (of the front blocks) had the same task first. We proceeded to distribute the cards in such a way that the front right block received the non-incentivized cards while the front left block, separated by the corridor, received the incentivized cards. On the desks, only the card, a pen, a pencil and a rubber were allowed.

At the end of the 10 minutes we collected the card, and the professor asked me to explain the motives and background of the experiment, which I did. Then I explained that all the people in the front right block were the control group, and they would not receive any compensation, while those in the front

left block were the treatment, and as they had read on their instructions. I also informed the students in the treatment group that I would return at the end of the lecture in order to pay them the money they had earned, according to their performance.

I had written, in pen, numbers from 1 to 20 in the top left corner of each card, and covered that with a post-it with the same number on it, that they could keep as a receipt (so that I could be able to identify the subjects while keeping the cards anonymous). The post-its in the control group were green, those in the treatment yellow, so that nobody in the control group could try to sneak in and ask for money. After my explanation at the end of the experiment, a student in the control group asked why I also told the control group to keep the receipts, and why those were there in the first place if they would not need it. I explained it was in order not to let them suspect any difference in their cards (the colors of the post-its were similar enough that the difference could not be noticed from the distance separating the two groups). This question proved to me that this technique worked.

At the end of the lecture, as agreed with the professor, I came back in and the class, where the treatment group was waiting for me, and gave them the money, for a total expense of 33 euro. Interestingly enough, a girl who had earned the second highest price (5 euro), told me she didn't feel comfortable accepting the money, and told me to keep them. As discussed in further detail below, this might mean intrinsic motivation being higher than extrinsic motivation for her at least).

For the remaining two sessions of the experiment, the ones involving technical students, I went to the campus of the same university where the I.T. faculty is. In those building rooms are smaller, they are built with a capacity of around 30 students. For this reason I performed the treatment and control sessions (namely the incentivized and non-incentivized sessions) with two separate classes of students, both attending the same mandatory course of their second year of master in I.T. When they got enrolled for that course, the faculty software randomly selected them in one of the two time schedule, which should exclude any selection bias due to a student being in one of the two classes. The first class, the one with non-incentivized subject, was composed by exactly 20 students, while the second one by 25. In here, again, the professor had to ask to 5 people (part of the last row) to wait until the experiment was over. In both classes, after an introduction by the professors akin of the one of the previous day, I made my pitch, akin to the one exposed above, in seven points. The only difference was in the non-incentivized class, where this time anyway, as no deception regarding some subjects receiving compensation was necessary, I did not put the post-its on the cards of the subjects in the class not receiving rewards, and consequently did not mention the seventh point.

Again, at the end of the lecture of the incentivized class, I came back and distributed the monetary compensations to the students showing me their numbered post-its. One student showed up giving

back the ticket stating he was not interested in the money but wanted to know more about the experiment, and three students did not show up in order to collect their money, once again suggesting that the extrinsic motivation provided could have been not high enough.

### 3.5 Data

After the three sessions were completed, the data from the cards were gathered. For every subject, cards provided the scores for each of the two tasks, and the answers to the set of questions I asked the students to ask. These included the experience for each task, the age of the respondent, the field of expertise in terms of academic background (technical or non-technical), whether or not the subject was in the incentivized (treatment) group, whether or not the payment was collected after the experiment, and which of the two tasks was performed first.

Table 1 presents a summary of the variables considered for the different groups of subjects.

**Table 1**. Summary statistics for the main variable involved in the analysis

| | **I.T. Students** | | | | | | **Law Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Control | | | Treatment | | | Control | | | Treatment | | |
| **Variable** | Mean | S.D. | N | Mean | S.D. | N | Mean | S.D. | N | Mean | S.D. | N |
| Sudoku Score | 5.15 | 2.85 | 20 | 3.75 | 2.25 | 20 | 4.60 | 1.96 | 20 | 3.85 | 2.28 | 20 |
| Word Puzzle Score | 34.4 | 15.9 | 20 | 27.25 | 16.62 | 20 | 28.43 | 15.69 | 20 | 25.45 | 16.11 | 20 |
| Sudoku Experience | 1.95 | 0.94 | 20 | 1.55 | 0.76 | 20 | 1.50 | 0.51 | 20 | 1.50 | 0.69 | 20 |
| Word Puzzle Experience | 1.90 | 0.97 | 20 | 1.45 | 0.60 | 20 | 1.55 | 0.69 | 20 | 1.35 | 0.59 | 20 |
| Age | 23.3 | 1.34 | 20 | 24.45 | 2.01 | 20 | 20.75 | 1.25 | 20 | 21.4 | 1.79 | 20 |
| Payment refused | | | - | | | 5 | | | - | | | 6 |

## 4. Analysis

In order to verify the hypothesis with the data gathered I decided to proceed with two different approaches. The first one consists of a mean comparison approach, while the second one is a regression-based one, with increasingly complex specifications. As regarding the means comparisons, I started with non-parametric tests, then after verifying that the conditions for parametric ones held

moved to the parametric version of the ANOVA, the most popular one, which requires data at least at the ordinal level.

## 4.1 Means Comparisons

According to the hypothesis to be tested, the goal of this analysis was to verify whether the effect of the extrinsic rewards as extrinsic motivators was higher on tasks that matched the background of the subjects. This means that I expected the difference in score for Sudoku – the technical task – between extrinsically incentivized and non-extrinsically-incentivized subjects to be higher for I.T. students – technical students. Vice versa the difference in Word puzzle – non-technical task – between treatment and control group was expected to be higher for the non-technical students, the Law students.

### 4.1.1 Pairs' Comparisons

The first step was to verify whether for each task there was a statistically significant difference in the scores between the incentivized and non-incentivized groups within each of the students' backgrounds. Namely, I needed to measure the effect of the treatment. In order to do so I compared the scores of each task for the treatment groups to their control counterparts.

I opted for a non-parametric test, as this kind of test require far less assumptions compared to their parametric counterparts: observations need to be independent and drawn from an underlying continuous distribution, (eleven classes, as in the Sudoku score, are considered a good enough approximation to continuity). Furthermore, non-parametric tests offer some advantages, as the reduced impact of outliers and a higher suitability to small samples. On the other hand, they are less powerful in a statistical sense compared to their parametric counterparts. I implemented the Mann-Whitney U Test, commonly used to verify whether two samples are drawn from the same population. This is a rank-based test that has the identity of the medians as null hypothesis against a symmetrical alternative, and requires data at least at the ordinal level.

### 4.1.2 Overall Comparison

Pairs' comparisons allowed me to observe the magnitude and direction of the treatment effect within each group of students' background. A further step consisted in a between-groups analysis in order to address the actual research question. The statistical tool I used for this part of the analysis is the two-

way ANOVA. More specifically, I implemented this tool twice, separately for the Sudoku scores and the word puzzle scores. This kind of analysis of variance is used when the goal is to conjointly investigate the effect of two different factors on a dependent variable. This setup perfectly suits the case for the current paper: in this part of the analysis in facts the objective was to analyze how belonging to a certain field of expertise (being a technical versus non-technical student) and receiving or not the treatment (being extrinsically rewarded or not) would affect the task score. The most relevant part should have been the interaction between these two factors, and for this reason, I introduced an interaction term in both the analyses of the scores of the two tasks. The analysis of variance builds on three assumptions: independence of observations, normality of the residuals, and homoscedasticity. In the results session I explained how these were verified.

