CBS
COPENHAGEN
BUSINESS SCHOOL
HANDELSHØJSKOLEN

THE EFFECT OF REWARDS ON PERFORMANCE

- IN A STANDARDISED EXPERIMENT.

A master thesis by:

Jens Lev Nytoft Berlin

Cand.Merc. Applied Economics and Finance.

Supervisor: Aleksandra Gregoric

Number of pages: 75

Number of characters: 171,531

Date of hand-in: December 1st 2016

Abstract

There has been a significant increase in the use of monetary incentives for promoting performance in recent years in spite of a growing body of literature within psychology that find the well-functioning of incentives to be more limited than what is prescribed from standard economic theories. The nature of the task in terms of performance type (quality vs. quantity, or complex vs. simple) has been proposed as a significant moderator for the effect of incentives on performance. Simple quantity-type tasks have provided the most consistent positive effect from performance-contingent incentives. To test if the nature of the task is a moderator for the effect of providing incentives, an experiment was performed with 40 students who were randomly assigned a bonus or no bonus treatment. The students solved two complex tasks and one simple task. The hypotheses were that providing incentives would have a positive effect for the simple tasks and a negative effect for the complex tasks. These hypotheses were tested in independent-two-samples t-tests for task scores dependent on incentive treatment. No significant effect of the incentive was found on the performance for the complex tasks. The simple task provided a significantly negative effect of the incentive. This was surprising as it was the task that, from a theoretical perspective, was the least likely to produce a negative effect of the incentive. The experiment tested three specific tasks and the results are specific to these tasks, without being able to determine if the findings generalise to other tasks or task performance in work settings in general.

Key Words: Motivation, Incentives, Pay for performance, Standardised Experiment, Herzberg, The Porter-Lawler model, Cognitive Evaluation Theory, Self- Determination Theory, Principal Agent Theory, Measuring performance, Extrinsic and intrinsic motivation, Crowding Theory, Yerkes-Dodson law.

Table of Content

1. Introduction

Despite falling profits Wall Street banks handed out $28.5 billion in bonuses in 2014 which translates into an average bonus of $172,860 per person, that is a 2% increase from the previous year (McSpadden, 2015)[1]. The extent and especially the size of bonuses that has been given to CEOs and security industry jobs have received a lot of attention since the financial crisis. In 1993 the Clinton administration made CEO compensation above $1million non-tax deductible in Section 162(m) of the Internal Revenue Code in an effort to put a cap on skyrocketing compensations (Holmberg & Schmitt, 2014). However, it had several loopholes but the main one was the exception for pay that was linked to performance (Holmberg & Schmitt, 2014)[2]. Nobel laureate Joseph Stieglitz, who was working at the Clinton administration at the time, noted that it in extreme cases it leads to distorted accounting (University of California Television, 2015). Furthermore, executives would be rewarded if the Federal Reserve would lower interest rates that lead to higher stock prices. Or if you are the executive of an airline company falling oil prices leads to higher stock prices and other phenomena that has little to do with the performance of the executives in various firms as Stieglitz puts it (University of California Television, 2015). This has been supported by Bertrand and Mullainathan (2001):"CEO pay in fact responds as much to a lucky dollar as to a general dollar" (901).

What these figures say about the actual performance and whether CEO pay is actually inefficient is very hard to conclude and many studies have tried to evaluate it empirically with varying conclusions (for more see Edmans & Gabaix, 2009; Camerer & Malmendier, 2007). It does invariably show a very strong belief in providing bonuses as a management tool for mitigating agency problems. The use of incentives is not limited to top executives but can be found at all levels of all sectors. Ninety percent of companies now require employees to participate in variable pay plans, up from about 50 percent two decades ago, according to a survey of 1,100 U.S. companies by human resources consulting firm Aon Hewitt (Kelleher, 2013). When asked, in a survey of 205 executives from 30 countries, the executives believe that incentives have only a slight ability to motivate executive performance. When the same executives are asked why they apply it for their employees the number one reason was to improve motivation. It is quite paradoxical that the interviewed executives are ambiguous with regards to the motivational effect of bonuses depending

[1] For more on The New York City State Comptroller's estimate see (DiNapoli, 2015).

[2] Based on a study of the top fifty companies in the Standard & Poor's 500-stock index with the highest market values (Sloan, 2016) found that: "In 1992, only 35 percent of the people in our study … had more than $1 million of income in the categories subject to deductibility limits… But in 2014, the last year for which corporate salary income is available, the number had risen to 95 percent" (Sloan, 2016).

on who it is intended for (Beer & Katz, 2003). The executives do not perceive the incentives to a have an influence on their performance but believe that it does motivate their employees. There is a strong belief in the price-effect in economic theory "if you reward, you get more" and "more is good" which adds up to the *logic conclusion* to reward desired behavior. However, is the widespread use of pay for performance empirically backed? If people get more motivated when given cash bonuses does it ultimately predict better performance?

2. Research question

What will be the effect of performance-contingent incentives on performance in an experiment with students?

A cornerstone of the theory in personnel economics is that workers respond to incentives. Paying on the basis of output will induce workers to supply more output (Lazear, 2000). Rebitzer and Taylor (2010) contrarily state: "Indeed, much of the literature showing that "incentives works" does so by exploiting the measured consequences of poorly designed incentives. That is, they demonstrate that organizations – at least in some cases – do *not* choose incentives optimally" (17). Based on the example provided by Lazear (2000), Rebitzer and Taylor (2010) further state, that the introduction of a piece rate system had a selection effect. Workers had to choose between a lower compensation or a higher work rate, and the workers who disliked this proposition left, while the firm was able to attract workers drawn to this income-effort tradeoff (Rebitzer & Taylor, 2010, 24). A positive effect from incentivising is consistent with the relative price effect and have been replicated numerous times: "Important work by Theodore Groves and John McMillan and their co-authors shows that strengthened incentives led to substantial productivity increases in Chinese industry and agriculture" (Rebitzer & Taylor, 2010, 17). According to Lazear (2000) it is very important to understand that the findings of a positive result of performance-contingent pay on performance is obtained within manufacturing. The characteristics of jobs within manufacturing are changing. The more classical assembly line jobs within production of different goods perhaps most well known from the production of Ford T and Charlie Chaplin's movies are disappearing. They are not just disappearing from the United States and Western Europe to China which has been a popular belief but they are disappearing all together: "In 1970, more than a quarter of U.S. employees worked in manufacturing. By 2010, only one in 10 did" (Kenny, 2014). "Pretty much every economy around

the world has a low or declining share of manufacturing jobs" (Kenny, 2014). Even China has been facing declining employment rates within manufacturing (Kenny, 2014) mainly attributable to technological advances and "Globalization" in the sense that everything is coming from everywhere as Economist Robert Reich puts it, leaving less and less influence of proximity of manufacturing to the assembly line (The Georgia Straight, 2013). The typical job description of a white-collar worker has changed, so has the nature of tasks from more simple tasks with simple rules to more "right-brain" creative and conceptually demanding tasks (Pink, 2010). The typical job characteristics are changing but management tools which can be considered as a "sticks and carrots" approach has still prevailed in the fight for a motivated workforce (Pink, 2010). Does pay for performance also predict a positive effect on performance when the tasks cannot be characterised as simple with effort being the main predictor?

Research by psychologists has sought to identify the range of situations and psychological mechanisms that can produce a positive or perverse relationship between motivation and performance. For example, one mechanism through which increased motivation can backfire is when it leads to greater self-consciousness. When performance on a task relies on highly practiced, automatic skills (Baumeister,1984; Langer and Imber,1979) increasing awareness, competition, introducing a cash incentive or audience or ego-relevant threats, can cause people, involuntarily, to consciously think about the task , shifting control from "automatic" to "controlled" processes that are less effective (Ariely, Gneezy, Loewenstein & Mazar, 2009).

A second mechanism by which increased motivation is likely to have a negative effect on performance relates to a general focus of attention. Attentional focus can be detrimental for tasks that involve insight or creativity, since increased motivation tends to narrow individuals' focus of attention (Easterbrook, 1959), and creativity and insight require drawing unusual connections between elements. McGraw and McCullers (1979) provided support for this mechanism by showing that the introduction of monetary rewards for tasks that involved problem-solving had detrimental effects on performance.

The strengthening of performance by rewards can cause an unpleasant experience of being controlled by others and reduce task interest and creativity. Several studies testing the additivity of intrinsic and extrinsic motivation found tangible extrinsic rewards to be undermining intrinsic motivation whereas verbal rewards enhanced intrinsic motivation (Deci, 1971). Suggesting, that intrinsic and extrinsic motivation has altering interactive effects instead of the simple additive relation previously suggested.

The expectation that people will improve their performance when given performance-contingent incentives rests on two subsidiary assumptions: 1) that providing performance-contingent incentives will increase overall motivation and effort, and 2) that this increase in motivation and effort will result in improved performance.
The scope of this thesis will be on investigating whether and under which circumstances the last assumption holds.

The thesis investigates the effects of performance-contingent incentives on performance by conducting experiments with economics students under two different conditions: one with performance-contingent bonuses and one without. The experiment was conducted in Denmark with students attending Copenhagen Business School. The results obtained from the experiment provide an indication of whether there is a significant difference between the performance under the two different treatments bonus or no bonus. The insight obtained through the literary review and the empirical research, provides the basis for how the experiment is constructed and which indications it is expected to provide. The research question is clarified through the formulation of two following hypotheses in section 8, which will be tested in the experiment. The data from the experiment is then subjected to an analysis that provides an answer to the hypotheses and the research question.

2.1 Delimitations

In order to narrow the focus and create a coherent structure for the current thesis important choices have been made regarding the focus of investigation, what theories are presented and which have been left out, how the theories interact, the choice of research method, the experimental design, the social setting and how to analyse the data.
Studies have shown that tangible rewards tend to undermine intrinsic motivation, but they have also shown that these rewards have negative effects on other important variables, including learning, well-being and prosocial or generative behavior (Sansone & Harackiewitcz, 2000). In spite of these potential effects the focus of this thesis is on the effect of providing the incentive on the performance for three specific tasks in an experiment.
Many other theories could have been presented besides the ones in question. The narrow focus of this thesis has caused the exclusion of many contributions that behavioral economics has made. There will not be a presentation of behavioral economic topics as prospect theory, status-quo biases, bounded rationality and several other topics.

The presented theories and research in this thesis have been chosen based on their ability to provide valuable insight when used in relation to each other. There are two main different theoretical approaches to what drives desired behavior and performance with the proponents of intrinsic motivation on one side and the proponents of extrinsic motivation on the other side and both set forward substantial critique of the other.

When deciding on the best research method for investigating the effect of performance-contingent bonuses on performance several issues have been taken into consideration. An alternative method of inquiry could have been based on interviews with key personnel in charge of incentive systems in companies that apply different incentive systems with a top down approach. Asking them a series of questions that would shed light on their reasoning for implementing it, if it works as intended, if there has been some undesired or desired consequences that were not anticipated and most importantly how it effects the performance of the subjects or groups of subjects who partake in these pay-for-performance plans. This method of inquiry is expected to come with one general undesired feature independent from the kind of salary structure, how it is administrated etc. namely that it would be based on rationalisations of why the system in place is favourable. According to (Beer & Katz, 2003):"It seems reasonable, as mentioned above, that HR directors may provide an overly-positive assessment of a company's compensation plan, inasmuch they have a direct investment in the perceived success of that plan [Brenner]" (11).

This implies that some sort of justification of previous decisions would prevail in this qualitative approach. Another approach could be to perform interviews with personnel who are currently receiving pay-for-performance and try to make them assess the effects of these plans however, with this approach it is expected that employees would focus on inadequacies in the incentive design mainly providing suggestions to possible incremental upgrades to their current system. Further it can be noted that a potential selection effect where personnel who have a personal inclination towards pay-for-performance are attracted by the incentive structure in place, and thereby amplify this justification process. One can possibly overcome such rationalisations or post-hoc justifications during the course of such interviews if structured and developed in an optimal way, however it makes it a difficult research method for inquiring into the actual effect of performance-contingent bonuses on performance. A research method based on questionnaires or e-surveys that represent a quantitative inquiry method based on proportions would be facing the very same difficulties. It would be a stacking of these rationalisations providing a picture of the proportion of the practitioners who prefer one incentive strategy to another instead of providing an indication of the

actual relation between bonuses and performance. It is how the subjects perform when given the prospect of a bonus compared to a control group in a standardised experiment, which is the central method of exploration during the course of this thesis.

The experiment was conducted at Copenhagen Business School (CBS) Dalgas Have 15. The requirement for participating in the experiment was that subjects were enrolled at the CBS and the recruiting process and actual tests took place at the location. All participants were in the range 20-29 years of age and there was close to an even ratio of men and women.

The reasoning why students and not an actual workplace setting has been selected is due to a lack of sufficient funds in order to make the potential rewards significant to the participants. The experiment represents a focused sample as all the individuals who participated have relatively comparable demographics. More importantly the students represent a group of people who will likely face the prospects of bonuses in the professional careers that lay ahead of them. This line of reasoning leads to the assertion, that it is not unlikely that it is the second best social setting, after an actual work setting, for such an experiment. It can be noted that due to the scarcity of funds the total number of participants was limited to 40 students and it would be desirable to perform the test on a larger sample in order to decrease the error measures and improve the empirical data or perform a similar or a closely related experiment in an actual work setting.

2.2 Inquiry paradigm

This thesis professes itself to a "position" close to the postpositivistic paradigm. The position is the lens of the inquirer when dealing with aspects of cognition. The method of inquiry as it reflects basic belief systems or paradigms in how it guides the disciplined inquiry and the responses to ontological, epistemological and methodical questions (Guba, 1990, 18).

According to postpositivists there are only two main sources of cognition, namely what we can observe using our senses and what we can determine using our logic (Thurén, 2008). The postpositivistic paradigm is an adjusted version of positivism or a response to the critique raised against it. The postpositivistic paradigm challenges the traditional notion of the absolute truth of knowledge and that one cannot be positive when studying the actions of humans (Creswell, 2014). The ontology is realist at its core however it is a critical realism. It is critical realist, as reality exists and it is driven by real causes, however one can never be sure that this reality has been fully uncovered due to imperfect sensory and intellectual deficiencies (Guba, 1990, 20). It is for this reason that researcher's state that they do not prove a hypothesis, instead they indicate a failure to

reject the hypothesis (Creswell, 2014). The postpositivists find a non-interactive position to be merely a hypothetical position as it is impossible for the inquirer to perform any inquiry without interacting with what is inquired (Guba, 1990, 20). The epistemology of postpositivism is modified objective, as objectivity is the ultimate ideal. However, it cannot be reached in any absolute sense, but should be approximated by taking as neutral a position as feasible, and disclosing the predispositions of the inquirer so that readers can make adjustments accordingly (Guba, 1990, 21). The research process in a postpositivistic tradition is described by Creswell (2014) as: "Thus, in the scientific method- the accepted approach to research by postpositivists - a researcher begins with a theory, collects data that either supports or refutes the theory, and then makes necessary revisions and conducts additional tests" (7). The postpositivists follow a modified experimental or manipulated methodology where the emphasis is on critical multiplism, to decrease the risk of distorted interpretations: "If human sensory and intellective mechanisms cannot be relied upon, it is important that the "findings" of an inquiry be based on as many sources- of data, investigators, theories, and methods- as possible" (Guba, 1990, 21).

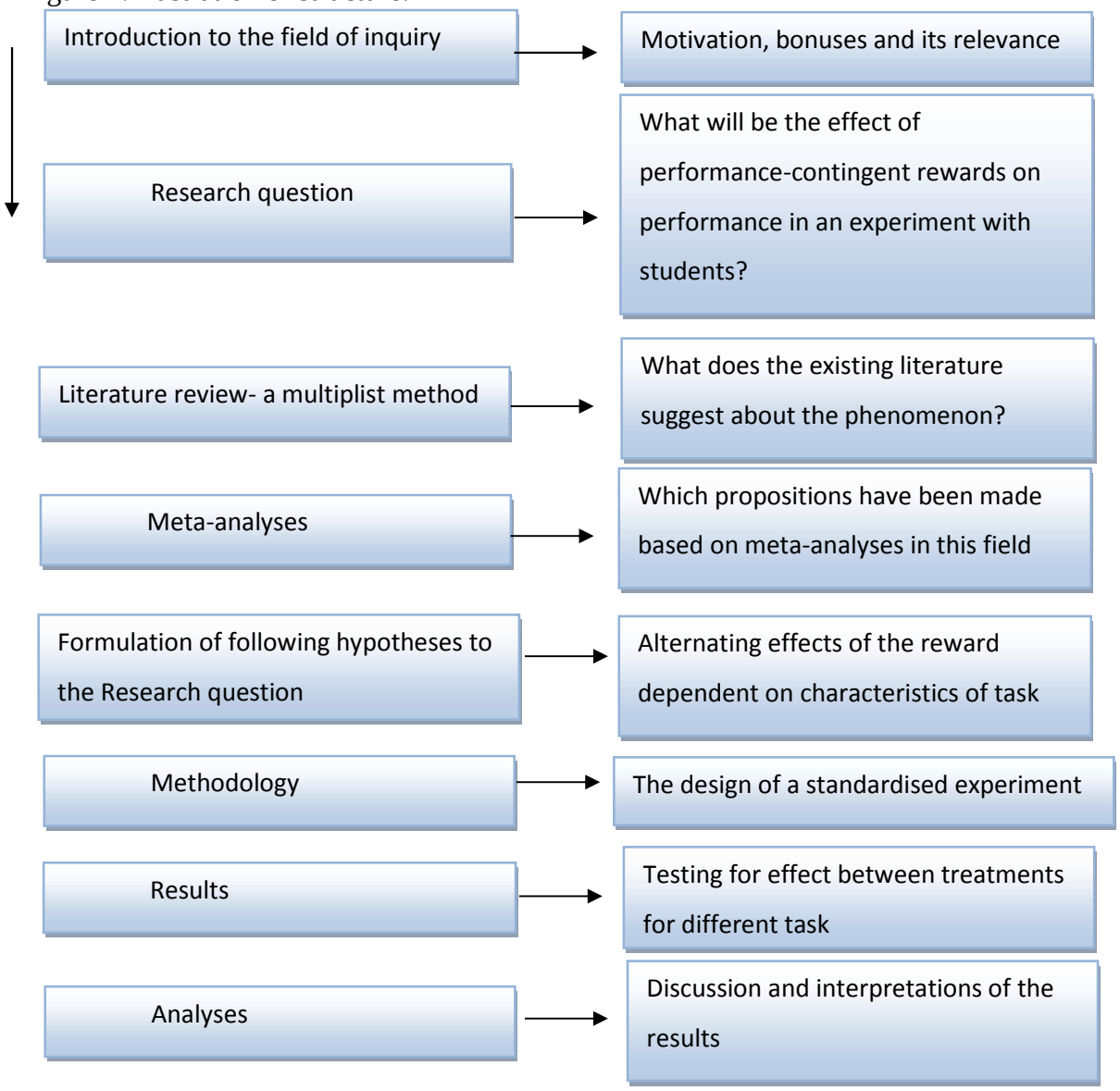
3. Structure

This section provides an insight to the driving force of the research and how this inquiry has manifested itself in the structure. The following line of questions represents the line of inquiry into the topic or the driving force of the research.

What is motivation? What is pay for performance? Why has it become this influential as a corporate governance tool? Firstly, what does the economic literature say about the strengths and weaknesses of performance-contingent incentives? Secondly, what does the psychological literature say about paying people for something they are intrinsically interested in? If the task is boring or simple you might get better results from incentivising, but what if the task is complex or serves intrinsic human aspects as competence affirmation or self realisation potential? What does the empirical evidence suggest? Does incentivising and adding extrinsic motivation improve performance compared to the intrinsically motivated subjects when performing the same task when tested in an experiment with students? How do you explain the outcome of these tests?

The progression of inquiry is structured in different sections and can be seen as the structural body of the thesis. The following figure is a representation of the chosen structure and how the elements provides insight to the inquiry. The structure of the thesis is depicted in Figure 1:

Figure 1.Illustration of structure.



The Introduction presents the ideas that sparked the curiosity of the researcher and framed the relevance of further investigation into this topic. The development of the research question is a natural prolonging of the inquiry into the topic followed by the choices made to narrow the playing field of interest and choices made as a result of the selected research method and its actual design. This is followed by a presentation of the selected theory. It consists of several different theories and theoretical approaches but serves the purpose of providing insights into the concepts of extrinsic and intrinsic motivation and their interaction with performance. The theoretical framework is extended with a presentation of the selected meta-analysis on the subject that provides important insights into the results, that has been obtained in the extensive research in the field of intrinsic motivation, extrinsic motivation and performance and the hypotheses that have been formulated

based on these meta-analyses. The meta-analyses further provides some very useful guidelines and considerations regarding the experimental design and the strengths and weaknesses that may arise from certain decisions, as one of the main aspects of meta-analyses is selecting and omitting results within a clearly defined area of research, based on methodological inadequacies. The insight obtained at this point leads to the formulation of two following hypotheses that focus the research question and serves as predictions for the test. This section leads to a presentation of the methodology and considerations that result from choosing a standardised experiment as exploration method with important aspects of validity and reliability. This is followed by a presentation of the particular choices made regarding the experimental design and the reasoning for these particular choices. Next, is a presentation of the results found in the analyses of the data from the experiment. These results are then discussed given the knowledge obtained in previous research and in the theoretical framework. Finally I will provide a conclusion on this investigation and answer the research question and the two following hypotheses.

4. Literature review

4.1 Relevant theory

The theory chosen for this thesis all serves the purpose of illuminating the curious case of what extrinsic and intrinsic motivation is and in turn, how it affects performance.

4.2 Definition of motivation, performance and incentives

Motivation, performance and incentives are crucial concepts at the centre of this thesis. Everyone has a concept of what the term motivation embodies on a general level without having one clear meaning. For this reason, it is very important first to conceptually clarify what is meant by the term motivation before one can proceed to an understanding of derived terms as extrinsic or intrinsic motivation. The word motivation is etymologically derived from the Latin words movere (move, stir, agitate), motivus (stirred, moved), and motare (set in motion, shake, stir). These words seem to share a common focus on movement. The everyday usage of the term motivation describes why a person does something however this is very vague. Within the academic field several authors have defined motivation in different ways for instance as: "a need or desire that serves to energize behavior and to direct it towards a goal" (David Myers, 1996, 297).
Hawkins (1993) defines motivation as:

> what drives or induces a person to behave in a particular fashion [...] the internal force which initiates, directs, sustains and terminates all important activities. It influences the level of performance, the efficiency achieved and the time spent on an activity. (132)

These related definitions of motivation have an emphasis on the different factors that drive and influence, e.g. directing, initiating or sustaining behavior and the intensity and quality of this behavior.

These different factors that drive behavior have widely been recognised and categorised as internal and external factors. Porter and Lawler (1968), here cited from Gagne and Deci (2005), defined intrinsic and extrinsic work motivation as:

> Intrinsic motivation involves people doing an activity because they find it interesting and derive spontaneous satisfaction from the activity itself. Extrinsic motivation, in contrast, requires an instrumentality between the activity and some separable consequences such as tangible or verbal rewards, so satisfaction comes not from the activity itself but rather from the extrinsic consequences to which the activity leads. (331)

Performance is a term that refers to behavior that can be observed and assessed. The term performance as Cerasoli, Ford and Nicklin (2014) present it:

> The term performance is “synonymous with behavior . . . it is something that people actually do and can be observed” (J. P. Campbell, McCloy, Oppler, & Sager, 1993, p. 40). Performance is achievement-related behavior, with some evaluative component (Motowildo, Borman, & Schmit, 1997). For example, performance in academic settings may be operationalized as presentation quality, while in a sports setting it may be the number of goals scored. (981)

Incentives in economic theories are mainly concerned with the possible sources that motivate agents to choose between different actions. According to Cerasoli et al. (2014):

> Formally defined, incentives “are plans that have predetermined criteria and standards, as well as understood policies for determining and allocating rewards” (Greene, 2011, p. 219). Although “money is probably the most widely used incentive” (Pinder, 2011,

p. 396), incentives include anything provided by an external agent contingent on performance of particular standards of behavior(s). Thus, promotions, grades, awards, health benefits, praise, and recognition are all incentives (981).

4.3 A historical account of the theoretic field of motivation

There are a two very central underlying assumptions that vary across the different theories of motivation namely whether behavior is based on a *natural force* e.g. needs or desires, or some kind of *rationality* e.g. instrumentality or meaningfulness and whether the focus is on *process*, how motivation takes place, or, the *content*, what motivates.

In a historical perspective hedonism has been a dominant influence on the psychological approach to motivation. Hedonism refers to several related theories about what is good for us, how we should behave, and what motivates us to behave in the way that we do. All hedonistic theories identify pleasure and pain as the only important elements of whatever phenomena they are designed to describe. Within the philosophical arena *value hedonism* holds that pleasure is intrinsically valuable and pain is disvaluable. Something is "intrinsically valuable" if it is valuable for its own sake, whereas instrumentally valuable goods as money only holds value because of what we can do with it, which in turn gives us pleasure. Much non-philosophical literature and everyday talk of hedonism, refers to a person who seeks out personal pleasure without any regard for the well-being of others and in some cases not even their own future well-being. This version is often termed *folk hedonism* and is very closely related to *hedonistic egoism* which states that we ought, morally speaking, to do whatever provides us with the most net pleasure after pain is subtracted at any given time. *Motivational Hedonism* or *Psychological Hedonism* is a theory that describes the desire for pleasure and avoidance of pain as the guidance of all our behavior. Prominent theorists as Sygmund Freud, Jeremy Bentham and John Stuart Mill amongst others have argued for versions of motivational hedonism. Bentham and Mill used the idea to support their theory of *Utilitarianism*, which states that the right course of action is the one that is most likely to produce the greatest net happiness of all the concerned people (Overskeid, 2002, 77). These different accounts of hedonism are at the centre of the development of a multitude of theories that relates to human needs. The psychoanalytic Freud suggested only two basic needs, the life and death instinct, which gave inspiration to Drive reduction theory formulated by Clark Hull (1943). It stated that humans have four basic psychological needs; hunger, thirst, sex and avoidance of pain, whereas (Maslow, 1943) suggest a hierarchy from lower deficiency needs as hunger and safety to the higher growth needs as

the need to achieve and to realize one's potential. The early "prototypical" theories, stimulus-response theories i.e. psychoanalytic of Freud, Hull's Learning Theory, Classical Conditioning known for the Pavlovian dogs, Operant Conditioning by B.F Skinner (1984) or Lewin's field theoretical approach, all portrayed humans as machine like reactive organisms obliged to act by internal or external forces beyond our control e.g. instincts, needs, drivers or reinforcers etc. These theories all imply that motivation will result if the right button is pushed (Oettingen & Gollwitzer, 2001, 329). More modern theories as the Social Learning theory, Albert Bandura (1977) advocated that behavior is chosen based on what is expected to lead to the most personally rewarding goals. It expands on traditional pure stimulus-response behavioral theories by emphasizing various internal processes as the cognitive process and that motivation can arise from thoughts, beliefs, morals, feedback, and not only rewards.

4.4 Content theory

Content theories provide for a static "snapshot" of how people are motivated to perform various activities by describing what motivates people to work. Content theories hold that certain internal factors direct, initiate and sustain behavior. These internal factors are described as needs and the theories try to identify these needs and their fulfilment.

4.4.1 Herzberg – two factor theory or motivation-hygiene theory

Frederick Herzberg[3] was heavily influenced by Maslow's hierarchy of needs but relates more specifically to the motivation of individuals in a work setting. Herzberg developed the motivation-hygiene theory based on an analysis of more than 200 interviews with engineers and accountants in the late 1950'ies. Herzberg applied a critical incident technique where the respondents were asked to describe anytime when they felt either exceptionally good or bad about their job (Herzberg, Mausner & Snyderman, 1959). Herzberg theory states that job satisfaction and dissatisfaction are affected by two different sets of factors and hence cannot be measured on the same continuum, which was the popular view at the time (Behling, Labovitz, & Kosmo, 1968). Herzberg argued that meeting *hygiene factors* of individuals, or the lower-level needs as Maslow coined them, would not motivate them to exert effort, but would only prevent them from being dissatisfied. Only if the

[3] For the sake of simplicity this line of work is referred to as Herzberg's work as he continuously developed it even though there were several others co-authoring the initial publication.

motivators, or higher-level needs, were met would individuals be motivated. The hygiene factors such as company policy, administrative practices, supervision, interpersonal relationships, working conditions, and salary contribute very little to job satisfaction. The motivators that lead to satisfaction are achievement, intrinsic interest in the work, responsibility, and advancement (Herzberg et al. 1959). This implies that managers must focus on changing the intrinsic nature and content of jobs themselves by "enriching" them to increase employees' autonomy and their opportunities to take on additional responsibility, gain recognition, and develop their skills and careers. The theory distinguishes between the nature of the work itself and the job environment or inner and outer factors. This is very important for the further inquiry as it is highly related to the terminology of intrinsic and extrinsic motivation in *Self-Determination Theory* by Deci and Ryan which will be presented in more detail later. Further Herzberg did not find salaries to be a motivating factor based on his study.

Critics have cited that the theory is a need based theory which leaves very little or no influence to the decisions of individuals. Content theories suggest that the drive or desire to fulfil needs stems from need deficiencies. The theory does not take personal differences in motivational factors into account. However, the theories points to motivational factors that relates to other factors than salaries and bonuses. The theories puts an emphasis on traits as autonomy, self-realisation and mastery in the job context which is of importance to a growing group of researchers on motivation and especially as unskilled manual labour jobs with lower complexity is in a decline.

4.5 Process theory

Process theories seek to explain the dynamic process of how people make choices under certain types of processes or situations in an effort to achieve desired rewards.
Process or cognitive theories describes how individual behavior is energized, directed and maintained as a result of a conscious decision-making process based on beliefs and expectations of the future. The theories rest on an assumption that individual behavior is based on conscious and rational choices. The focus on expectations is central to *Expectancy Theory* formulated by Victor Vroom and later extended by Porter and Lawler, which will be presented next.