## 4.2 Regression Analysis

I decided to conclude the analysis by slightly changing the perspective I was observing the scores from. For completeness sake, I wanted to analyze how much all of the different factors I obtained from the data collected determined the scores. It was important to me to make sure that the fact I was observing the scores focusing on the contribution of the extrinsic rewards and the relation between the background of the students and the field of the task, was not making me loose the bigger picture. In other words, I wanted to make sure that extrinsic rewards, the interaction of the tasks' and students' field of competence, together with the other variables I obtained, were actually relevant explanatory variables able to explain a significant portion of the observed variation in scores. An OLS regression model provides adequate insights for this goal.

## 4.3 Results

### 4.3.1 Mann-Whitney U Test

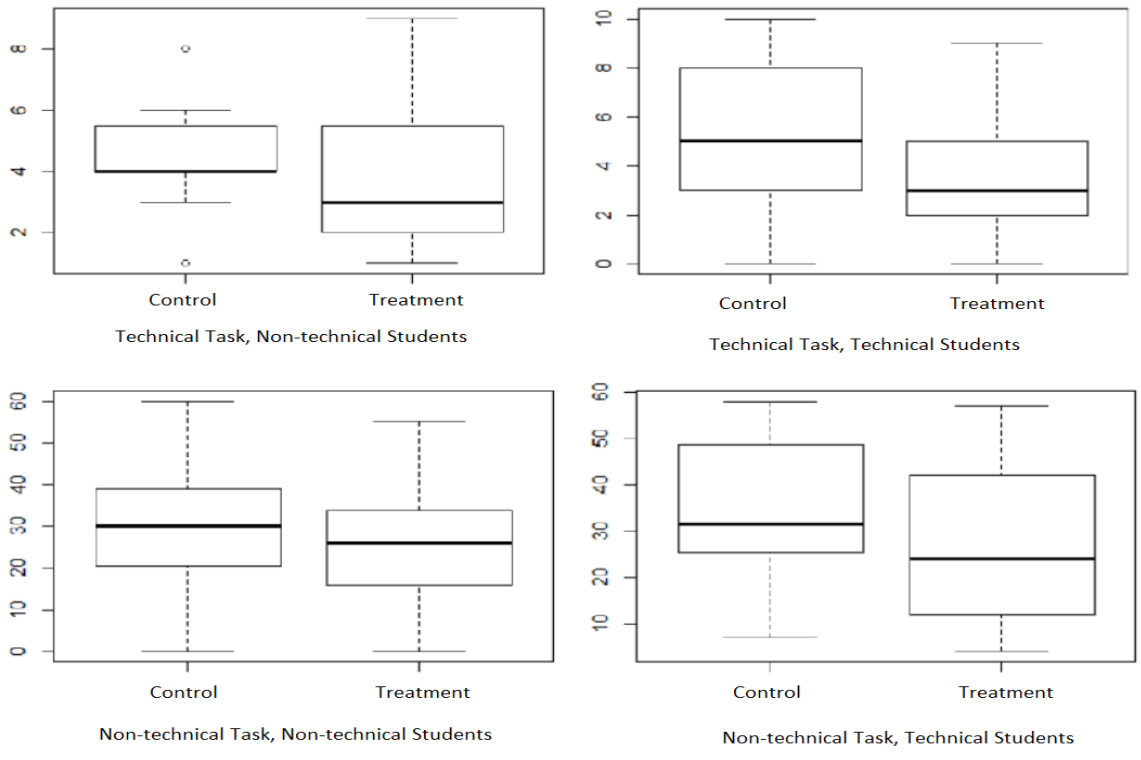
Table 2 presents the output from the Mann-Whitney U tests performed on the different control groups – treatment groups' pairs in order to investigate the magnitude, direction and significance of the treatment effect. Figure 2 provide a graphical representation of the pairs comparison with box plots.

**Table 2**. Mann-Whitney U Test comparisons regarding Sudoku scores and word puzzle scores.

| Score | Group | Student Type | N | Rank Sum | U | P-value |
|---|---|---|---|---|---|---|
| Sudoku non-rewarded | Control | Technical | 20 | 470 | 260 | 0.103 |
| Sudoku rewarded | Treatment | Technical | 20 | 350 | | |
| Sudoku non-rewarded | Control | Non-Technical | 20 | 463.5 | 253.5 | 0.147 |
| Sudoku rewarded | Treatment | Non-Technical | 20 | 356.5 | | |
| Word puzzle non-rewarded | Control | Technical | 20 | 463 | 253 | 0.155 |
| Word puzzle rewarded | Treatment | Technical | 20 | 357 | | |
| Word puzzle non-rewarded | Control | Non-Technical | 20 | 439.5 | 229.5 | 0.432 |
| Word puzzle rewarded | Treatment | Non-Technical | 20 | 380.5 | | |

The table shows how the test did not provide any evidence in favor of a significant treatment effect. None of the p-values results smaller than the conventional value of .05, which leads not to reject the null hypothesis of medians identity (p-values are equal respectively to .15, .10, .43 and .16). If anything, a comparison of the mean ranks anticipates what Figure 2 shows more clearly: the extrinsically rewarded students performed worse than their non-incentivized counterparts did, albeit not to a significant extent.

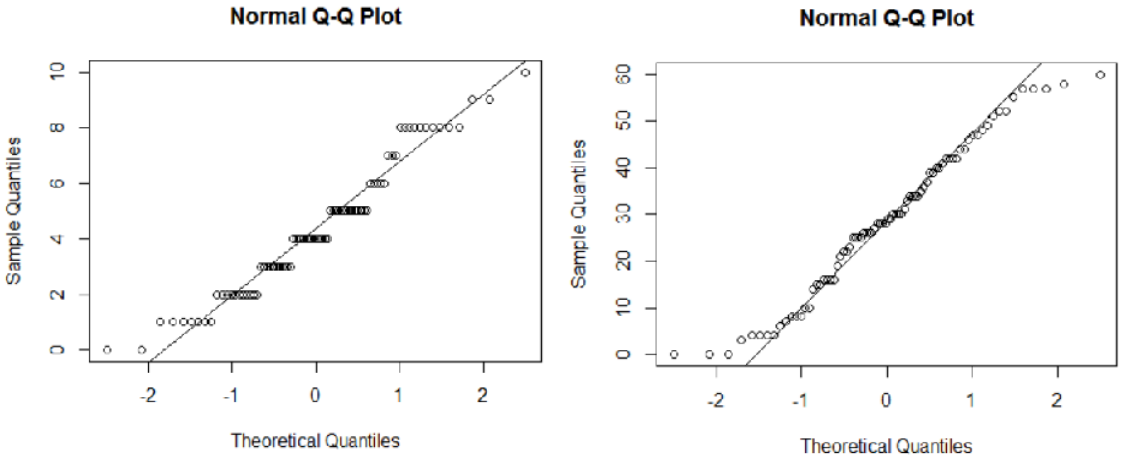
**Figure 2**. Box plots of the comparisons regarding Sudoku scores and word puzzle scores.