4.5.1 Expectancy theory by Victor Vroom

The expectancy theory by Victor Vroom was formulated in 1964 based on the concepts developed by psychologist Kurt Lewin. Lewin's field theory stated that behavior is a function of both the person and the environment and the individual plans behavior according to how valuable the desired outcomes are and how able the person is to achieve these goals. Vroom (1964) applied these concepts directly to a work setting and sets forward this proposition to explain motivation:

> The force on a person to exert a given amount of effort in performance of his job is a monotonically increasing function of the algebraic sum of the products of the valences of different levels of performance and his expectancies that this amount of effort will be followed by their attainment. (284)

The proposition states that when individuals decide how much effort to put into work behavior it depends on three different sets of factors (Vroom, 1964):

- **Expectancy**, that putting forth effort will lead to a given level of performance, a sort of action-outcome link.
- **Instrumentality**, the degree to which the individual believes that a given level of performance will lead to the desired reward, a sort of performance-reward link.
- **Valence**, which is the extent to which the expected outcomes are attractive or unattractive.

These relations were also stated in a formula:

Motivation = *Expectancy* * *Instrumentality* * *Valence*

Managers should attempt to ensure that their employees believe that increased effort will improve performance and that performance will lead to valued rewards. The theory has an underlying assumption that increased effort leads to increased performance, which is the central theme of investigation for this entire thesis. Furthermore, motivation for action is treated as a unitary concept that varies in amount rather than kind.

4.5.2 The Porter-Lawler Model

Porter and Lawler published an extension of the Vroom expectancy model in 1968. The basic premise of the Porter-Lawler model is the same as for Vroom's model, since an individual's motivation to complete a task is affected by the reward they expect to receive for completing the

task. However, the Porter-Lawler model is more complex in a number of ways. It suggests that increased effort does not automatically lead to improved performance because individuals may not possess the necessary abilities needed to achieve high levels of performance, or because they may have an inadequate or vague perception of how to perform necessary tasks. Without an understanding of how to direct effort effectively, individuals may exert considerable effort without a corresponding increase in performance. Porter and Lawler thereby stress that an individual's *ability* to perform a task and their *perception* of the task influence motivation and ultimately the performance.

Successful performance leads to rewards. Porter and Lawler distinguish the rewards as either *intrinsic* or *extrinsic*. Intrinsic rewards are the positive feelings that the individual experiences from completing the task e.g. satisfaction and feeling of accomplishment. Extrinsic rewards are rewards emanating from outside the individual such as raises, bonuses or recognition. In accordance with this view the intrinsic rewards are more closely related to successful performance than the extrinsic since the intrinsic rewards are is directly linked to performing a task whereas the extrinsic rewards are a result of a company's management and administration. Porter and Lawler's model suggested that an individual's view regarding the attractiveness and fairness of the extrinsic rewards will affect motivation (Porter & Lawler, 1968). Porter and Lawler (1968) advocated structuring the work environment so that effective performance would lead to both intrinsic and extrinsic rewards, which would lead to total job satisfaction. This was to be accomplished by enlarging jobs to make them more interesting, and thus more intrinsically rewarding, and by making extrinsic rewards such as higher pay and promotions clearly contingent upon effective performance.

The theory is a model of behavioral choice. It provides an explanation of why individuals choose one behavioral option over others, thereby focusing on the behavioral direction process. It barely attempts to explain what motivates individuals, instead it focuses on how individuals make decisions to achieve their end value. Furthermore, the theory assume that these decisions are based on conscious and rational choices, which is an assumption that has been largely contested from many sides during the course of time, without going into details on the dispute on rational choice theory.

The process theories are criticised for treating motivation for action as a unitary concept that varies in amount rather than kind. It is simplistic, because the total motivation a person has is determined by various factors, but is represented in a single variable, that provides the basis for making

predictions. Even theories that distinguish intrinsic motivation and extrinsic motivation consider them to be additive, with total motivation being the critical motivational predictor or as Gagné and Deci (2005) comment on the Porter-Lawler model: “Implicit in this model is the assumption that intrinsic and extrinsic rewards are additive, yielding total job satisfaction” (331).

4.6 Cognitive Evaluation Theory

Building on Vroom’s (1964) expectancy-valence theory of motivation, Porter and Lawler (1968) proposed a model of intrinsic and extrinsic work motivation. Implicit in the Porter and Lawler model is the assumption that intrinsic and extrinsic rewards are additive, yielding total job satisfaction. Early studies testing the additivity hypothesis found that tangible extrinsic rewards undermined intrinsic motivation whereas verbal rewards enhanced it (Deci, 1971). Implying that intrinsic and extrinsic motivation can be both positively and negatively interactive rather than additive. The initial finding of decreased intrinsic motivation for an interesting activity following the experience of being rewarded for doing it has been referred to as the *undermining effect* (Deci & Ryan, 1980; 1985b). Numerous experiments replicated and extended the finding and highlighted its limiting conditions and the results gave birth to the Cognitive Evaluation Theory (CET). CET specifically focuses on the conditions that diminish or enhance intrinsic motivation. CET presumes that people have an inherent need for self-determination and competence. CET further asserts that all external events have both a *controlling* aspect and an *informational* aspect (Deci & Ryan, 1985a). The theory argues that events that negatively affect a person's experience of autonomy or competence diminish, whereas events that support perceived autonomy or competence enhance intrinsic motivation. The effect of rewards on intrinsic motivation depends on whether the informational or controlling aspect is more salient. The informational aspect relates to competence affirmation whereas the controlling aspect undermines the sense of self-determination.

CET introduces the term *perceived locus of causality* (PLOC) which refers to a person’s perception of whether the reasoning leading to a behavior originates from within the person or externally. CET suggests that external factors that prompt a change in PLOC from internal to external factors are tangible rewards, deadlines, surveillance, and evaluations which tend to diminish the sense of autonomy and undermine intrinsic motivation (Amabile, DeJong & Lepper, 1976; Lepper & Greene, 1975; Smith, 1975; DeCharms, 1968; Heider, 1958). CET suggest that some external factors can have the opposite effect such as providing choice about aspects of task engagement, which tend to enhance feelings of autonomy and therefore can prompt a shift in PLOC from

external to internal and increase intrinsic motivation (Zuckerman, Porac, Lathin, Smith & Deci, 1978).

Critics of CET have stated that the studies testing CET were laboratory experiments instead of organisational. Many activities in work organisations are not intrinsically interesting which makes it somewhat inadequate. Most people who work have to earn money, so using monetary rewards as a central motivational strategy seems practical and appealing. There seems to be an either-or tendency of promoting intrinsic motivation through participation and empowerment while reducing the use of extrinsic factors. Alternatively, a tendency of using rewards and other extrinsic contingencies to maximise extrinsic motivation while ignoring the importance of intrinsic motivation.
Self-determination theory, which is a further development of CET but broader in scope, seeks to explain how extrinsically motivated behavior can become autonomous and introduces individual differences in causality orientations amongst other additions to CET (Deci & Ryan, 1985a, 1985b, 2000).

4.7 Self-Determination Theory

Self-Determination Theory (SDT) is a further development of CET. It is based on an assumption that people have an inherent growth tendency, or put differently an inner motivation to expand and exercise competencies, explore new things and to learn (Ryan & Deci, 2000a, 70). This assumption posits that all people given the right circumstances will strive to give their best and "fulfil" themselves. This idea of an inherent growth tendency has been the driving force behind extensive research into the circumstances that lets inner motivation flourish and which that diminish inner motivation. People have an inclination to integrate external regulations into self-regulations and try hard to integrate into a larger social whole, which is an idea that is shared by numerous theories (Ryan, 1995). However, the proposition that these evolved integrative or actualizing tendencies operate in combination with basic psychological needs for autonomy, competence, and relatedness is specific to SDT (Sansone & Harackiewicz, 2000). Several studies indicate that needs are closely related with intrinsic motivation. The concepts of autonomy and competence corresponds with early need based theories as well as CET. SDT is comprised of six mini-theories however for the purpose of this thesis the theory is treated as a unified theory.
Deci and Ryan (2000) define needs as "…innate psychological nutriments that are essential for ongoing psychological growth, integrity, and well-being" (229).

According to SDT, people have three core needs; *autonomy, competence* and *relatedness*, which leads to persisted motivation if fulfilled (Deci & Ryan, 2000).

- **Autonomy** refers to the motivation of individuals to pursue their personal values and interests; however, this does not mean that people want to act independently from others.
- **Competence** refers to the development of key skills and abilities a sense of mastery and that one's behavior results in the intended outcomes and effects.
- **Relatedness** refers to a sense of belonging in a group or in close relationship with friends and family.

These three needs are essential for peoples optimal functioning and none of them can be set aside without negative consequences for the well-being of a human. Needs are typically defined as individual or individually varying in strength. However, SDT posits that they are universal and the fulfilment is essential for human development. SDT focus on the extent that people can get their needs satisfied in their social context. SDT posits that needs are universal; however people have varying orientations about the initiation and regulation of their behavior. People have different levels of *general causality orientations* and these orientations indexes to which extent people are *autonomy oriented, control oriented or impersonally oriented* (Gagne & Deci, 2005, 339).

- **Autonomy orientation** reflects a general tendency to experience social contexts as autonomy supportive and to be self-determined.
- **Control orientation** reflects a general tendency to experience social contexts as controlling and to be controlled.
- **Impersonal orientation** reflects a general tendency to be amotivated.

These orientations describe individual differences in people's tendencies to orient towards environments and regulate behavior in various ways. The determination of a person's predominant causality orientation can be used to decide which motivational form that will be successful for activating a person to take action.
Autonomy is one of the three human core needs according to Deci and Ryan (2000). As mentioned above it doesn't refer to independency but to the term *perceived locus of causality* (PLOC), that CET introduced, which relates to a person's perception of whether the reasoning leading to behavior originates from within the person or externally. Albeit the large focus on intrinsic

motivation, SDT acknowledges that only a small proportion of behavior is a result of pure self-interest and self-satisfaction. SDT suggests that behaviors can be characterised in terms of the degree they are autonomous versus controlled. Humans act in social contexts where different behaviors are expected with a sense of having to engage in various actions. Therefore it is highly relevant to see which factors that can motivate people to engage in these activities that results from a sort of social contextual compliance. A behavior is considered to be *externally regulated* when people act with the intention of obtaining a desired result or avoiding an undesired one (Gagne & Deci, 2005, 334). They are performing the behavior only when it is instrumental to some separable consequence. When an employee is given a task by the immediate superior the motivation for carrying out the task can vary significantly. The motivation to perform a task can be everything from active personal engagement to unwilling compliance with the rules or a simple lack of motivation. Gagne and Deci (2005) describe the process of taking in external regulations: "Internalisation is defined as people taking in values, attitudes, or regulatory structures, such that the external regulation of a behavior is transformed into an internal regulation and thus no longer requires the presence of an external contingency" (334).

The more internalised the extrinsic motivation the more autonomous the person will be when enacting the behaviors. SDT describes the process of internalization of external regulation by a controlled-to-autonomous continuum unlike other theories that see it as a dichotomy. Internalisation in the SDT framework is an overarching term that refers to three different processes *introjection, identification* and *integration*:

- **Introjection** describes regulation that is internal to the person, but hasn't been accepted as one's own. It describes a type of internal regulation that is still quite controlling because people perform such action to avoid guilt or anxiety or to attain ego-enhancement or pride.
- **Identification** of the values that drives behavior as similar to the subjects own goals, desires and identity leads to a higher perceived freedom. People feel greater freedom and volition because the behavior is more congruent with their personal goals and identities. They perceive the cause of their behavior to reflect an aspect of themselves.
- **Integration** describes how the person perceives a behavior to be an essential part of who they are and engagement is self-determined.

Integrated regulation represents the highest level of internalisation of extrinsic motivation, and it shares some qualities with intrinsic motivation which is another type of autonomous motivation.

Integrated regulation does not become intrinsic motivation but is still considered extrinsic motivation because it is characterized not by the person being *interested* in the activity but by the activity being *instrumentally* important for personal goals. The theory describes these different types of regulation in order to index the extent to which people have integrated them. SDT posits a self-determination continuum between amotivation and intrinsic motivation. Along this descriptive continuum, are the four types of extrinsic motivation, with external being the most controlled, and thus the least self-determined, type of extrinsic motivation, and introjected, identified, and integrated being progressively more self-determined (Gagne & Deci, 2005). The SDT model of internalization does not suggest that people invariably move through these "stages" with respect to particular behaviors however it is possible under the right conditions.

SDT distinguishes between two forms of extrinsic motivation; controlled and autonomous. Controlled motivation results from governance through rules and rewards, whereas autonomous motivation occurs when people feel they act out their own free will. Intrinsic motivation is per definition always autonomous and the type of extrinsic motivation which is autonomous bear some similarities. Nonetheless, there are two factors which are essential for the internalisation of extrinsic motivation but not necessary in order to maintain intrinsic motivation (Gagné & Deci, 2005). Firstly, restrictions and structures are essential for the internalisation of extrinsic motivation but not necessary for maintaining intrinsic motivation. Second, it is essential for the process of internalisation that key figures close to the subject is supportive of the behavior a sense of *relatedness*, while this is of lesser importance for maintaining intrinsic motivation (Gagné & Deci, 2005, 339).

The reasoning behind distinguishing these factors interaction with the two forms of autonomous motivation is that there is a difference in the behavior they predict. SDT posits that an autonomously motivated employee will perform better at complex tasks, whilst controlled motivation yields the same results or slightly better results on simple tasks (Gagné & Deci, 2005, 346). This is a very interesting prediction, which is examined further in the experiment.

Rooted in three psychological core needs SDT predicts certain aspects in the social context that will support intrinsic motivation and promote internalisation of extrinsic regulations. Based on research these three specific factors have been detected (Gagne and Deci, 2005, 338):

1. A meaningful rationale for completing the given task
2. Recognition that the subjects didn't necessarily find the task intrinsically interesting

3. A focus on options rather than control

An employer should strive to make these factors present at the work place if he is interested in achieving a climate that gives the best possible conditions for motivated and well performing employees.

4.8 Partial Conclusion

When people are intrinsically motivated for an activity they will take it on and derive satisfaction from the activity itself. Extrinsic motivation requires an instrumentality between the activity and some separable consequences and therefore satisfaction comes not from the activity itself but from the extrinsic consequences to which the activity leads. Many activities which agents are expected to perform in the workspace results from external regulations i.e. direct-, social- or legal regulations. SDT explains the process of how these regulations can be internalised so that the agent feels a larger extent of self-determination while acting according to internalised regulations. The outcome of this process is highly dependent on whether the person perceives that it satisfies the core human needs of competence, relations and autonomy. Autonomy is highly related to what CET coins *perceived locus of causality* (PLOC) which refers to a person's perception of whether the reasoning leading to behavior originates from within the person or externally.
SDT states that the type of motivation that is most effective is highly dependent on the nature of the task. Autonomously motivated subjects perform better at complex tasks while controlled motivation has a small or no advantage with simpler tasks.
When an employer wants to promote the motivation of his employees SDT emphasises three aspects in the social context that will support intrinsic motivation and promote internalisation of extrinsic regulations; a meaningful rationale for completing the given task, recognition that the subjects did not necessarily find the task intrinsically interesting and a focus on options instead of control.

5. Principal Agent theory

An underlying assumption of the Principal Agent theory (P/A-theory) is that the agent is self-serving and strives to maximize his own utility, hedonistic egoism, whereas the principal wants the

agent to maximize his utility often represented as shareholder value, the profit of the firm etc. The principal is anyone who hires an agent to solve a task. A typical case is a company traded on the stock exchange where the board of directors, the principals, are elected to represent the shareholders interests who hires the Chief Executive Officer (CEO), the agent, to run the company. At the next step the CEO acts as the principal to the managers who are the next in the chain of command and so on for every step of the hierarchy. For the sake of this thesis the employer will be referred to as the principal and the employees as agents. Both parties are assumed to act rational when they strive to maximize their own utility. In this setting the principal will try to maximize the difference between the productivity of the agents and the costs of running the company including wages paid to the agents. The agents will in turn try to maximize the difference between the compensation including wage, career options, competence development etc. and the loss of time, physical and mental exertion. The P/A-theory is by large focused on this conflict of interest between the principal and the agents and the alignment of these interests through compensation schemes that can take a multitude of forms: piece rates, options, discretionary bonuses, promotions and profits sharing and many more (Prendergast, 1999, 7). Performance-contingent incentives are often viewed as one of the main solutions to the problem of risk sharing between the agent and the principal. The literature within this field has an underlying assumption that individuals respond to contracts that reward performance. Assuming a response one would have to assess whether these responses are always in the interest of the firm.

There is substantial empirical literature testing the trade-off between risk and incentives. The premise of this literature is that relating pay to performance increases output, but at the cost of imposing risk on workers, which is reflected in higher wages (Prendergast, 1999, 8).

Another difficulty when selecting incentive structures is that contracts cannot completely specify all relevant aspects for an agent. As a result contracts offering incentives can give rise to dysfunctional behavioral responses, where agents emphasize only those aspects of performance that are rewarded.

5.1 Contracts and Economic incentives

According to P/A-theory the contradicting interests of agents and principals make it necessary for the principal to find a mechanism to regulate the behavior of the agent. The traditional solution is a contract that specifies what is expected in different situations. If a contract could specify all the actions expected from an agent, down to the last detail, there would not be any problems. Even in the simplest of job descriptions “complete” contracts are purely hypothetical due to varying

uncertainties and actions that are hidden from the principal (Besanko, Dranove, Shanley & Schaefer, 2007, 457). Further, it is very costly for the principal to monitor and ensure that the agent lives up to the obligations specified in such a contract. The hidden actions of the agent or asymmetric information provide the basis for the occurrence of *moral hazard.* When any of the two parties stand to gain from acting contrary to the principles laid out by the agreement or simply by taking on extra risks because the other party bears a part or the entire risk. The hidden information is related to other agency problems as *adverse selection*, where an agent does not shed light on information that is in the best interest of the principal but does not serve the agent's interests. This occurs when performance-contingent incentives lead to underreporting of undesired outcomes. Two of the classical methods to solve these problems are rules and regulations accompanied with sanctions or control and surveillance. However, rules and regulations have side effects for the agent that will feel constrained and unmotivated and hence less productive. These negative enforcement strategies are costly to implement and maintain. Instead, the principal can provide economic incentives in the agents contract to motivate the agent to act in the interest of the principal. Pay for performance is a common way of creating this alignment of interests between the principal and the agent. There has been empirical evidence showing that employees take their compensation into consideration when they make a decision at work (see Besanko et al., 2007, 480). Another option for the principal is to revise the agent's contract periodically to assess the work rate of the agent and see if there should be made adjustments to the contract. This ensures a linkage between putting forth effort and being rewarded as effort in one period is being rewarded in the next period and so on.

5.2 Risk

When the principal lets the agent's pay depend on performance he imposes risk on the agent. There can be a number of factors that influence the performance of the agent that are beyond the control of the agent. A sales person's performance is dependent on several factors including the quality of the product he is selling and the overall demand in the market he is selling to. This implies that his performance-based income is volatile beyond the effort exerted and hence far more risky than a fixed salary. Since the agent most typically only has one source of income he is assumed to be risk averse, meaning he has to be compensated to take on risk opposed to a risk neutral agent who is indifferent to risks (Besanko et al., 2007, 469). The principal on the other hand is expected to be risk neutral since the risk of the principal is divided amongst many agents (Bregn, 2004a, 10).

When the principal decides to let some or the entire pay of the agents depend on his performance, the agent needs to be compensated for this risk. On the other hand the principal gets more of the behavior that is rewarded. The performance-contingent incentive comes with a risk and the optimal level of incentives depends on the balance of these two considerations (Besanko et al., 2007, 473). The risk averseness of the agent follows the assumption that marginal utility has a diminishing relation to his wealth; the last dollar earned is less important than the one before and far less important than the first one. Agents are assumed to be risk averse however the level of compensation demanded for adding risk varies from agent to agent as humans deal differently with uncertainty (Besanko et al., 2007, 467). Since the principal doesn't know these individual risk profiles it is impossible to make optimal incentive contracts for every employee. By choosing performance goals with little variation the principal can reduce the risk exposure of the agents. Another option is to use a team compensation practice which reduces the risk of the individual agents however this approach comes with the downside of potential "free-riding".

5.3 Dynamic contracts

Classical P/A-theory only accounts for a single contract period however real life accounts for a dynamic setting with periodical renegotiations. These periodical renegotiation opportunities are also coined *career concerns* (Prendergast, 1999, 51). A dynamic setting undermines the importance of the hedonistic egoism of both principal and agent since they have reputational concerns with regards to the periods to come. The principal has to treat the agent "fair" and in accordance with the contractual agreement and the agent has the career concerns in mind regardless of whether the agent's future is within the company or in another firm. In a dynamic setting the agent might have an incentive to reduce his effort if he is remunerated based on performance measures with threshold bonuses that are either set too high or too low. The high target, a sort of all or nothing bonus, seems unrealistic, and a low measure induces the agent not to excel the target by too much, since this can lead to higher target values for the periods to come. This second argument falls in line with a tendency that has been referred to as the "ratchet principle" of economic planning where current performance is used as a criterion in determining future goals. Agents face a dynamic trade-off between rewards from attaining aims in the current period and future losses from the assignment of higher targets in the periods to come (Weitzman, 1980, 303).

5.4 The measurement of performance

The measurement of "true" performance can have problems when faced with distorting factors, that are beyond the control of the agent, as mentioned earlier. Further, problems arise when the performance measures do not reflect all the job aspects expected from the agent. Some behaviors are measured and rewarded and some are not. This issue is at the core of the *multitasking* principle. Rebitzer and Taylor (2010) conclude on the Holmström and Milgrom (1991) paper on multitasking that:

> The central point of their paper is both simple and profound: when an agent performs multiple tasks, incentives must perform the double duty of inducing appropriately high levels of effort generally and inducing a desirable allocation of an agent's attention across the various tasks inherent in the job. (30)

The agent is faced with the scarcity of time and will therefore allocate his effort to the tasks that are measured and rewarded (Besanko et al., 2007, 476). The principal signalises importance of the measured job aspects and the agent stand to gain from performing well on these assignments. However, performance measures might not necessarily reflect the overall goals and desired purposes and not all aspects are equally suitable for measuring in the first place. This issue is particularly important for performance-contingent incentives that often reflect very specific job aspects. There is one obvious solution to this problem, which is to simply not use performance-contingent incentives if this leads to a more desirable allocation of work effort. This would in turn lead to a lower work rate according to P/A theory because of the lack of alignment of interests and incentive effect. Another solution is to adjust the expectations of the agent so that all job aspects are equally measurable or to design the job in a fashion so that every agent is responsible for one or very few aspects somewhat similar to what Taylor (1947) advised in *Principles of Scientific Management*. Taylor (1947) provided a full management theory build on the measurement of agents every movement to eliminate inefficiencies mainly at the assembly line. This approach comes with other undesired effects as the meaningfulness or sense of purpose is highly reduced with extreme specialisation, which in turn leads to less motivated workers and lower work effort. Extreme specialisation was a practise that was highly popular and produced great results following the invention of the assembly line. However, as presented previously and worth repeating: "Pretty much every economy around the world has a low or declining share of manufacturing jobs" (Kenny, 2014).

When selecting the performance measures for more complex jobs one has to take several aspects into consideration; complete contracts are impossible, hidden information, setting precise goals for all tasks, the variation in measurability of tasks and the difficulty in setting measures that truly reflect the underlying long term visions and goals of the principal without rewarding undesirable behavior. Furthermore as stated in this maxim by Cameron (1963):
"Not everything that counts can be counted, and not everything that can be counted counts".

The maxim beautifully reflect that aspects of high importance are not necessarily measurable, and that it might be compelling to focus on quantitative aspects, however they may not really hold much value especially in a long term perspective. This leads to another problem built in to the discipline of measuring. There is an inherent short-termism in defining various metrics of measurements of performance since they are often evaluated on annual, biannual or quarterly basis. Quantitative aspects are often preferred since they are more easily counted, analysed and displayed in charts and graphs over qualitative aspects which per definition is harder to define and is based on subjective evaluation.
The difficulty of setting performance measures that "truly" reflect the interest of the principal is challenging. This issue was at the centre of the Kerr (1975) article with the compelling title: "On the folly of Rewarding A, While Hoping for B". Kerr (1975) demonstrates that many companies, organisations and society as a whole very often reward one type of behavior while in fact hoping for a different behavior.
Keeping the issues from the previous sections in mind the following aspects should be considered and ideally met when setting performance measures:

- The amount of external influences on the measure should be limited, to factor out as much noise from the "true" ability and performance of the agent which further limits the risk transferred to the agent.
- The measure should reflect all the aspects the principal wants the agent to take on or there should be individual measures for every single task, to avoid downgrading of unrewarded tasks.
- The principal must ensure that the agents can't engage in activities which benefit the agent at the cost of the principal.

Unfortunately, setting such measures relies on a tough and costly analysis which should be considered carefully before deciding to start the implementation of pay for performance.

5.4.1 Absolute or relative performance measures?

When designing the performance measures the principal has to decide whether the measures should be absolute or relative. An absolute measure is set ex ante and is a fixed target value. This makes it easy to administer since it is objective and verifiable to a third party. As mentioned previously, setting fixed target values imposes risk on the agent, since factors beyond the control of the agent influences the chances of reaching them. The "ratchet principle" of economic planning states that agents faces a dynamic trade-off between rewards from attaining aims in the current period and future losses from the assignment of higher targets in the periods to come (Weitzman, 1980, 303). By setting threshold values the agent has very little incentive to exceed a target value without the prospect of reaching the next threshold. Furthermore, the agent would not want to exceed a target with a large margin and thus the agent has an incentive to speculate in the timing of producing results.

The other option is to use relative performance measures by relating the performance of the individual agent to the overall performance on comparable tasks. This method factors out risk aspects relating to overall market fluctuations. Such relative measures provide other issues. The underperformance of co-workers leads to higher pay levels for the individual agent and hence creates incentives that go against information sharing and cooperation in general (Besanko et al., 2007, 478). Active negative behaviors are more easily avoided if the actions of individuals are observed by the colleagues (Bregn 2004a, 16). Applying relative performance measures can lead to collective consensus of low performance of agents to more easily meet individual performance goals (Bregn 2004a, 16).

5.4.2 Objective or subjective goals?

As mentioned above the advantage of objective measures are that they are verifiable and can be tested by a third party. However, this is not always feasible and in these instances a third party can be substituted with an informed person who can perform post hoc evaluations. This implies that the performance is observable however not necessarily verifiable since it is based on a subjective assessment by a supervisor or some other informed party. This leaves some uncertainty regarding

the reliability of the evaluations. There is considerable evidence that subjective assessment result in biases. When evaluations are subjective, workers are likely to waste valuable resources as work time when currying favours with their bosses (Prendergast, 1999, 9). One problem is *leniency biases*, where supervisors are reluctant to give bad ratings to workers leading to somewhat useless assessments (Prendergast, 1999, 30). Another, related problem with subjective evaluations is the *centrality bias*, where supervisors compress ratings around some norm instead of truly distinguishing good and bad performance (Prendergast, 1999, 30). This leads to lower levels of motivation since putting forth effort and being rewarded is disconnected with a clear undermining of what the *expectancy theory* would refer to as the *instrumentality*.

Explicit contracts result in agents optimising relative to the contract. While subjective assessments may be tainted by biases or workers currying favours. According to some theorists a solution is to use objective and subjective criteria's in a complementary way thereby reducing the pitfalls in both assessment strategies (Bregn, 2004a, 16).

5.4.3 Broad or narrow goals?

The considerations related to absolute or relative measures are by large the same when choosing individual or collective goals. Broader measures that focus on many aspects tend to reflect factors beyond the control of the agent and hence impose risk on the agent for which he needs compensation. Broader measures will encourage cooperation and suggestions for improvements of practises, as agents are rewarded on the basis of the revenue of a division or company profits etc. Broad measures can also lead to "free riding" as agents only receive a marginal utility for putting forth extra effort. This negative effect is reduced if the actions of individuals can be observed by the colleagues. However, collective surveillance and enforcement is moving focus from the actual work and is growing in significance with size of peer groups.

A particular problem with narrow measures arises from employees having specific knowledge because, in such cases, narrow measures do not provide incentives for the employee to utilise her information (Raith, 2008). This trade-off between risk and distortion effects the performance measures used and the weights put to different measures in incentive plans.

Surveillance and subjective evaluations can be used to reduce the above mentioned problems related to risk and multitasking as a result of performance measures. However, surveillance is cumbersome and costly. Therefore, it is necessary for any company to put consideration into how these instruments – surveillance, subjective evaluations and performance measures – are used.

5.5 Partial conclusion

This chapter provided an account of the P/A-theory's view on motivating subjects by providing performance-contingent incentives with the different related issues. Both principal and agent are expected to be utility maximizing with contradicting interests. The interests are aligned through contracts where economic incentives can create a link between putting forth effort and receiving a reward. However, some aspects are usually beyond the control of the agent which imposes risk on him for which he needs compensation. The employer gets more of the rewarded behavior and it is therefore a trade-off between risk and incentives. When the principal decides how to optimally construct contracts he is faced with two main issues relating to the available information regarding the effort of the agent and the possibility to set "good" performance measures. The informational aspect occurs when there is information asymmetry leading to problems of *hidden information* and *hidden actions* with agents taking excessive risks, *moral hazard,* or keeping information from the principal also coined *adverse selection.* Three main aspects should be considered and ideally met when setting performance measures. The amount of external influences on the measure should be minimised. The measure should reflect all the aspects the principal wants the agent to take on or there should be individual measures for every single task, to avoid *multitasking.* The principal must ensure that the agents cannot engage in activities which benefit the agent at the cost of the principal. A proper analysis of all of these aspects is costly and should especially be considered when introducing pay for performance. The performance measures should reflect the desired goals, and the considerations for different designs is summarised in the Table 1.