### 4.3.2 Two-way ANOVA

The Analysis Of Variance, even in its two-way form, is considered a parametric statistical tool. As such, it requires for the data it is performed on to satisfy the normality assumption, which refers to the distribution of errors. In order to check for this assumption, I performed Shapiro-Wilk Tests for Normality separately for Sudoku scores and word puzzle scores. Table 3 in Appendix 9 contains the output of the tests, while Figure 3 and Figure 4 below provide the Q-Q plots for both scores. The Shapiro-Wilk test for normality has as null hypothesis that the sample is drawn from a normally distributed population. As regarding the Sudoku scores, the p-value is far below the conventionally accepted alpha value of 0.05 (p-value = .007), which should induce to reject the normality assumed in the null hypothesis. Nevertheless, the Q-Q plot suggests that the greater part of the responsibility for such an un-doubtful result in term of p-value should be addressable to the fact that the possible scores obtainable only took integer values from 0 to 10. Indeed, when considering this aspect, the plot seems to be consistent with the Q-Q line. As regarding the word puzzle score instead, the p-value from the Shapiro Wilk test is above 0.05, and equals .07, thus not allowing the rejection of normality stated in the null hypothesis. The Q-Q plots visually confirms the results.

**Figures 3 and 4.** Q-Q Plots respectively for Sudoku Scores and Word Puzzles Scores.




Sudoku Scores

Word puzzle Scores

The two-way version of the ANOVA suits better than the one-way version in case there is the suspect that more than one factor could be affecting the observed variability in outcome between two samples. This is precisely the case for the current analysis. More specifically, in this part of the analysis I wanted to investigate both the effects of the treatment (extrinsic reward) and the students'

background on the score. ANOVA also allows to include interaction terms, which was particular interesting in this case.

Tables 4 and 5 present the results of the two-way ANOVA of both scores.

**Table 4**. The effects of students' background, of extrinsic reward and the interaction term on Sudoku score. Two-way ANOVA.

| Source | df | Sum of Square | Mean of Square | F value | P-value |
|---|---|---|---|---|---|
| Technical student | 1 | 1.0 | 1.012 | 0.182 | 0.670 |
| Incentivized | 1 | 23.1 | 23.112 | 4.166 | 0.045* |
| Interaction | 1 | 2.1 | 2.112 | 0.381 | 0.539 |
| Residuals | 76 | 421.6 | 5.548 | | |

**Table 5**. The effects of student's background, of extrinsic reward and the interaction term on word puzzle score. Two-way ANOVA.

| Source | df | Sum of Square | Mean of Square | F value | P-value |
|---|---|---|---|---|---|
| Technical student | 1 | 300 | 300.3 | 1.164 | 0.284 |
| Incentivized | 1 | 515 | 515.1 | 1.997 | 0.162 |
| Interaction | 1 | 86 | 86.1 | 0.334 | 0.565 |
| Residuals | 76 | 19,608 | 258.0 | | |

Table 4 shows that as far as Sudoku scores are concerned, neither the background of the student nor the interaction term significantly affect the difference in outcomes (p-values are respectively equal to .67 and .54), while the presence of the incentive appears to have a significant effect (p-value is equal to .04). This result can be interpreted as evidence in favor of the crowding out effect, as it was shown in the Section 3.5 that the scores of the extrinsically rewarded subjects on the Sudoku task were lower than their non-rewarded counterparts were. Nevertheless, still no evidence appears in favor of a different extent of the crowding-out being narrowed down in case of a reward paid on a task coherent with the background of a subject.

Table 5 presenting the output for the word puzzle score tells an even more extreme story. Indeed none of the variables results statistically significant (p-values are equal to .28, .16 and .57). This mean that neither the treatment, the students' background or the interaction term explain the difference between the two samples. These results not only do not provide evidence in favor of the hypothesis at the core of this paper, but they do not even provide evidence in favor of the motivation crowding-out theory.

### 4.3.3 Regression analysis

Before performing the regression analysis I check the OLS assumptions for the estimators to be BLUE (Best Linear Unbiased Estimator). Best here means that the variance of the OLS estimator d'b for d'β has a lower variance than any other linear unbiased estimator does. In order to determine if the estimators are BLUE, the Gauss-Markov assumptions are checked:

1. The model is linear in its parameters.
2. There is a random sample of observations.
3. There exists no perfect collinearity amongst the independent variables.
4. The error term has an expected value of zero given any values of the independent variable.
5. The error term has the same variance given any values of the explanatory variables.

The correlation coefficient matrix presented in Figure 5 (see the Figures section after the appendices) excludes the presence of multicollinearity. None of the couples of variables indeed presents correlation above 0.6. Figure 6 (again see the Figures section after the appendices), plotting the standardized residuals over the fitted values from the regression model as specified in model 1 (see below) suggests some minor heteroscedasticity concerns that still leave the unbiasedness of the estimators safe. Following, the different specifications of the regression model.

Firstly, I implemented the mere additive model with the main explanatory variables. The specification of model 1 is the following:

$$y = \alpha + \beta_1 x_{\text{technical task}} + \beta_2 x_{\text{technical student}} + \beta_3 x_{\text{rewarded}} + \beta_4 x_{\text{experience}} + \beta_5 x_{\text{age}} + \varepsilon$$

The dependent variable y consists of the standardized scores of the two tasks (all the Sudoku scores have been standardized together, and all the Wordpuzzle scores have been standardized together, so that the resulting scores could be merged together and be compared). Beside the intercept, the first explanatory variable is a dummy, and provides info regarding whether the score presented is that of a technical task or of a non-technical task (x=0, the reference category,

defines a non-technical task). The second explanatory variable is another dummy, and provides information on the background of the student: the reference category x=0 stands for non-technical students. The third one is again a dummy taking the values 1 and 0 as the previous ones, and the reference category x=0 stands for the student being non extrinsically rewarded, while the value x=0 stands for the student receiving the treatment, namely the monetary reward. The fourth explanatory variable is a discrete variable taking the integer values between 1 and 5, providing information on the self-reported experience. Interestingly enough, it does not only provide information on how important previous experience is on the outcome of the task, but it could also be seen as a proxy for intrinsic motivation. Indeed, since Sudoku is mainly a recreational activity, it can be assumed that subjects who already practiced Sudoku before the experiment did so because they enjoy it. The last explanatory variable is age. The reason why I did not opt for the inclusion of the square of age, a quite common behavior in the vast majority of social sciences papers which have age within the dataset, is that the min-max range is very narrow (between 20 and 29, where 29 is an outlier). The difference between the linear and quadratic relation would consequently be highly unlikely to affect the results.