Table 1. Summary of considerations when choosing performance measures

Performance measure	**Pros**	**Cons**
Absolute	Easily verifiable to a third party at a low cost and it is easy to enforce.	External factors can influence the measure which imposes risk for which the agent needs compensation. The "ratchet principle".
Relative	Factors-out risk aspects relating to overall market fluctuations	Discourages agents from helping co-workers. Collective consensuses of low performance to more easily meet individual targets.
Objective	Easily verifiable to a third party at a low cost and it is easy to enforce.	It is not always possible to set objective measures.
Subjective	Can be used when objective measures can't.	Agent can be currying favours for the evaluating part. Leniency bias where supervisors are reluctant to give bad ratings. Centrality bias where ratings are compressed around a norm.

Broad	Can support cooperation for shared goals	External factors may influence leading to risk and extra compensation. Potential free riding.
Narrow	Provides the agent with a clear purpose.	Potential multitasking where non-rewarded aspects are neglected and lack of incentives for cooperation

This chapter has been focused on arguments for and against applying performance-contingent incentives, and the considerations for an actual design from the perspective of the P/A-theory which is a classical economic theory. The opposing theories that have been presented, are founded within a psychological or sociological tradition, and provides a very different view on how to motivate subjects to perform better. This next section is dedicated to this clash between these different strands.

6. Extrinsic and Intrinsic motivation

There are two different theoretical approaches to what drives, directs and maintains behavior. SDT is a proponent of intrinsic motivation whereas P/A theory is a proponent of extrinsic motivation and both set forward substantial critique of the other. *The Crowding Theory* formulated by Bruno Frey is focused on how extrinsic motivation influence intrinsic motivation. This is followed by the defence of P/A-theory and the arguments put forward by Kunz and Pfaff (2002), as a response to the growing empirical evidence suggesting negative effects of monetary rewards.

6.1 The Crowding theory

Classical economic theories consider extrinsic motivation with an underlying assumption that comes close to an economic law. Namely, that providing monetary incentives increase performance. Major schools in psychology, on the other hand, emphasise the motives that come from within the person.

The *crowding theory* states that there is a systematic interaction between extrinsic and intrinsic motivation and it is generalised in two respects:

1. All interventions emanating from outside the person considered, both positive and negative, may affect intrinsic motivation.
2. External interventions may *crowd-out* or *crowd-in* intrinsic motivation (Frey & Jegen, 2000, 4)

The crowding-in effects occur when the external regulation is perceived as informational or supportive as is the case of positive feedback. The intrinsic motivation of the subject is either unaffected or in some cases even supported. This effect is explained by SDT as regulations that support the human needs for competence affirmation and sense of relatedness, which are aspects that also facilitate the internalisation of external regulations.
The crowding-out effect suggests that there are situations where it is not sensible to use the price mechanism to obtain a higher supply provided that intrinsic motivation is valuable or Frey and Jegen (2000) put it: "Monetary incentives crowding out the motivation to undertake an activity may be considered a major anomaly because it predicts the exactly reverse reaction that the relative price effect on which much of economics is grounded" (3).
If an agent perceive a regulation to be controlling and it diminishes the feeling of autonomy, the external regulation undermines the intrinsic motivation. Further, it will "push" the reasoning for the behavior from being intrinsically to being extrinsically driven. These instances of crowding-out or negative consequence from providing rewards has been called *"The hidden cost of rewards"* (Lepper & Greene, 1978), *"Overjustification hypothesis"* (Lepper, Greene & Nisbett, 1973) or simply *"Corruption effect"* (Deci, 1975). This extensive line of work on the subject has been accepted by economists as a part of a wider concept of human motivation but of little significance empirically and insignificant in relation to standard economic P/A-theory (e.g. Alchian & Demsetz, 1972; Fama & Jensen, 1983). The price effect holds that, external intervention raises performance by imposing higher marginal cost on shirking, a sort of disciplining effect, without any effects on the marginal benefit of performing thereby ignoring any changes in intrinsic motivation (Frey & Jegen, 2000, 5). If you reward a behavior you get more of it: "the most fundamental economic "law", namely that raising monetary incentives increases supply" (Frey & Jegen, 2000, 3).

Laws, rules, rewards and fines, a sort of sticks and carrots, approach is to a very large extent used to guide our behavior in the workspace and in society as a whole.
The *crowding theory* views the behavioral regulations as an expression of extrinsic motivation at work and distinguishes these regulations as either rules or prices. A performance-contingent incentive is per definition a prize, expected to motivate a subject to achieve the goals that has been assigned. The crowding theory states that external intervention in the typical case is expected to have two separate effects on the agent's performance. On one hand, the price effect has a positive effect on the agent's performance and if the external intervention also positively effects the intrinsic

motivation then the two effects work in the same direction. On the other hand, when the external intervention undermines intrinsic motivation the *crowding out effect* is at work with a negative effect on the agent's performance. When the price effect is dominated by the crowding out effect the overall effect of external interventions reduces the agents performance level (Frey & Jegen, 2000, 5). Whether the intervention is beneficial from the principal's point of view is decided by the magnitude of these two opposite effects.

According to the *overjustification hypothesis* reported by Lepper, Greene and Nisbett (1973) there is another reason for the crowding out effect. The self-perception theory suggests, that people do not have direct access to the motives behind partaking in an activity and therefore observe their own behavior and thereafter try to justify them. If the extrinsic reasons for the behavior are salient the person will attribute the behavior accordingly, even if the person was intrinsically interested to begin with. The focus is shifted from the activity itself to the reward and the attractiveness of this reward. The person will be less inclined to partake in the activity in the future, if engaging is not rewarded and hence the crowding out of intrinsic motivation in the short run and the long run.

The crowding out of intrinsic motivation is not necessarily a problem when applying pay for performance. If the price effect outweighs the negative effect it leads to more overall motivation. However, the different types of motivation predict different effects on performance in relation to the nature of the task. The autonomously motivated employee is predicted to perform better on more complex tasks whereas controlled motivation yields the same results or slightly better results on simple tasks (Gagné & Deci, 2005, 346). If one further takes the problems discussed earlier in the section on P/A-theory into consideration e.g. the difficulty of setting "good" performance measures and the thereof potential problems of multitasking, hidden information etc. and the crowding out effect all stacks against pay for performance or at least calls for considerable caution when introducing pay for performance**.**

6.1.1 Pay for performance and the crowding theory

The crowding theory suggest that there are two main psychological processes through which external interventions influence intrinsic motivation which is; 1) *impaired self-determination* and 2) *impaired self-esteem* (Frey & Jegen, 2000, 6). Impaired self-determination originates from a perception that the external regulation forces the individual to behave in a specific manner, thus substituting intrinsic motivation with extrinsic control. Impaired self-esteem originates from a lack of acknowledging the intrinsic motivation that the individual carries in the first place. The

possibility to show this interest is removed when given a reward or commanded to comply (Frey & Jegen, 2000, 7).

The interpersonal climate within which rewards are administered has significant influence on the effect of rewards. The agent's perception of how the regulation is administered decides how a regulation effects the intrinsic motivation. This is similar to the process, described in SDT, of internalization of external regulations when the needs of the employee are supported. When the regulation is administered in a controlling fashion it undermines intrinsic motivation. Contrarily, if the regulation is administered in an informational or supportive way it will tend to be autonomy supportive and increase intrinsic motivation (Frey & Jegen, 2000, 7).

When rewards are given for some particular behavior it tends to have an undermining effect on intrinsic motivation due to the crowd-out effect. When rewards are given as a sign of the employees overall contribution and competences it tends to enhance intrinsic motivation. Pay for performance is by large dependent on the measurability and tends to be very specific in terms of reaching specific goals and standards and thus tends to have an undermining effect.

Whether agents consider a reward as "fair" depends on the relative difference in rewards given to an employee and his or her peers. If the principal provides the same rewards to a peer group in spite of relative differences in performance, it tends to have a negative effect on high performing subjects because of the absent recognition. This is a problem that occurs in work settings with rigorous pay systems when there is no option to provide a differentiated pay. This is quite similar to the problems related to the centrality bias, described in the previous section (Frey, 1997, 433).

Frey (1997) argues that different regulations have different effects. Rules or sanctions or a combination of the two has a larger tendency to undermine the sense of self-determination than prizes and rewards. This is based on the notion that a prize provides the subjects with a choice of not pursuing the reward whereas a rule does not provide this option of self-determination. However, different employees perceive such prizes differently. Some subjects might actually respond to the prize in the same fashion as with rules leading to a crowding out effect (Frey, 1992, 168; Frey 1997, 432). It is worth mentioning that many rules are enforced with prizes or negative prizes and thus tend to have the same undermining effect.

According to the crowding theory, pay for performance is perceived as potential rewards the employee can decide to pursue or choose to ignore.

6.2 The effects of extrinsic motivation on intrinsic motivation

Alexis Kunz and Dieter Pfaff published an article in 2002 based on their concern with the growing body of work on the detrimental effects of rewards. They discuss how and if this should influence the assumptions of the P/A-theory and if theories on intrinsic motivation should ultimately be incorporated into the P/A-theory (Kunz & Pfaff, 2002, 276).

According to Kunz and Pfaff (2002) the core issue of intrinsic motivation is that it has too many definitions. It is essentially a hypothetical constructed term that tries to relate unobservable psychological processes inside subjects by observing behavior and then draw conclusions without any "real" knowledge of the relation (Kunz & Pfaff, 2002, 279).

They criticise the empirical research for not trying to distinguish the needs for competence and self-determination from the sensation of total satisfaction from a task. In order to test CET they find it necessary to make intrinsic motivation operational so that it can be tested independently from the behavior it initiates. In contrast, the tests have been on the effect that cannot be attributed to extrinsic motivation and therefore was inferred to be intrinsic motivation. According Kunz and Pfaff (2002): "this is a circular theoretical deduction which lacks predictive power and empirical backing "(284). Kunz and Pfaff (2002) consider the theories on intrinsic motivation and the crowding theory to be lacking the essential predictive power of outcome. The theories cannot predict which assumptions are essential for a factor to be considered to be intrinsically or extrinsically motivating. The crowding theory does not provide criteria's that ex-ante can assess if a reward will be perceived as controlling, informational or insignificant. The theory can only post-hoc analyse the effect of a reward, when it has been presented which makes the theory difficult to work with operationally. The overjustifaction theory also lacks criteria's that will enable to decide the optimal level of affirmation of certain behavior. When an employee is not rewarded for an activity, the reasoning for performing that activity can easily come from social or organisational concerns instead of intrinsic motivation. By making a clear distinction between intrinsically and extrinsically driven behavior the theories on intrinsic motivation create an understanding of motivation as either internally driven or a result of the surrounding environment. However, it is often the case that it is a mix of the person and the situation. This leads Kunz and Pfaff to conclude that neither intrinsic motivation nor the crowding effect can be used to explain the empirically proven detrimental effect of rewards on performance in economic theory (Kunz & Pfaff, 2002, 283).

Kunz and Pfaff (2002) criticise the theories on intrinsic motivation on several accounts. One point of critique is that they do not have a clear definition of intrinsic motivation. However, this is not necessarily the case, for instance the SDT has a clear definition. SDT have made intrinsic motivation operational in the need for competence, relations and autonomy and shown a positive relation between fulfilling these needs and persisted motivation. This could arguably provide SDT with predictive power. Further SDT posits a self-determination continuum between amotivation and intrinsic motivation. This is not a rigorous distinction between internally and externally driven motivations, but a descriptive continuum that follows the degree of internalisation.

Kunz and Pfaff (2002) criticise on an empirical basis the theories that emphasise intrinsic motivation relies on. The theories lack external *validity* and they cannot be used for *generalisations*. Many of the investigations in the field have methodical deficiencies since they have been performed in experimental research environments and not in the setting, they proclaim to describe. Some of the subjects in the experiments are children who cannot be expected to explain their behavior on the level that is required given the complexity of the theories. Further, there is no proof of the transferability of results obtained with children to adults. There is no or very little consideration for the long term effect rewards have on performance even though some investigations show a lessening of the detrimental effects over time. The results obtained are open to *different interpretations*. Other psychological theories can interpret the results from the experiments in a different way due to the fact that conclusions are drawn on unobservable psychological processes. The *results are inconclusive*. There has been considerable empirical research that does not support the theories on intrinsic motivation and research that finds that the detrimental effect of rewards on performance is highly dependent on how it is administrated and the language used (Kunz & Pfaff, 2002, 284). A further critique is that the results of the experiments by large reflects the choices of the researchers which explains the altering results obtained in both experiments and the meta-analyses on the topic.

6.3 Partial conclusion

On one hand, the crowding theory focuses on the effect of external regulations on intrinsic motivation. A regulation which is perceived as controlling will crowd-out the intrinsic motivation, whereas an informational or supportive regulation will crowd-in or support the intrinsic motivation. The effect of a regulation is hence determined of how it is perceived. The effect of a reward is

highly influenced by the extent that it is dependent on specific results. The more dependent on results the larger the crowding-out of intrinsic motivation. This effect is also influenced by the context and how it is implemented and administrated. This is much in line with SDT which states that successful administration of rewards depends on how it interacts with basic psychological needs for competence, autonomy and relations.

On the other hand, Kunz and Pfaff (2002) criticise the term intrinsic motivation. According to Kunz and Pfaff (2002) intrinsic motivation is a hypothetical construct of unobservable psychological processes supposedly at play when there is no extrinsic motivation based on a sort of circular deduction. Kunz and Pfaff (2002) accept that rewards can indeed have detrimental effects on performance. However, the assumptions for these detrimental effects do not collide with P/A-theory which should not be revised. Further, they do not deem it necessary to include the terms intrinsic motivation and the crowding effect into P/A-theory (Kunz & Pfaff, 2002, 290). Kunz and Pfaff (2002) raise some points of critique of the empirical basis that the theories on intrinsic motivation are founded on. They criticise the empirical research for lacking external validity since they are based on experimental settings. Therefore, it is difficult to make generalisations applicable to work settings, and experiments performed with children are not necessarily transferable to adults. Furthermore, the results are open to interpretations and they are inconclusive, since the results are varying significantly across different experiments. These points of critique are not necessarily unwarranted and will be brought up in the following section on meta-analysis by Eisenberger and Cameron (1996).

7. Research studies and Meta-analyses on intrinsic motivation

There has been an extensive research in the field of intrinsic motivation. There has been considerable empirical research in support of the conclusions that have led to the formulation of the CET and SDT and many experiments thereafter replicating these findings. Kohn (1993a) and Kohn (1993b) find that in the field business and education the use of rewards have often had a detrimental impact. However, there has been considerable controversy in the validity and reliability of these findings. The main dispute regards how results obtained in experiments relate to work settings and whether there is any transferability of results obtained with children to adults. This section is dedicated to the different views on the effect of rewards. The literature on psychology is typically concerned with the importance of intrinsic motivation whereas several economists find these effects

to be either limited or highly dependent on specific conditions which can easily be avoided. A proponent of this view is Eisenberger and Cameron[4] (1996) who calls for considerable reservations regarding the detrimental effects of rewards.

7.1 "Detrimental effects of rewards – Reality or myth" by Eisenberger and Cameron (1996)

Eisenberger and Cameron (1996) put forward a general critique of humanistic psychologists and cognitive-social psychologist to be highly influenced by a Romantic sensibility and thus putting an emphasis on aspects as individual freedom, self-expression and self-fulfilment (Eisenberger & Cameron, 1996, 1155). Eisenberger and Cameron (1996) state that theories of intrinsic motivation have to a large extent been driven by the researchers implicit Romantic conceptions about human motivation: "Systems of reward for improved task performance as promoted by behaviourally oriented psychologists, are seen as inherently self-defeating because they interfere with the desire to explore one's own potential" (1155).