When looking at the interpretation of the coefficients, it is important to remember how the reference categories were defined. This means that the parameter $\alpha$ provides the average score for the non-technical task, performed by a non-technical student, non-extrinsically rewarded, with no experience on the task and of hypothetical age of zero.
The other parameters instead refer to the dummies, and measure the difference in score with the relative baseline categories. Thus $\beta_1$ denotes the difference between the scores on technical task and non-technical task for all subjects. In the same fashion, $\beta_2$ denotes the difference between the scores of technical students compared to non-technical students. $\beta_3$ describes the difference in score between incentivized and non-incentivized subjects. The coefficient $\beta_4$ provides information on the effect of experience on the score, and $\beta_3$ of the effect of age on the score.
As column (1) in Table 6 below shows, the scores appear to be slightly lower for technical tasks, for incentivized students and for older student. It is worth mentioning that the coefficient for the treatment is negative, albeit not significant, which barely supports the crowding-out theory. The negative sign of the age coefficient might be due to older students possibly being those who failed exams for more times, thus being forced to be still enrolled, and it might be that they are less “skilled” for these types of tasks. Never the less, it is important to notice that only the experience coefficient appears to be statistically significant and positive. Furthermore every unitary increase on the experience scale – as defined above – corresponds to almost 47% increase in the score.
The model itself can explain, according to the adjusted $R^2$, 14% of the total variance in the score, which

for social sciences does not appear to be a particularly low coefficient. Some doubts arise by the distribution of residuals, very high at the skews.

In the following models, I gradually introduced different interaction terms.
In model 2 an interaction term was added, describing the effect that belonging to the "technical students" category has on the technical task score. The specification of the model is the following:

$$y = \alpha + \beta_1 x_{\text{technical task}} + \beta_2 x_{\text{technical student}} + \beta_3 x_{\text{rewarded}} + \beta_4 x_{\text{experience}} + \beta_5 x_{\text{age}} + \beta_6 x_{\text{technical task}} x_{\text{technical student}} + \varepsilon$$

On the contrary of what is expected from the hypothesis, the coefficient of the interaction term is negative, albeit not significant (p-value equal to .59), showing that technical students performed worse than their non-technical colleagues did on the task belonging to their own field of expertise. Interestingly enough, the sign of the "technical task" coefficient, $\beta_1$ became positive. Considering that the total $\Delta$ of the coefficient amounts to 0.07, and the coefficient does not become significant, it does not raise particular concerns. The reason of this change should be addressed to the fact that technical students are the ones who performed worse on technical tasks, and now that the interaction term captures this feature, the $\beta_1$ coefficient is no more including this hidden dynamic.
Also, the fact that the adjust $R^2$ decreases slightly and the distribution of the residuals becomes even more skewed, suggests that this interaction term does not add any explicative value to the model.

In model 3 another interaction term is added, describing the effect that being incentivized had on technical task's score. The specification of the model is the following, and below column (3) from Table 6 presents the output summary:

$$y = \alpha + \beta_1 x_{\text{technical task}} + \beta_2 x_{\text{technical student}} + \beta_3 x_{\text{rewarded}} + \beta_4 x_{\text{experience}} + \beta_5 x_{\text{age}} + \beta_6 x_{\text{technical task}} x_{\text{technical student}} + \beta_6 x_{\text{technical task}} x_{\text{incentivized}} + \varepsilon$$

It appears from the output that the incentives had a worse effect on technical tasks compared to non-technical tasks, but once again, the coefficient is non-significant. The incentivized coefficient, $\beta_3$, is halved by the introduction of this interaction term, meaning that part of the negative effect of the incentive is now captured by this interaction term. The $\beta_1$ coefficient for technical task was instead raised by 0.10, showing the negative effect had on the technical task score. The adjusted $R^2$ decreased slightly, while the distribution of the residuals on the skews dropped to an even lower amount than the original additive model.

In the last model, I introduced a third interaction term in order to describe the effect of incentives on technical students' scores. Below, the specification and the output summary in column (4) from Table 6:

$$y = \alpha + \beta_1 x_{\text{technical task}} + \beta_2 x_{\text{technical student}} + \beta_3 x_{\text{rewarded}} + \beta_4 x_{\text{experience}} + \beta_5 x_{\text{age}} + \beta_6 x_{\text{technical task}} x_{\text{technical student}} + \beta_6 x_{\text{technical task}} x_{\text{incentivized}} + \beta_6 x_{\text{technical student}} x_{\text{incentivized}} \varepsilon$$

To a non-statistically significant extent, it seems like the effect was worse than on the non-technical ones. The introduction of this term also decreased substantially the coefficient for incentives. The adjusted $R^2$ dropped by another small amount.

Table 6 provides a summary of all the specifications exposed above (see next page).

**Table 6.** Summary of the different specifications of the regression models.

Dependent Variable: Standardized task score.

| **Parameter** | **(1)** | **(2)** | **(3)** | **(4)** |
|---|---|---|---|---|
| Intercept | 0.683 | 0.643 | 0.589 | 0.553 |
| | (0.483) | (0.511) | (0.549) | (0.578) |
| Technical Task | -0.030 | 0.049 | 0.146 | 0.146 |
| | (0.837) | (0.814) | (0.565) | (0.566) |
| Technical Student | 0.235 | 0.313 | 0.313 | 0.352 |
| | (0.233) | (0.202) | (0.204) | (0.217) |
| Extrinsic Reward | -0.204 | -0.204 | -0.105 | -0.066 |
| | (0.187) | (0.189) | (0.624) | (0.798) |
| Experience | 0.467*** | 0.467*** | 0.470*** | 0.467*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Age | -0.064 | -0.064 | -0.064 | -0.063 |
| | (0.167) | (0.168) | (0.169) | (0.179) |
| Technical Task * Technical Student | _ | -0.158 | -0.158 | -0.158 |
| | | (0.591) | (0.591) | (0.592) |
| Technical Task * Incentivized | _ | _ | -0.195 | -0.195 |
| | | | (0.506) | (0.508) |
| Technical Student * Incentivized | _ | _ | _ | -0.082 |
| | | | | (0.782) |
| Adj. $R^2$ | 0.143 | 0.140 | 0.136 | 0.131 |
| F-statistic | 6.325 | 5.295 | 4.586 | 3.998 |
| N | 160 | 160 | 160 | 160 |

Notes: P-values in parentheses.

# 5. Discussion and Conclusion

## 5.1 Key Findings

The evidence resulting from the statistical analysis presented within section 4 does not support the hypothesis of a positive effect of a coherent match between the nature of a task and the background of the subject on the performance. Consequently, the answer to the research question seems to be simply: no, they are not. If anything, the result that happened to emerge from the analysis points in support of the motivation crowding-out theory. Indeed, the incentivized groups performed on average worse than their non-incentivized counterparts did, albeit according to the majority of the tests even this result did not reach a statistically significant extent.

Evidence resulting from this specific experiment did not provide evidence in favor of the hypothesis setting, in the next session I will discuss some of the limitations regarding the experimental set up and the analysis that might have contributed to this outcome.

## 5.2 Limitations

The analysis performed on the data obtained by the experiment failed to reject the null hypothesis. Primarily, this should induce not to accept the hypothesis suggested in this paper. Nevertheless, following are some limitations that might have also played a role in leading to this non-acceptance of the hypothesis, or at least provided the experiment and analysis presented with some opportunities for improvements. Recognizing these limitations does not by itself constitute a solution to those, but firstly it is important in order to suggest future research paths, and secondly it provides further guidance in the overall understanding of the results.