Eisenberger and Cameron (1996) state that the claims that reinforcement of behavior negatively effects intrinsic motivation and creativity, have led many psychologists to assert that:

1. intrinsic interest and creative behavior depend on personally directed exploration, free from social control;
2. the supposed pragmatic benefits of behavioral technology for education, business, psychotherapy are often negated by inherent negative side effects;
3. behaviorism is flawed by basic misconceptions of human nature. (1154)

According to Eisenberger and Cameron (1996) the acceptance, that reinforcement negatively effects intrinsic motivation and creativity, relies on a preponderance of empirical evidence assumed to be in support of these claims, is questionable. Eisenberger and Cameron (1996) find the statistical evidence to be inconsistent with these assertions and their general validity:

> However, an examination of almost 100 relevant studies carried out over the last quarter of century revealed considerable variability of results; reports that reward have decremental

[4] Eisenberger and Cameron (1996) revised the Cameron and Pierce (1994) publication. The two articles and the subsequent conclusions are based on the same meta-analysis and therefore treated as one meta-analysis, as noted by Deci, Koestner and Ryan (1999, 632).

> effect, no effect, and an incremental effect on intrinsic task interest have all been frequently obtained. This diversity of findings is likely to result from one or more of the following conditions: (a) Differences between rewarded and nonrewarded groups are small relative to individual differences within groups, (b) group differences are actually random variations from a true difference that falls close to zero, or (c) group differences are greatly influenced by the details of how rewards is administered. (1157)

Eisenberger and Cameron (1996) criticise the research literature and find that the detrimental effect of rewards to be more limited and conditional than commonly supposed:

> "… (a) detrimental effects of reward occur under highly restricted, easily avoidable conditions; (b) mechanisms of instrumental and classical conditioning are basic for understanding incremental and decremental effects of reward on task motivation; and (c) positive effects of reward on generalized creativity are easily attainable using procedures derived from behavior theory" (1154).

In the midst of Eisenberger and Cameron (1996) meta-analysis, several articles criticised it for methodological inadequacies (Kohn, 1996; Lepper, Keavney,& Drake, 1996; Ryan & Deci, 1996). The most significant points of critique were summarized by Deci et al. (1999) and they are:

1. collapsing across cells where there were significant interactions for theoretically meaningful variables for example, initial task interest, positive versus negative feedback, and informational versus controlling administration of rewards without doing moderator analyses for any of those variables,
2. using inappropriate control groups in several comparisons,
3. omitting nearly 20% of the studies as outliers rather than attempting to isolate the cause of the variability in effect sizes. (633)

While Eisenberger and Cameron (1996) is criticising the empirical evidence suggesting a negative effect from rewards, Eisenberger and Cameron (1996) and their research is itself the subject of substantial critique.

7.2 "A meta-analytic review of Experiments Examining the effect of extrinsic rewards on intrinsic motivation" by Deci et al. (1999)

Previous to the Eisenberger and Cameron (1996) meta-analysis there had been three meta-analyses (Rummel and Feinberg, 1988; Wiersma, 1992; Tang & Hall 1995). These three meta-analyses found considerable evidence in support of a general hypothesis that tangible rewards undermine intrinsic motivation for that activity. Deci et al. (1999) finds that: "Although each of these studies had methodological shortcomings, the consistency in their results is noteworthy" (632). Contrarily, Eisenberger and Cameron (1996) find the undermining effect to be largely a myth.

Deci et al. (1999) found all four previous meta-analyses to have significant methodological problems. Due to the controversy regarding the conclusions of the four previous meta-analyses, Deci et al. (1999) re-examined the research literature. The examination led to several interesting insights.

Tangible rewards have a significant negative effect on intrinsic motivation whereas positive feedback has a positive effect on intrinsic motivation. Positive feedback enhance intrinsic motivation and affirms competence if administered in an informational way. Conversely, if administered in a controlling way it diminishes intrinsic motivation and counteracts the positive effects of the information, leading to an undermining of intrinsic motivation (Deci et al., 1999). Numerous aspects of the interpersonal environment were shown to affect intrinsic motivation depending of being perceived as controlling or informational. Deadlines, evaluations, and imposed goals were found to undermine intrinsic motivation, whereas the provision of choice and the acknowledgment of feelings have been found to enhance intrinsic motivation (Amabile et al, 1976; Smith, 1975; Mossholder, 1980; Zuckerman et al., 1978; Koestner et al., 1984).

Based on previous research[5], Deci et al. (1999) suggests that making rewards more informational requires:

1. minimizing the use of authoritarian style and pressuring locution,
2. acknowledging good performance but not using rewards to try to strengthen or control the behavior,
3. providing choice about how to do the tasks, and
4. emphasizing the interesting or challenging aspects of the tasks. (656)

[5] By Deci, Nezlek, et al. (1981), Koestner, Ryan, Bernieri, and Holt (1984), and Deci et al. (1994).

When institutions, as for instance families, schools, businesses, and athletic teams, put an emphasis on the short term and try to control people's behavior, it may be having a substantially negative long-term effect (Deci et al., 1999, 659). Further, when organizations select rewards to control behavior they are likely to come with greater surveillance, evaluation, and competition, which have all been found to undermine intrinsic motivation (Kohn, 1993a; Deci & Ryan, 1985; Deci et al., 1999).

7.3 "Intrinsic motivation and extrinsic motivation jointly predicts performance a 40- Year Meta-Analysis" by Cerasoli, Nicklin and Ford (2014)

The most recent meta-analysis by Cerasoli et al. (2014) focus on two moderators for the interrelationship among intrinsic motivation, extrinsic motivation and performance. The moderators are quality vs. quantity with regards to performance type and the incentive contingency directly performance-salient vs. indirectly performance-salient. According to Cerasoli et al. (2014):"The distinction that may be the most critical is that of "quality" and "quantity" (J. P. Campbell et al., 1993)" (982). Intrinsic motivation was more influential for *quality* than extrinsic incentives and extrinsic incentives accounted for more of the difference in *quantity* performance criteria than did intrinsic motivation (Cerasoli et al., 2014, 996). Tasks emphasizing performance *quality* will have a strong link to intrinsic motivation. The reason is that quality-type tasks tend to require a higher degree of complexity and engagement of more skill, which commands a greater deal of personal investment (Cerasoli et al., 2014, 982). Quantity-type tasks can also be interesting (Cerasoli et al., 2014, 982), but tend to be lower in complexity, and require less personal cognitive investment (Gilliland & Landis, 1992). Cerasoli et al. (2014) found that:

> when extrinsic incentives were present *but only indirectly salient to performance*, intrinsic motivation were a better predictor of performance. Conversely, when incentives were present *and were directly salient to performance*, intrinsic motivation was a poorer predictor of performance. (996)

Interestingly they found support for the hypothesis that incentives and intrinsic motivation are not necessarily antagonistic. Depending on the type of performance and the contingency of the incentive, incentives coexist with intrinsic motivation. Cerasoli et al. (2014) takes it a step further: "Those arguing against these types of incentive programs do not question their effectiveness: in fact, many note that the incentives are almost *too* effective (G. P. Baker, 1993)" (997). Referring to

the concern that the intrinsic motivation is undermined while an incentive is in place, but once it is gone it leaves no motivation for that activity.

The suitable degree of salience of an incentive should be considered based on the desired behavior. Their research demonstrates the combined impact of incentives and intrinsic motivation is critical to performance (Cerasoli et al., 2014, 1001).

Based on their investigation Cerasoli et al. (2014) bring certain recommendation regarding how and when incentives should be applied, instead of an "either or" approach to incentives:

> Tasks that are straightforward, highly repetitive, and perhaps even less inherently enjoyable, should be more closely linked to extrinsic incentives. On the other hand, tasks that require a great deal of absorption, personal investment, complexity, and overall quality should be less linked to incentives and much more closely linked to intrinsic motivation. Instead, when creativity, autonomy, teamwork, learning, ethical behavior, well-being, and quality are valued, incentives should be framed as less salient. (997)

8. Following hypotheses to the Research Question

As mentioned previously, in the section with the research question, a potential positive effect from providing performance-contingent incentives rests on the assumption that it will increase overall motivation and effort, and that this increase in motivation and effort will result in improved performance. The experiment in this thesis is on the second part of this reasoning namely the effect of introducing incentives compared to a no incentives control group. The literature review provided the expectation that different types of motivation predict different performance in relation to the nature of the task. Intrinsic motivation yielded better performance on tasks that are interesting and has a quality performance measure. Extrinsic motivation yielded better performance on tasks that are simple and has a quantity performance measure. This distinction is central to the formulation of the following hypotheses to the research question. The research question is: **What will be the effect of performance-contingent incentives on performance in an experiment with students?**

The research question is clarified by the formulations of two following hypothesis that will be tested in the experiment and subsequently answered.

- **There is a positive effect from providing performance-contingent rewards when the task can be characterised as simple.**

- **There is a negative effect from providing performance-contingent rewards when the task can be characterised as complex.**

9. Methodology

The following section is a presentation of the methodical considerations regarding the experiment that serves as the empirical aspect of this thesis. It consists of both general considerations regarding the research instrument and particular considerations regarding the design and execution of the standardised experiment.

9.1 Validity, Reliability and Methodological considerations

Validity is a term developed in a positivistic tradition, which relates to empirical concepts as evidence, objectivity, truth and universal laws amongst others. Within quantitative research validity is defined by Joppe, here cited from Golafshani (2003), as:

> Validity determines whether the research truly measures that which it was intended to measure or how truthful the research results are. In other words, does the research instrument allow you to hit "the bull's eye" of your research object? Researchers generally determine validity by asking a series of questions, and will often look for the answers in the research of others. [Joppe, 2000, 1]. (599)

The validity is hence a matter of, whether the means of measurement are accurate and whether they are actually measuring what they are intended to measure.

When performing these standardised tests, it comes with the underlying assumption that the derived conclusions can be applied to subjects in similar situations. But how *reliable* are the results if repeated later in time is of crucial importance. Reliability within quantitative research is defined by Joppe, here cited from Golafshani (2003), as:

> The extent to which results are consistent over time and an accurate representation of the total population under study is referred to as reliability and if the results of a study can be reproduced under a similar methodology, then the research instrument can be considered as reliable. [Joppe, 2000, 1]. (598).

The *reliability* of the results according to Kirk and Miller (1986) relate to:

1. the degree to which a measurement, given repeatedly, remains the same
2. the stability of a measurement over time
3. the similarity of measurement within a given time period. (41)

The experiment as a scientific research method has as the aim and the advantage that it seeks to reveal the causal relation between variables. This is approximated by controlling all variables except from the independent and by randomized participation (Aronson, Wilson & Brewer, 1998, 102). In an experiment, it is important to make efforts to ensure that all subjects have the same information in an attempt to minimize any confounding effects.
Standardized experiments have the advantage that they make it possible to repeat previously conducted experiments. This can be beneficial in several ways, partly because it can reinforce a theory by confirming a hypothesis by replicating it and partly because it may point to deficiencies in the theory. The selected design is crucial to whether the obtained results are useful when performing a standardised experiment. A good experimental design is crucial to ensure high validity.
In some experiments, it may be of essential importance to the experiment that the researchers retain information about what the real experiment is about (Aronson et al., 1998, 120). However, some of the problems related to this deception can be offset by debriefing the participants after the experiment (Aronson et al., 1998).

Experiments can be criticized for lacking internal validity if there are too many sources of error - as the results achieved cannot be assumed to be due to the bonus, but may be due to various error sources. A prerequisite for internal validity is that the participants have been randomly allocated. According to Field and Hole (2003, 63) the generalisability of the experimental method refers to, whether the findings achieved in an experiment can be generalised to other subjects in other trials, in another time and another place. The external validity is related to what other populations and settings that the treatment and measurement variables that a potential effect can generalised to. Mook (1983) points out that there often prevails a misconception regarding the importance of generalisation. According to Mook (1983) it is particularly significant that many experiments enrich our insights about what might happen - for example, the behavior that could follow a given influence. It is not the same as using experiments to conclude and point out what will happen (Mook, 1983, 382). As Mook (1983) further specifies:

> Rather than making predictions about the real world from the laboratory, we may test predictions that specify what ought to happen in the lab. We may regard even "artificial" findings as interesting because they show what can occur, even if it rarely does. (379)

The discipline of hypothesis testing with statistical analysis is an analysis based on assumptions. The assumption for a two-sided significance test for comparing two population means, assuming equal population standard deviations are:

- Independent random sampling either from random sampling or randomized experiment
- Approximately normal population distribution for each group (this is mainly important for small sample sizes, and even then the confidence interval and two-sided test are usually robust to violations of this assumption.)
- $\sigma^1 = \sigma^2$ (In practice, they are not usually relied on if one *sample* standard deviation is more than double the other.) (Agresti & Franklin, 2009, 495)

The significance test of the difference in means is based on two independent samples as the overall sample was randomly selected the two groups can be treated as independent random samples (Agresti & Franklin, 2009, 469). For two-sided tests, like the ones in question, the last assumption is important because the two sub-groups consist of 20 observations so the central limit theorem does not hold. However, when performing two-sided tests the model is robust against violations of the normal assumption even when the sample sizes are small (Agresti & Franklin, 2009, 485). The assumption that the two samples have equal standard deviations, so that a pooled standard deviation estimate is used, is robust (Agresti & Franklin, 2009, 495). According to Winter (2013): “The dictum “more is better” certainly applies to statistical inference. According to the law of large numbers, a larger sample size implies that confidence intervals are narrower and that more reliable conclusions can be reached” (1). When Winter (2013) tested the applicability of the t-test on samples as small as N=3 he found that the t-test was valid if the effect sizes are relatively large. The sample size N=20 in thesis is beyond the definition of extremely small sample sizes of five or less observations Winter (2013, 1) operates with. However, the considerations he raises are very important as: “… large effect sizes are uncommon in the behavioral/psychological sciences” (Winter, 2013, 7). There are two common pitfalls of hypothesis testing. The type I error is the incorrect rejection of a true null hypothesis. The type II error, retaining a false null hypothesis or

failing to detect an effect that is present. For the current thesis all significance test are reported with an alpha level of .05 as it is considered the best way to balance these two potential errors (Winter, 2013, 2). Researchers strive to minimize both Type I and Type II errors, and methodologists have cautioned that a small sample size implies low statistical power with a correspondingly high probability of a Type II error (Winter, 2013, 2). The sample is not "extremely small" but not "especially big" either, which makes it more unlikely to detect small effect differences and likely to cause type II errors of failing to detect an effect that is present. The type I error is also more likely to occur in smaller samples as differences can be inflated "…when the sample size is smaller, a statistically significant finding is more likely to be a false positive" (Winter, 2013, 8). As stated above, the samples are not extremely small. However, any conclusions derived from the hypothesis testing should be treated with caution and be assessed critically and in relation to existing evidence according to Winter (2013):

> it can be argued that if a psychologist observes a statistically significant effect based on an extremely small sample size, it is probably grossly inflated with respect to the true effect, because effect sizes in psychological research are typically small. Accordingly, researchers should always do a comprehensive literature study, think critically, and investigate whether their results are credible in line with existing evidence in the research field. (8)

The purpose of this thesis is not with indisputable certainty to predict whether the same effect takes place in the real world. It is assumed, nonetheless, that the hypotheses and the results can be discussed in relation to existing research, in a way that can act as a contribution to how performance-contingent bonuses relate to performance, while recognizing that there is other ways that the results may be discussed.