Firstly, I will present the limitations regarding the experimental set up. I will start by the nature of the task and environment: as it was not possible for me to have access to a company's human resources department, which would have allowed me to manipulate the actual payment schemes of real employees in their work environment, I had to find a second-best solution. As discussed above, I opted for university students in order to get as close as possible to a sample of people with a strong background identity, which in this case was academic instead of professional. Dealing with students and not with workers, made it also more complicated to find a pair of tasks that were easily recognizable for them as belonging to their academic background, while still being accessible to the other sample of students, and could be corrected quickly and objectively by me, not belonging to neither of those fields of studies. Thus, one could argue that some of the students could have perceived

both tasks as in general pertaining to the sphere of the games, and not as something easily identifiable as part of their background. In order to limit this eventuality, I framed the task by introducing the one matching the background of the students receiving the card as "one for which it is likely you developed some specific competences during your course of studies". Approaching the students during lectures was meant to contribute to same goal. A very specific reason why not being in a real working situation might be particular detrimental to this paper is that it is hard to recreate out of a work environment the precise mix of intrinsic and extrinsic motivation driving actions, and that is exactly the matter of the current study.

A second limitation to the experimental set up is one shared with the vast majority of the literature studying incentives: limited funding. As discussed in section 3.1, for incentives to serve the purpose of providing control, the precept of dominance must be met, which implies that the rewards must be high enough to induce the subjects to exert effort in tasks. This aspect plays an even more important role in studies, as the present, which investigate the interrelations between intrinsic and extrinsic motivation. In facts, there is more to a reward than the monetary – or in general extrinsic – value itself. As discussed in Gneezy and Rustichini (2000a), a reward or a fine to low could be perceived as insulting, and provoke the subjects of an experiment in ways deviating from the utility maximization through the task itself. For example, they might be willing to prove to disregard such a low reward, and thus act on purpose below their actual potential.

A third limitation might concern the heterogeneity between the groups of technical and non-technical students, in terms of age (the IT students were master students, while the Law students were attending their first year of bachelor) and experience on the different tasks. Albeit not an ideal condition, this could not affect the first part of the analysis, which involved pairs comparisons within the same academic background, and the fact that the later part of the analysis gave results in line with that part, heterogeneity between groups of student should not be a major concern. A similar argument could arise on the difference between room structures. As a final remark on these limitations, it is worth mentioning that these were the results of the time constraints of my permanence in Italy, together with banking holidays and some problems with some of the university rooms' heating system leaking, which forced the professors to reschedule the lectures in other rooms.

Another, quite minor, limitation, pertains part of the methodology. More specifically, as in further detail explained in section 4, the normality assumption required for ANOVA was not met in an orthodox way by the Sudoku scores. This was due to the fact that the variable "Sudoku Scores" was not a continuous variable, and indeed could only take the integer values between "0" and "10".

Nevertheless, the reason why I consider this a minor issue is that I performed the ANOVA after not finding significant within-samples differences, and provided results in line with the previous.

## 5.3 Running the Experiment: Some Anecdotes

Running this experiment was my first practical experience as far as experimental economics is concerned. It was thrilling and made me enthusiastic in all of the phases. In particular, in this section I will share some anecdotes regarding the field phase – the one that involved direct contact with the subjects –, which I deem relevant for a more complete understanding of the results.

More specifically, it made me proud to have students (and the professors who allowed me to take the sessions at the beginning of their lectures) following me after the experimental sessions in the corridors, asking about the experiment. Their interest lead to interesting discussions. Firstly, all of the students who approached me asked me about the goal of the experiment, none of them had guessed. This reinsured me regarding the absence of any possible desirability bias in their behavior. Furthermore, some of them pointed out that even if they did read the introduction where the task corresponding to their background was framed as a task of their competence, they did not quite feel like that. Some others said they did. This might push toward the research of a more suiting pair of task, or a more incisive framing.

## 5.4 Practical Implications

The results of the analysis did not support the hypothesis. If that would have been the case, it would have provided an insight to the payment schemes designers suggesting to make sure that the employees perceived the tasks they were incentivized for as belonging to their field of expertise. This indeed would have allowed for the extrinsic rewards to be perceived as identity reinforcing instead of controlling, thus lowering the possibility of a motivation crowding-out. Nevertheless, the results from the current analysis provide some pale evidence in support of the classical crowding-out theory. The experiment at the core of this paper was designed specifically in order to analyze a specific aspect of that theory, in relation to the cognitive evaluation theory. For this reason it might not be the wiser option to provide practical suggestion on the application of other aspects of the motivation crowding-out theory building on the result of this specific experiment. Nevertheless, a general remark I feel confident to extend as an externally valid suggestion to practitioners is to design rewards with high enough monetary stakes, in order to avoid any possible behavior deviating from high effort.

## 5.5 Future Research

As discussed in section 5.2, many limitations affected this empirical research. As such, they suggest the first opportunities for improvement in further research. Firstly, in order to more precisely recreate the actual mix of intrinsic and extrinsic motivation driving workers behavior under monetary rewards, an actual field experiment would fit better. A proper design would include subjects whose assigned job includes two tasks, one of them being significantly more linked to the workers background. The task should be of comparable difficulty, and the precepts mentioned in section 3.1 should still hold. Secondly, for further research to move toward a more effective research design, it will be important to obtain larger funds for this kind of experiments. The reason for this have been vastly argued throughout the previous sections. A third direction for further research to pursuit is a higher extent of homogeneity among the subject groups in terms of all of the characteristics that could affect the level of performance of the subjects in the task assigned. This is a fundamental aspect in order for the analysis on the treatment to satisfy the *ceteris paribus* condition. Lastly, the last direction suggested for future research comes directly from the results of the analysis. The hypothesis tested was not accepted. It was an attempt to apply the cognitive evaluation theory's explanation of the motivation crowding out phenomenon. A challenge for future research will thus be to find other possible applications of that explanation. As from this analysis it emerged that matching tasks with the field of expertise of the subject was not an effective way to narrow down the crowding out effect, it will be worth for further research trying to explore other aspect of incentives design that, based on the cognitive evaluation theory, could.

## 5.6 Conclusion

The empirical analysis conducted within this paper, based on the data obtain with a framed field experiment involving students, had the aim to investigate the following research question: *Are subjects more positively responsive to extrinsic rewards when these are paid on tasks they perceive as belonging to their field of expertise?* This research question stemmed from the cognitive evaluation theory, more specifically from the major role that this theory attributes to the perception of an extrinsic reward as controlling or identity-reinforcing, when it comes to enhancing or reducing performance. In order for monetary rewards to effectively extrinsically motivate agents they need to be perceived as identity-enhancing and not controlling, otherwise they will crowd-out intrinsic motivation and will result in lower performance. The analysis did not provide evidence in support of the hypothesis, which consequently is rejected. Some limitations of the experimental set up might have affected the results of the analysis, thus suggesting hints for the directions further research might follow. A field experiment with actual workers on their usual tasks has been suggested as a possible better suiting

set up. More in general, rejecting this hypothesis means it is necessary to explore other possible applications of the cognitive evaluation theory's explanation of the motivation crowding out theory.