9.2 Research method

When analysing the experiment I will apply both inductive reasoning, which is based on empirical observations and deductive reasoning based on logic. Induction is a process of making general reasoning from empirical facts. Whereas, deduction is a method of logic reasoning that is considered valid if it is logically consistent. However, it does not need to be true in the sense that it corresponds with reality only that it is logically consistent with its premises (Thurén, 2008). When used in a combination it is said to follow a hypothetic-deductive method which is a favoured

method in the positivistic or postpositivistic paradigm (Thurén, 2008, 18). The hypothetic-deductive course of this thesis is provided in Table. 2. The scientific method of findings which are considered to be true and valid through data collections, can be hypothesized and thereafter tested follows a hypothetic-deductive method.

The method should be seen from a popperian point of view, where the scientific theory can never be proved to be true since no number of confirming observations can verify a universal generalisation, but are tested by attempts to falsify them (Thurén, 2008, 58).

The experiment was carried out with a control group of a comparable demographic to make it probable that the observed effect, on average, was that of providing a cash incentive on performance. If the hypotheses do not hold we can logically falsify that giving cash incentives leads to increased performance for simple tasks and decreased performance for complex tasks as universal theorems and to be more limited in their scope.

Table 2. The hypothetic-deductive process of this thesis.

1) Definition of problem or question 2) Collection of information 3) Formulation of hypotheses 4) Conduct experiment 5) Do observations 6) Draw conclusion 7) Confirmation => repetition of 4) experiment; if Affirmation => Reformulate 3) hypothesis	1)The effect of performance-contingent bonuses on performance 2) Relevant literature on motivation and performance 3) Does the effect of the incentives positively affect the performance for simple tasks and negatively affect performance on complex tasks. 4) One experiment with three tasks of different nature with 20 students receiving performance-contingent rewards in one group and 20 students in a no-reward control group. 5) Are there any significant differences in mean scores across the two groups. 6) What in the theory and previous experiments can explain the results. 7) Indication that rewards has altering effects on performance dependent on the nature of the task.

9.3 Experimental design

Initially I intended that all participants were enrolled in one Master's Program namely Applied Economics and Finance. However, I disregarded this idea, as the volunteering rate was very low after presenting the experiment to the students. I presented the experiment, both on their internal group on the social media Facebook and in person in the beginning of one of their lectures. Due to a lack of interest from this group of students and the actual time constraints on the production of this thesis I chose a broader criterion for participation. The only requirement for participating is that the

subjects are currently students at CBS. The loosening of this criterion for participation, unequivocally, meant that the two groups were less homogenous, than would have been the case with only students attending the same program and the same semester. However, in spite of this alteration the groups can still be characterised as relatively homogenous on average in terms of age and their current progression of education.

The experiment was conducted with 40 students during one week in October 2016 at Copenhagen Business School (CBS). The students were recruited from the ones who were present during those days and volunteered when presented with the experiment. The only requirement was that they attended CBS. They were told that they were partaking in an experiment with three different tasks and that they should solve them to their best ability with an expected duration of 15-20 minutes based on previous experiences. They were informed that participation was anonymous and that the average scores would be used to compare the CBS students' abilities with Russian students for whom there had already been a similar experiment. There had not been any prior experiment in Russia. This information was provided so there would seem to be a meaningful rationale for partaking without giving away the actual inquiry intended. All participants were debriefed about the actual inquiry after completing the experiment. The participants were "deceived" as they were told that there had been a prior experiment in Russia even though it was not the case. There is an element of deception related to this misguidance of participants. I debriefed all the participants after the experiment and informed of the actual line of inquiry, in accordance with their interest, as some were more interested than others were. Further, all participants were informed of the importance of not giving away information to other potential participants to minimize the potential risk of invalidating the task scores for other participants.

Further they were informed that the experiment consists of three different tasks that fall into three broad categories based on the skills they predominantly require, namely creativity, spatial insight and concentration.

The recruiting process for the rewarded group was the same as for the no-reward group. However once seated in the test-room, they were informed that based on their performance they could receive a bonus. The participants were not informed that the performance levels that triggered a bonus was based on average scores. They were only provided with the raw scores that they needed to match to receive payment. The participants could not see the clock while performing the tasks. This approach was chosen to avoid excessive focus on the clock. Further, to avoid that participants would lose interest in completing a task when having lost the opportunity of a reward. The performance levels

were stated as: “if you solve this problem within 3 minutes you will receive full payment 200DKK or if you solve it within 4 minutes you will receive half of that payment 100DKK”.
The expectation that the rewards have an effect on the motivation and more importantly on the performance of the participants is that the potential rewards of some significance to the participants. Even amongst a relatively homogenous group of people as students, I expect, that there is some dispersion in their disposable income as some work more than others do, some have cheap rental apartments provided by their parents’ etc. meaning that they have different marginal utilities of potential rewards. However, on average, it is expected that the possibility to earn up to 600DKK for 15-20 minutes of task solving, that the participants volunteered for prior to learning anything about possible bonuses should, at least on average, be expected to influence their motivation.
The incentives was provided directly salient to performance which would provide a setting where the incentives might have an undermining effect on the intrinsic motivation the subjects carried in the first place.

The experiment was conducted with one participant at a time due to logistic reasons as there was only one experimenter. Further, to avoid interactions between subjects in an effort to minimise the potential of a confounding effect. The no-reward group was tested first. The average scores from the no-reward group were used for setting performance targets for the rewarded group. The level “good” was reached if the incentivised subjects performed the tasks within the average time of the non-rewarded group. The level excellent was defined as the average minus 25 percent when considering time scores as is the case for the first two tasks e.g. the creative task and the spatial insight task. There was only one trial in both of these two tasks. There was provided a maximum time limit of 10 minutes for the first two tasks, which limits the influence on the mean scores of “slow” subject. This was done to keep the duration of the overall experiment close to the intended time interval of 15-20 minutes. Further, it is as a way of avoiding potential outliers, which needs to be removed from the sample after an appropriate inspection. By choosing this approach the “slow” individuals were included without having a large impact on mean scores. It can be noted, at this point, that all subjects completed the given tasks with the exception of two subjects when performing the *Figure test* or spatial insight task. There was one subject on the bonus treatment and one subject on the no bonus treatment who did not complete the task. These two subjects were assigned a time of 10 minutes.

The concentration task, *Recall last-4 digits*, was based on 14 trials with the highest number of correct trials being the desired aim. For this task the level "good" was defined as the average number of correct trials for the no-reward group and the level "excellent" was defined as the average of the no-reward group plus 25 percent rounded to the nearest full number.

9.3.1 Tasks

The *Candle problem* was chosen as the creative task. It was developed by psychologist Karl Dunker at Clark University and published posthumously in 1945. It centres on functional fixedness where one needs to break this fixedness in order to solve the task. Karl Dunker defined functional fixedness as being a mental block against using an object in a new way that is required to solve a problem. The *Candle problem* has been revitalised by Pink (2010) where he hammers the idea that economic incentives harms creativity and complex problem solving. Based on Glucksberg (1962) finding that rewards magnifies the problem of functional fixedness with a negative effect on performance, Ramm et al. (2013) decided to replicate his initial finding without finding any evidence for an undermining effect.
The participants are given: a small box with *thumbtacks*, a box of *matches* and a *candle*. They are told: "fix a lit candle to the wall (bulletin board) in a way so it will not drip onto the table below it".

The substantial element in providing a valid solution to this problem comes from realising the value of the containers with its own set of abilities and quite literally "thinking out of the box". That in turn leads to emptying one of the boxes and pinning it to the bulletin board and placing the candle in the box. For the rewarded group the level excellent could be reached within 179 seconds and the level good within 239 seconds.

The second task was a spatial insight problem, which bear some similarity to the tasks typically provided on IQ tests within this discipline of spatial reasoning. The *Figure test* requires some reasoning skills connected to the capacity to think about objects in three dimensions and to draw conclusions from information about these objects. The task is presented visually in Figure 2. (the solution was not provided) accompanied with the following text: "Show how you can divide this figure into four equal parts that has the same size and shape, and has the shape of the figure".

Figure 2. *The Figure Test*

If the subjects decide to replicate the figure on the squared A4 and use the given information to draw the smaller figures in proportion the tasks starts to resemble a puzzle of lower complexity where one needs to move the pieces until they fit. For the rewarded group the level excellent could be reached within 178 seconds and the level good within 238 seconds.

The third task, *Recall last-4 digits*, was chosen as the task that focuses mainly on concentration. Ariely et al. (2009) performed a quite similar task, coined Recall last-3 digits, because: "it requires tiresome concentration and we thought that subject who were more highly motivated might be more likely to maintain high levels of concentration. We did not, however, observe any such difference" (12). The subjects were told that they could use pen and paper if they wanted. The game goes as follows, the experimenter reads a sequence of digits in a quick pace, and stops at an unannounced point, and asks the participant to provide the last four digits before the stop[6]. The participants are given 14 trials at the task. For the rewarded group the performance level good was defined as 10 or 11 correct trials and the level excellent was reached if the participant had 12 or more correct trials.

The *Candle problem* (Glucksberg, 1962; Ramm et. al. 2013) and a different version of *Recall last-4 digits* (Ariely et. al, 2009) have been used in previous experiments. These tasks were selected because they represent respectively a complex and a simple task. Further, because these two tasks have been used previously, the results from the current experiment can be interpreted in relation to the previous results. The *Figure test* was chosen because it represents an IQ type of task that requires some cognitive skills. IQ tests are in general believed to provide valuable information regarding the abilities of an individual. IQ tests typically reflect several different aspects besides spatial reasoning. However, the task was chosen, amongst three options[7], because it provides a first impression of being very simple and yet many people do not solve the task as quick as they presume they would. The first two tasks were standardised to the extent that seemed feasible. However, for

[6] See Appendix 1. for more on the presentation of tasks in the experiment and the sequences of digits.
[7] Before the "test trial" I tested two other tasks on two friends in order to find one additional task to include in the final experiment.

the concentration task, *Recall last-4 digits*, the reading of these sequences was repeated numerous times prior to the actual experiment in order to obtain a standard pace that all participants would face. This was done to ensure that all subjects were faced with essentially the same task.
Further to examine that the three final tasks would work as intended there was a "test trial" with four individuals of a similar age, all attending CBS. The test trial was conducted to ensure, that the assignments would be understood in a similar way by all participants. Further the test trial was conducted to ensure that the tasks had the desired degree of difficulty. Furthermore, to ensure that it was feasible to conduct all three tasks in 15-20 minutes on average. This time frame was chosen based on the expectation that the likelihood of getting people to volunteer is negatively correlated to how much time participants have to spend, with this interval as a seemingly fine compromise.

10. Findings and analysis

The main aim of the analysis is to investigate the effect of performance-contingent rewards on the performance compared to the performance of a no-reward control group. The students participating in the experiment solved two tasks of a more complex character and one task of a more simple character. This allows for a comparison for the different incentive treatments dependent on the character of the task. Firstly, there will be an inspection of potential influences on performance from the demographics of the participants. This is done through multiple regression analyses. Secondly, there will be an inspection of the effect of the incentive on the performance for the three tasks. This will provide an indication if the data can support the hypotheses. The hypotheses are that the incentive has a positive influence on performance for the simple task and a negative influence on performance for the two complex tasks. The two hypotheses will be tested by performing independent-two-samples t-tests on task scores dependent on incentive treatment.

The recorded demographics are age, sex and years of studying. In order to determine if these demographics are linearly associated to solving the different tasks, I inspected the variables that could be relevant in the multiple regression models for each task analysed. A multiple regression analysis was estimated, where the outcome of each of the three tasks was related to age, sex and years of studying[8]. When including these three variables the predictive power of the models was

[8] See Appendix 2. for more detailed data output for all statistical analyses performed in SAS JMP.

overall low. This general pattern was not surprising as it was expected that the unobserved ability of the participants would be the main source of the variability in the results.

There was no significant coefficients for *Candle problem* ($F(3,36)= 0.26$, $p= .85$) with an R^2 of .02. None of the included variables seem to predict the outcome significantly for *Figure test* ($F(3,36)= .28$, $p= .84$) with an R^2 of .02. Also for *Recall last-4 digits* there were no significant coefficients ($F(3,36)= 0.74$, $p= .54$) with an R^2 of .06.

These results indicate several important insights. The variables age, sex and years of studying were irrelevant for these overall models for each task. All participants were in the range 20-29 years of age and there was no systematic interaction between age and performance in this experiment. The fact that age did not provide any valuable explanation for the observed variation in scores for the three tasks, implies that, for the given age interval, adding another year probably does not improve the abilities required when solving these specific tasks. The fact that years of studying did not explain the variation in the performance for these three specific task, can be understood in a similar fashion, i.e. that additional years of studying at CBS for overall well educated students, does not interact in a systematic way with the performance on these specific tasks. There was a perfectly even ratio of 10 males and 10 females for the bonus treatment, and a close to even ratio for the no bonus group with 11 males and 9 females. There was no significant difference in task scores across all tasks for gender independent from the treatment and more importantly no significant difference in task scores dependent on the treatment[9]. This means that, overall, the males and females did not perform different dependent on the treatment. The most important finding was that the overall models only accounted for a small proportion of the variability. This was anticipated as the unobserved ability of the participants is expected to account for the main variability in scores.

To get an indication if subjects performed in a systematic way across tasks, I inspected the correlation between scores for the three tasks. The closer the correlation coefficient (*r)* is to 1 or -1 the higher the degree of relationship between scores for the tasks.

[9] Additionally, to the overall multiple regressions I also performed independent-two-samples t-test to compare task scores in the bonus condition across gender for all tasks. The effect of gender dependent on the treatment was not significant across all tasks dependent on treatment with the following presentation of the example of the effect of gender for *Recall last-4 digits* dependent on bonus. An independent-two-samples t-test was conducted to compare amount of correct trials in the bonus condition across gender. There was an insignificant difference dependent on bonus in scores for males (*M*=7.4, *SD*=3.6) and females (*M*=7.3, *SD*=2.2) conditions; t (18) = 0.08, p= .94.

Table 3. Pairwise Correlations and Significance Probabilities for the three tasks

Variable	By Variable	*Correlation*	*Count*	*Significance Probability*
Figure Test	*Candle Problem*	.1999	40	.2163
Recall last-4 digits	*Candle Problem*	-.2172	40	.1783
Recall last-4 digits	*Figure Test*	-.4181	40	.0073

The two tasks that were the "strongest" correlated were *Recall last-4 digits* and *Figure test*, $r(40) = -.42$. The Significance Probability reported in the last column of Table 3. tests a null hypothesis of no correlation (r=0). The Significance Probability provided that the correlation coefficient between scores for *Recall last-4 digits* and *Figure test* (p=.0073) is statistically significant.
The correlation coefficient was not statistically significant for the two other pairwise correlations. This implied that doing well on *Recall last-4 digits*, getting a high amount of correct trials, is related to a shorter solution time in the *Figure test* and vice versa. This indicates that there is a relationship between the performances for these two tasks, however it does not provide a systematic picture of the performance across all three tasks.