# 6. References


Arrow, K. J. (1972). *Gifts and Exchanges*. Philosophy and Public Affairs 1 (Summer): 343-362

Bem, D. J. (1967a). *Self-Perception: The Dependent Variable of Human Performance.* Organizational Behavior and Human Performance 2, 105-121

Bem, D. J. (1967b). *Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena*. Psychological Review 74, 183-200.

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. *Experimental economics*, *11*(3), 268-281.

Cleave, B. L., Nikiforakis, N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics*, *16*(3), 372-382.

Deci, E. L. (1971). *The Effects of Externally Mediated Rewards on Intrinsic Motivation*. Journal of Personality and Social Psychology 18, 105-115

Deci, E. L. and Ryan, R. M. (1980). *The Empirical Exploration of Intrinsic Motivational Processes*. In L. Berkowitz (Ed.) Advances in Experimental Social Psychology 13, 39-80. New York: Academic Press.

Deci, E. and Ryan, R.M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press

Fehr, E., & Falk, A. (2002). *Psychological foundations of incentives*. European economic review, 46(4), 687-724.

Fiske, A. P. (1992). *The four elementary forms of sociality: framework for a unified theory of social relations*. *Psychological review*, *99*(4), 689.

Frey, B. S., Oberholzer-Gee, F. and Eichenberger, R. (1996). *The Old Lady Visits Your Backyard: A Tale of Morals and Markets.* Journal of Political Economy 104 (6), 12971313.

Frey, B. S. and Oberholzer-Gee, F. (1997). *The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding Out*. American Economic Review 87, 746-755

Frey, B. S., & Jegen, R. (2000). *Motivation Crowding Theory: A Survey of Empirical Evidence*.

Gneezy, U. and Rustichini, A. (2000a.) *A Fine is a Price.* Journal of Legal Studies 29, 1-17.

Gneezy, U. and Rustichini, A. (2000b). *Pay Enough or Don't Pay at All*. Quarterly Journal of Economcis 115(2), 791-810

Heyman, J., & Ariely, D. (2004). *Effort for payment: A tale of two markets*. Psychological Science, 15, 787–793. doi:10.1111/j.0956-7976.2004.00757.x

Huffman, D., & Bognanno, M. (2015). *Performance Pay and Workers' Non-Monetary Motivations: Evidence from a Natural Field Experiment*.

Kreps, D. M. (1997). *Intrinsic motivation and extrinsic incentives*. The American Economic Review, 87(2), 359-364.

Kruglanski, A. W., Friedman, I., & Zeevi, G. (1971). *The effects of extrinsic incentive on some qualitative aspects of task performance*. Journal of Personality, *39*(4), 606-617.

Lepper, M.R., Greene, D. and Nisbet, R. E. (1973). *Undermining Children's Intrinsic Interest with Extrinsic Rewards: A Test of the "Over Justification" Hypothesis*. Journal of Personality and Social Psychology 28, 129-137

Promberger, M., & Marteau, T. M. (2013). *When do financial incentives reduce intrinsic motivation? comparing behaviors studied in psychological and economic literatures*. *Health Psychology*, *32*(9), 950.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, *72*(5), 923-955.

Solow, R. S. (1971). Blood and Thunder. *Yale Law Journal* 80: 170-083

Titmuss, R. (1970). *The gift relationship: From human blood to social policy*. London, UK: George Allen and Unwin.

# 7. Appendices

This section presents all the different versions (in English) of the cards used during the experiment.

**Appendix 1:** Non-technical students, incentivized, version 1: Sudoku first.



**Below a numerical task, ten short versions of the Sudoku scheme.**

**You have 5 minute to solve as many as you can, and will be paid 0,50 € for every scheme correctly completed. The money will be paid in cash at the end of the lecture.**

Please state your age _____

How frequently do you play Sudoku, if ever?
(This answer will not affect your compensation)

1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | |
|---|---|---|---|
| | | | 3 |
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | | |
|---|---|---|---|
| | | 1 | |
| 4 | | | |
| | | | 2 |
| | 3 | | |

| | | | |
|---|---|---|---|
| 2 | | | |
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | 1 |
| 3 | | | |
| | | | 4 |
| | | | |

| | | | |
|---|---|---|---|
| 1 | | | 3 |
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | | |
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | | | |
|---|---|---|---|
| | 1 | 4 | |
| | 3 | | |
| | | 2 | |
| | | | |

| | | | |
|---|---|---|---|
| 3 | | 2 | |
| | | | |
| | | | |
| | 1 | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | |
| | | | |
| | 3 | | |
| | | | 2 |

**Below a verbal task, a word puzzle scheme. Being a student with a strong verbal background, it is likely you developed a high degree of expertise in verbal tasks like this.**

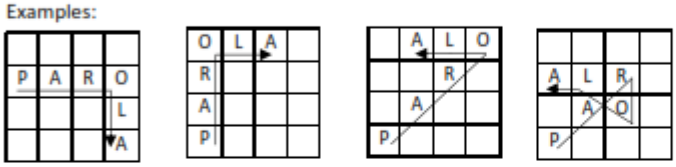
**You have 5 minute to find as many words as you can. The total of the words you will find will be counted, and you will be paid 0,10 € for every 3 letters words found, 0,15 € for every 4 letters word, 0,20 € for 5 letters, 0,25 € for 6 letters, 0,30 € for 7 letters, 0,35 € for 8 letters, 0,40 € for 9 letters words, 0,50 € for 10 letters words.**
**The money will be paid in cash at the end of the lecture.**

Please state your age _____

How frequently do you play word puzzles, if ever? (This answer will not affect your compensation) ☐ 1 Never ☐ 2 ☐ 3 ☐ 4 ☐ 5 Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:




| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____

**Appendix 2:** Non-technical students, incentivized, version 2: word puzzle first.



**Below a verbal task, a word puzzle scheme. Being a student with a strong verbal background, it is likely you developed a high degree of expertise in verbal tasks like this.**

**You have 5 minute to find as many words as you can. The total of the words you will find will be counted, and you will be paid 0,10 € for every 3 letters words found, 0,15 € for every 4 letters word, 0,20 € for 5 letters, 0,25 € for 6 letters, 0,30 € for 7 letters, 0,35 € for 8 letters, 0,40 € for 9 letters words, 0,50 € for 10 letters words.**
**The money will be paid in cash at the end of the lecture.**

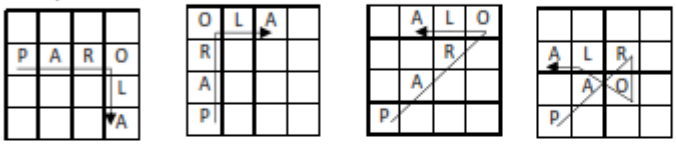
Please state your age _____

How frequently do you play word puzzles, if ever? (This answer will not affect your compensation)

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ☐ Never | ☐ | ☐ | ☐ | ☐ Daily |

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:




| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Below a numerical task, ten short versions of the Sudoku scheme.**

**You have 5 minute to solve as many as you can, and will be paid 0,50 € for every scheme correctly completed. The money will be paid in cash at the end of the lecture.**

Please state your age _____

How frequently do you play Sudoku, if ever? (This answer will not affect your compensation)

1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | 4 | |
|---|---|---|---|
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | 3 |
|---|---|---|---|
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | 1 | |
|---|---|---|---|
| 4 | | | |
| | | | 2 |
| | 3 | | |

| 2 | | | |
|---|---|---|---|
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | 4 | | 1 |
|---|---|---|---|
| 3 | | | |
| | | | 4 |
| | | | |

| 1 | | | 3 |
|---|---|---|---|
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | 1 | 4 | |
|---|---|---|---|
| | 3 | | |
| | | 2 | |
| | | | |

| 3 | | 2 | |
|---|---|---|---|
| | | | |
| | | | |
| | 1 | | 4 |

| | 4 | | |
|---|---|---|---|
| | | | |
| | 3 | | |
| | | | 2 |

**Appendix 3:** Non-technical students, non-incentivized, version 1: Sudoku first.



**Below a numerical task, ten short versions of the Sudoku scheme.**

**You have 5 minute to solve as many as you can.**

Please state your age _____

How frequently do you play Sudoku, if ever?

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |
| Never | | | | Daily |

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | |
|---|---|---|---|
| | | | 3 |
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | | |
|---|---|---|---|
| | | 1 | |
| 4 | | | |
| | | | 2 |
| | 3 | | |

| | | | |
|---|---|---|---|
| 2 | | | |
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | 1 |
| 3 | | | |
| | | | 4 |
| | | | |

| | | | |
|---|---|---|---|
| 1 | | | 3 |
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | | |
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | | | |
|---|---|---|---|
| | 1 | 4 | |
| | 3 | | |
| | | 2 | |
| | | | |

| | | | |
|---|---|---|---|
| 3 | | 2 | |
| | | | |
| | | | |
| | 1 | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | |
| | | | |
| | 3 | | |
| | | | 2 |

**Below a verbal task, a word puzzle scheme. Being a student with a strong verbal background, it is likely you developed a high degree of expertise in verbal tasks like this.**

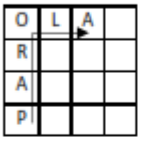
**You have 5 minute to find as many words as you can in the griddle.**

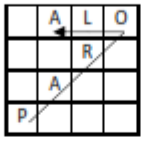
Please state your age _____

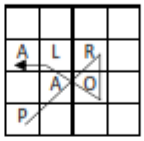
How frequently do you play word puzzles, if ever? 1 ☐ Never 2 ☐ 3 ☐ 4 ☐ 5 ☐ Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:


| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Appendix 4:** Non-technical students, non-incentivized, version 2: word puzzle first.



**Below a verbal task, a word puzzle scheme. Being a student with a strong verbal background, it is likely you developed a high degree of expertise in verbal tasks like this.**

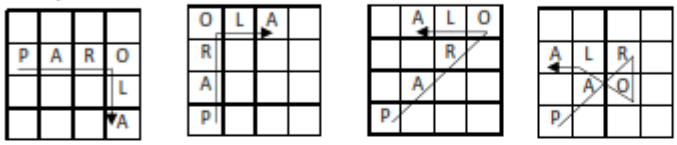
**You have 5 minute to find as many words as you can in the griddle.**

Please state your age _____

How frequently do you play word puzzles, if ever? 1 ☐ Never 2 ☐ 3 ☐ 4 ☐ 5 ☐ Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:




| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Below a numerical task, ten short versions of the Sudoku scheme.**

**You have 5 minute to solve as many as you can.**

Please state your age _____

How frequently do you play Sudoku, if ever?

1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | 4 | |
|---|---|---|---|
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | 3 |
|---|---|---|---|
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | 1 | |
|---|---|---|---|
| 4 | | | |
| | | | 2 |
| | 3 | | |

| 2 | | | |
|---|---|---|---|
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | 4 | | 1 |
|---|---|---|---|
| 3 | | | |
| | | | 4 |
| | | | |

| 1 | | | 3 |
|---|---|---|---|
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | 1 | 4 | |
|---|---|---|---|
| | 3 | | |
| | | 2 | |
| | | | |

| 3 | | 2 | |
|---|---|---|---|
| | | | |
| | | | |
| | 1 | | 4 |

| | 4 | | |
|---|---|---|---|
| | | | |
| | 3 | | |
| | | | 2 |

**Appendix 5:** Technical students, incentivized, version 1: Sudoku first.



**Below a numerical task, ten short versions of the Sudoku scheme. Being a student with a strong numerical background, it is likely you developed a high degree of expertise in numerical tasks like this.**

**You have 5 minute to solve as many as you can, and will be paid 0,50 € for every scheme correctly completed. The money will be paid in cash at the end of the lecture.**

Please state your age _____

How frequently do you play Sudoku, if ever?
(This answer will not affect your compensation)

1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | |
|---|---|---|---|
| | | | 3 |
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | | |
|---|---|---|---|
| | | 1 | |
| 4 | | | |
| | | | 2 |
| | 3 | | |

| | | | |
|---|---|---|---|
| 2 | | | |
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | 1 |
| 3 | | | |
| | | | 4 |
| | | | |

| | | | |
|---|---|---|---|
| 1 | | | 3 |
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | | |
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | | | |
|---|---|---|---|
| | 1 | 4 | |
| | 3 | | |
| | | 2 | |
| | | | |

| | | | |
|---|---|---|---|
| 3 | | 2 | |
| | | | |
| | | | |
| | 1 | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | |
| | | | |
| | 3 | | |
| | | | 2 |

**Below a verbal task, a word puzzle scheme.**

**You have 5 minute to find as many words as you can. The total of the words you will find will be counted, and you will be paid 0,10 € for every 3 letters words found, 0,15 € for every 4 letters word, 0,20 € for 5 letters, 0,25 € for 6 letters, 0,30 € for 7 letters, 0,35 € for 8 letters, 0,40 € for 9 letters words, 0,50 € for 10 letters words.**
**The money will be paid in cash at the end of the lecture.**

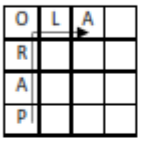
Please state your age _____

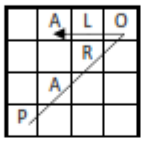
How frequently do you play word puzzles, if ever? (This answer will not affect your compensation)

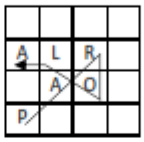
1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:


| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Appendix 6:** Technical students, incentivized, version 2: word puzzle first.



**Below a verbal task, a word puzzle scheme.**

**You have 5 minute to find as many words as you can. The total of the words you will find will be counted, and you will be paid 0,10 € for every 3 letters words found, 0,15 € for every 4 letters word, 0,20 € for 5 letters, 0,25 € for 6 letters, 0,30 € for 7 letters, 0,35 € for 8 letters, 0,40 € for 9 letters words, 0,50 € for 10 letters words.**
**The money will be paid in cash at the end of the lecture.**

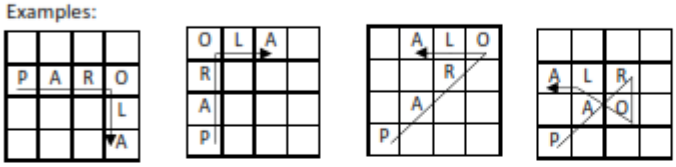
Please state your age _____

How frequently do you play word puzzles, if ever? (This answer will not affect your compensation)

1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:




| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Below a numerical task, ten short versions of the Sudoku scheme. Being a student with a strong numerical background, it is likely you developed a high degree of expertise in numerical tasks like this.**

**You have 5 minute to solve as many as you can, and will be paid 0,50 € for every scheme correctly completed. The money will be paid in cash at the end of the lecture.**

Please state your age _____

How frequently do you play Sudoku, if ever? (This answer will not affect your compensation)

1 ☐ Never  2 ☐  3 ☐  4 ☐  5 ☐ Daily

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | 4 | |
|---|---|---|---|
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | 3 |
|---|---|---|---|
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | 1 | |
|---|---|---|---|
| 4 | | | |
| | | | 2 |
| | 3 | | |

| 2 | | | |
|---|---|---|---|
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | 4 | | 1 |
|---|---|---|---|
| 3 | | | |
| | | | 4 |
| | | | |

| 1 | | | 3 |
|---|---|---|---|
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | 1 | 4 | |
|---|---|---|---|
| | 3 | | |
| | | 2 | |
| | | | |

| 3 | | 2 | |
|---|---|---|---|
| | | | |
| | | | |
| | 1 | | 4 |

| | 4 | | |
|---|---|---|---|
| | | | |
| | 3 | | |
| | | | 2 |

**Appendix 7:** Technical students, non-incentivized, version 1: Sudoku First



**Below a numerical task, ten short versions of the Sudoku scheme. Being a student with a strong numerical background, it is likely you developed a high degree of expertise in numerical tasks like this.**

**You have 5 minute to solve as many as you can.**

Please state your age _____

How frequently do you play Sudoku, if ever?

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |
| Never | | | | Daily |

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | |
|---|---|---|---|
| | | | 3 |
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | | |
|---|---|---|---|
| | | 1 | |
| 4 | | | |
| | | | 2 |
| | 3 | | |

| | | | |
|---|---|---|---|
| 2 | | | |
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | 1 |
| 3 | | | |
| | | | 4 |
| | | | |

| | | | |
|---|---|---|---|
| 1 | | | 3 |
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | | |
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | | | |
|---|---|---|---|
| | 1 | 4 | |
| | 3 | | |
| | | 2 | |
| | | | |

| | | | |
|---|---|---|---|
| 3 | | 2 | |
| | | | |
| | | | |
| | 1 | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | |
| | | | |
| | 3 | | |
| | | | 2 |

**Below a verbal task, a word puzzle scheme.**

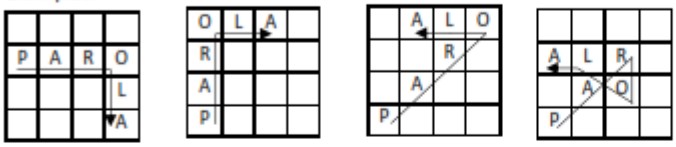
**You have 5 minute to find as many words as you can.**

Please state your age _____

How frequently do you play word puzzles, if ever? 1 ☐ Never 2 ☐ 3 ☐ 4 ☐ 5 ☐ Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:




| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Appendix 8:** Technical students, non-incentivized, version 2: Word puzzle first.



**Below a verbal task, a word puzzle scheme.**

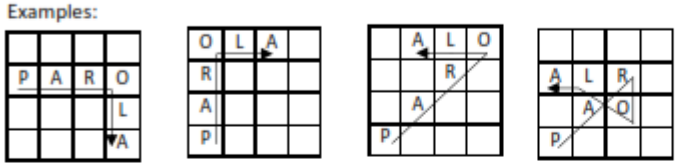
**You have 5 minute to find as many words as you can.**

Please state your age _____

How frequently do you play word puzzles, if ever? 1 ☐ Never 2 ☐ 3 ☐ 4 ☐ 5 ☐ Daily

Should this be the first time you solve a word puzzle-like task, the goal is to find as many words as you can in 5 minutes. You can move up, down, right, left and diagonal. You can use all directions in order to form words by putting together letters next to each other's, as in the examples below. Three letters is the minimum length.

Examples:




| | | | |
|---|---|---|---|
| T | A | L | R |
| F | E | E | D |
| E | F | N | L |
| S | N | O | F |

**Below a numerical task, ten short versions of the Sudoku scheme. Being a student with a strong numerical background, it is likely you developed a high degree of expertise in numerical tasks like this.**

**You have 5 minute to solve as many as you can.**

Please state your age _____

How frequently do you play Sudoku, if ever? 1 ☐ Never 2 ☐ 3 ☐ 4 ☐ 5 ☐ Daily

Should this be the first time you solve a Sudoku-like task, the rule is that all the figures from 1 to 4 should appear in each row, column and quarter of the griddle, one time and one time only, in any order, as in the example below.

Example:

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 4 | 3 | | |
| 2 | | | |
| 3 | | | |

| | | | |
|---|---|---|---|
| | | 4 | |
| 1 | | | |
| | | | 3 |
| | 1 | | |

| | | | |
|---|---|---|---|
| | | | 3 |
| | | | 1 |
| 1 | | | |
| 2 | | | |

| | | | |
|---|---|---|---|
| | | 1 | |
| 4 | | | |
| | | | 2 |
| | 3 | | |

| | | | |
|---|---|---|---|
| 2 | | | |
| | | 1 | |
| | 2 | | |
| | | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | 1 |
| 3 | | | |
| | | | 4 |
| | | | |

| | | | |
|---|---|---|---|
| 1 | | | 3 |
| | | | 1 |
| | 2 | | |
| | | | |

| | | | |
|---|---|---|---|
| | | | |
| | | 4 | |
| | | 2 | |
| 1 | | 3 | |

| | | | |
|---|---|---|---|
| | 1 | 4 | |
| | 3 | | |
| | | 2 | |
| | | | |

| | | | |
|---|---|---|---|
| 3 | | 2 | |
| | | | |
| | | | |
| | 1 | | 4 |

| | | | |
|---|---|---|---|
| | 4 | | |
| | | | |
| | 3 | | |
| | | | 2 |

**Appendix 9:**

Table 3: Test of Normality (Shapiro-Wilk) for Sudoku Scores and Word Puzzle Scores.

| | W | P |
|---|---|---|
| Sudoku Scores | 0.955 | 0.007 |
| Word Puzzle Scores | 0.972 | 0.073 |

Note: Significant results as per the Sudoku Scores suggest deviation from normality.

# 8. Figures

**Figure 5:** Correlation coefficient matrix

| | Technical Task | Technical Student | Rewarded | Experience | Age |
|---|---|---|---|---|---|
| Technical Task | 1 | | | | |
| Technical Student | 0 | 1 | | | |
| Rewarded | 0 | 0 | 1 | | |
| Experience | 0.042 | 0.160 | -0.176 | 1 | |
| Age | 0 | 0.646 | 0.208 | 0.058 | 1 |

**Figure 6:** Plot of standardized residuals over standardized predicted values.