Now I will proceed with analysing the effect of the bonus on the performance for the three tasks. Before analysing this effect, I will provide a repetition of the definition of the performance level good and excellent for the three tasks, and the size of the corresponding reward.
There is several ways of analysing the performance dependent on the incentive i.e. by examining the task scores by the probability to reach at least the good or the excellent performance level or the fraction of earnings from the total possible earnings or simply using raw scores. I will start by presenting the probabilities for reaching the different performance levels, the fraction of earnings and then turn to the raw scores. This is done to show that the general pattern of the results is the same regardless of how the data is analysed. The raw scores are used for the further analyses in independent-two-samples t-tests dependent on incentive treatment.

If a participant in the incentivised group reach the performance level excellent, that participant receives DKK 200. If a participant in the incentivised group reach the performance level good, that participants receives DKK 100. For the *Candle problem* the performance level excellent is reached if the participant solves the task within 179 seconds and the performance level good is reached within 239 seconds. For the *Figure test*, the performance level excellent is reached if the participant solves the task within 178 seconds and the performance level good is reached within 238 seconds. For *Recall last-4 digits* a participant has 14 trials. The performance level good is defined as 10 or 11

correct trials, and the level excellent is reached if the participant has 12 or more correct trials. In Table 4. is a presentation of the frequencies of reaching the different performance levels and the actual earnings of the bonus and the hypothetical earnings of the no bonus group.

Table 4. Performance by task and treatment presented as Amount of subjects reaching excellent, and at least good performance levels and earnings in DKK

	N= Excellent		N= at least good		Earnings in DKK	
Treatment	No Bonus	Bonus	No Bonus	Bonus	No Bonus	Bonus
Task						
Candle problem	6	9	13	11	1900	2000
Figure test	8	4	10	7	1800	1100
Recall last-4 digits	7	2	11	4	1800	600

At first glance, there seems to be a positive effect of providing a bonus for reaching the excellent performance level when solving the *Candle problem* and a negative effect for the *Figure test* and *Recall last-4 digits*. The amount of people, from 20, who at least reached the good level, is higher for the no bonus treatment across all three tasks. The actual earnings DKK 3,700 of the bonus group is the sum of the last column compared to a hypothetical earning of the no bonus group of DKK 5,500. This general pattern is presented in Table 5. where the frequencies are presented as percentages. The last column, the % earnings provides a picture of very a small positive difference in performance for the *Candle problem* and relatively larger negative effects for the *Figure test* and *Recall last-4 digits*.

Table 5. Performance by task and treatment presented as Percent of subjects reaching Excellent, and at least good performance levels and earnings in DKK

	% Excellent		% at least Good		% earnings	
Treatment	No Bonus	Bonus	No Bonus	Bonus	No Bonus	Bonus
Task						
Candle problem	30	45	65	55	48	50
Figure test	40	20	50	35	45	28
Recall last-4 digits	35	10	55	20	45	15

Table 4. and Table 5. provide a first impression of how the different groups performed across the three tasks, but one should keep in mind that the contingencies and the resulting earnings are hypothetical for the no bonus group. When you relate it directly to the compensation for instance by examining the data by the probability to reach the at least good or excellent performance level or the fraction of earnings from the total possible earnings, you compare the essential difference between the two groups since the first group did not face these contingencies. The group that was

tested first, the no bonus group, was not informed of any such performance levels making this approach seem far from ideal for actual comparisons. As an alternative, the data is analysed in terms of raw scores from the two groups, even though it does not relate directly to the compensation that the incentivised group received. Both groups were informed and it was emphasised from the experimenter that they should solve all the tasks to their best ability. Obviously, the incentivised group did receive additional information regarding how their performance translates into bonuses. However, they were not able to see the clock and would therefore be expected to solve the first two tasks as quickly as possible, instead of making inferences about their times relative to incentive levels. For *Recall last-4 digits*, the potential downside from using raw scores is that rewards can become unreachable because of too many incorrect trials, which might be more easily inferred by participants during the course of 14 trials compared to making inferences from a sense of time in the first two tasks. So for all three tasks there might be a potential negative effect from missing out on a reward whilst performing the task, however as mentioned above it was stressed from the experimenter that they should strive to solve every task to their best ability. For this reason the data will be analysed in terms of mean raw scores, which are presented in Table 6.

Table 6. Performance by task and treatment presented as Mean raw scores and standard deviations.

	Means		Standard deviation (Sd)	
Treatment	No Bonus	Bonus	No Bonus	Bonus
Task				
Candle problem	239,2	220,5	127,0	141,7
Figure test	237,9	319,4	148,3	168,4
Recall last-4 digits	9,7	7,4	3,1	2,9

Table 6. indicates a positive effect of providing a bonus when solving the *Candle problem* and a negative effect for the *Figure test* and *Recall last-4 digits*, on average, with relatively similar standard deviations within groups. Before performing independent two sided t-tests to compare task-solving scores for each of the three tasks in a no bonus and bonus condition, I tested the assumption of the normality and equality of the variance- for the sub-groups. The assumptions regarding randomisation is met as the overall sample was randomly selected the two groups can be treated as independent random samples.

The normality assumption of the distributions was tested, for the *Candle problem*, in a Shapiro-Wilks test of the no bonus (W=0.89, p= 0.03) and the bonus (W=0.92, p= 0.14) groups. It indicated

a rejection of null hypothesis that the data for the no bonus group follow a normal distribution whereas the null hypothesis was not rejected for the bonus group. When testing for the equality of variances, The Levene's test indicated that there is not a statistically significant difference (F=1.65, p= 0.21) in variances between the two groups when solving the *Candle problem.* Consequently, a pooled estimate for the error term is used. Next, an independent-two-samples t-test was conducted to compare the time needed to solve the *Candle problem* in bonus and no bonus conditions. There was an insignificant difference in scores for bonus (*M*=220, *SD*=142) and no bonus (*M*=239, *SD*=127) conditions; t (38) = -0.44, p= 0.66. These results indicate that there is not a statistically significant effect of the bonus on the time needed to solve the *Candle problem* as there was no basis to reject the null hypothesis of equal means. Additionally, I performed, a non-parametric test due to a possible violation of the normality distribution, Wilcoxon rank-sums test (Z=-.05, p = 0.62) that tests that the mean ranks of the groups are the same. This test confirmed a non-significant effect of the bonus on the time needed to solve the *Candle problem*. The p-values are relatively similar for both tests, which was expected, as the t-test is robust against violations of the normality assumption. The rewarded group was 19 seconds quicker on average, however this difference was not statistically significant.

For the *Figure test* the normality assumption was tested of the distributions in a Shapiro-Wilks test of the no bonus (W=0.92, p= 0.08) and the bonus (W=0.94, p= 0.26) groups. The results indicate that the data for both groups follow a normal distribution. When testing for the equality of variances, The Levene's test indicated that there is not a statistically significant difference (F=0.43, p=0.52) in variances between the two groups when solving the *Figure test*. Consequently, a pooled estimate for the error term is used. Next, an independent-two-samples t-test was conducted to compare the time needed to solve the *Figure test* in bonus and no bonus conditions. There was an insignificant difference in scores for bonus (*M*=319, *SD*=168) and no bonus (*M*=238, *SD*=148) conditions; t (38) = 1.62, p=0.11. These results indicate that there is no statistically significant effect of the bonus on the time needed to solve the *Figure test.* The rewarded group was 82 seconds slower on average, however this difference was not statistically significant.

The normality assumption was tested, for *Recall last-4 digits*, of the distributions in a Shapiro-Wilks test of the no bonus (W=0.94, p= 0.28) and the bonus (W=0.95, p= 0.35) groups. The results indicate that the data for both groups follow a normal distribution. When testing for the equality of variances, The Levene's test indicated that there is not statistically significant difference (F=0.17,

p=0.68) in variances between the two groups. An independent-two-samples t-test was conducted to compare the amount of correct trials in the *Recall last-4 digits* task in bonus and no bonus conditions. There was a significant difference in scores for bonus (*M*=7.4, *SD*=2.9) and no bonus (*M*=9.7, *SD*=3.1) conditions; t (38) = -2.45, p=0.02. These results indicate that there is an effect of the bonus on number of correct trials, as I could reject the null hypothesis of equal means. The rewarded group had a 2.3 fewer correct answers on average and this difference was statistically significant providing a negative effect of providing a reward. To test if this difference was driven by gender, I further tested whether the effect of the incentive was dependent on gender, using an independent-two-samples t-test. The normality and equality of variance assumptions was tested in the same manner described above, with no detection of violations. For women, I found a significant difference in scores for bonus (*M*=7.3, *SD*=0.9) and no bonus (*M*=10.2, *SD*=0.9) conditions; t (17) = -2.3, p=.035. For men, I found no significant difference in scores for bonus (*M*=7.4, *SD*=1.0) and no bonus (*M*=9.2, *SD*=1.0) conditions; t (19) = -1.2, p=.22. Implying that the number of correct trials in *Recall last-4 digits* for women decreased significantly when provided with a bonus. There was a difference for men, however this difference was not significant. This implies that the difference in the bonus and no bonus conditions, independent from sex, is in fact explained by the significant difference for women.

In order to check, if this was a general pattern across all three tasks, I performed a similar analysis of the *Candle problem* and the *Figure test*[10]. For the *Candle problem*, no such difference prevailed. For the *Figure test*, there was a significant difference in scores for men, while there was no significant effect for women. There was a significant decrease in performance in the bonus condition for men. However, the negative effect of the incentive for men was not large enough to make the effect of the incentive significant independent from gender. On an overall level, there was no significant difference in task scores for men and women dependent on the treatment. However, there was a significant difference across treatments in two instances dependent on gender. There was a significant negative effect from the bonus, for men performing the *Figure test* and women performing *Recall last-4 digits*. Implying that there was not a general pattern of the effect of the bonus dependent on gender.

[10] These additional tests are provided in Appendix 2. They are not a part of the main aim of the analysis, which is testing for the effect of the bonus for the three tasks independent from gender.

The findings showed on average a small positive effect from the bonus when solving the *Candle problem* and a larger negative effect for the *Figure test* compared to the control group however these differences did not provide a statistically significant indications that would provide basis for rejecting the null hypotheses of equal means. This answers the following hypothesis regarding a potential negative effect from incentivising when the tasks is of a more complex nature and provides some intrinsic task interest, which is a relation that there was not found any statistical evidence fore. There was a statistically significant difference in the performance levels between the two groups when performing the task *Recall last-4 digits*, which answers the second following hypothesis. However, the result was of the opposite effect that was anticipated, with the bonus providing a negative impact on performance for this simple task.

11. Discussion and implications

The literature review provided several possible explanations for the motivational effects at play in the experiment. These insights provide the basis for interpreting the observed performance in the experiment.

Motivation is a fundamental component of any model on performance. For many organisations, it was their number one problem (Watson, 1994). Kunz and Pfaff (2002) raise the point of critique that, intrinsic motivation is essentially a hypothetical constructed term. This claim is valid but it does not provide much insight. It can be asserted that untangling the isolated effect from intrinsic motivation can be difficult as: “intrinsic motivation rarely operates in isolation from other types of motivation” (Locke & Latham, 1990, 58). Admittedly, many activities are not intrinsically interesting (Locke & Latham, 1990) and that people who work, for the most part, have to earn money so using rewards as a central motivational strategy seems appealing and practical. However, ignoring the influence of intrinsic motivation on performance is a losing proposition as: “intrinsic motivation remains a moderate to strong predictor of performance regardless of whether incentives are present” (Cerasoli et al., 2014, 1001). The Porter- Lawler model assumed an additive effect of intrinsic and extrinsic rewards. On the contrary, The Cognitive Evaluation Theory is concerned with a possible undermining effect of intrinsic motivation from rewards, which is an effect that according to (Eisenberger & Cameron, 1996) is more limited and conditional than commonly supposed. The Crowding Theory suggested a systematic interaction between extrinsic and intrinsic motivation based on the interpersonal climate within which rewards are administered, with

controlling aspects crowding-out and informational aspects crowding in. Cerasoli et al. (2014) found support for the hypothesis that incentives and intrinsic motivation are not necessarily antagonistic in the most recent meta-analysis on the topic. Depending on the type of performance and the contingency of the incentive, incentives coexist with intrinsic motivation. Cerasoli et al. (2014) proposed the incentive contingency directly performance-salient vs. indirectly performance-salient to be a possible moderator for the interrelationship among intrinsic motivation, extrinsic motivation and performance. However, such an effect was not tested, as the incentives provided in the experiment were directly performance salient. Cerasoli et al. (2014) found that, when incentives were present *and were directly salient to performance*, intrinsic motivation was a poorer predictor of performance (Cerasoli et al., 2014, 996). This implies that the experiment was performed in the setting where there is a possible undermining effect as: "In a "crowding out" fashion, intrinsic motivation was less important to performance when incentives were directly tied to performance" (Cerasoli et al., 2014, 980). The Self-Determination Theory introduced a detailed distinction of the extrinsic and intrinsic motivation dichotomy by introducing a controlled-to-autonomous continuum with the prediction that an autonomously motivated employee will perform better at complex tasks, whilst controlled motivation yields the same results or slightly better results on simple tasks. Cerasoli et al. (2014) find that the joint impact of incentives and intrinsic motivation is critical to performance. Cerasoli et al. (2014) focus on the moderators for the interrelationship among intrinsic motivation, extrinsic motivation and performance with the most critical distinction being the performance type quality and quantity. The tasks that emphasise performance quality typically require a higher degree of complexity and engagement of more skill than quantity-type tasks, which can also be interesting, nevertheless tend to be lower in complexity (Cerasoli et al., 2014, 982). Intrinsic motivation was more influential for quality than extrinsic incentives and extrinsic incentives accounted for more of the difference in quantity performance criteria than did intrinsic motivation (Cerasoli et al., 2014, 982).

Based on these insights, the effect of providing incentives for the complex tasks was expected to be negative. The analysis of scores for the *Candle problem* did not provide a negative effect. The effect was positive but not significant. It can be noted that for the *Candle problem*, (Glucksberg, 1962) found a negative effect from incentivising. Ramm et al. (2013) replicated the experiment. When comparing the raw scores Ramm et al. (2013) found a negative effect from incentivising dependent of completion, however this effect was not significant. Ramm et al. (2013) inspected the relevance

of gender composition as Glucksberg (1962) only had male students participating. Ramm et al. (2013) found no significant difference dependent on gender. The current experiment provided a further support to this finding, as I found no significant differences for gender for the *Candle problem*. Nor Ramm et al. (2013) or the present experiment can indicate further evidence for a negative effect of incentives, when solving the *Candle problem,* that Glucksberg (1962) originally found and Pink (2010) has revitalised.

On average the *Figure test,* which was casted as a task of a more complex character, indicated a larger negative effect from the incentive. Nevertheless, the difference was not statistically significant. The negative effect from incentivising, is in line with the hypothesis for complex tasks. Yet, the magnitude was insignificant. The finding of non-significant differences is much in line with (Winter, 2013), who states that effect sizes in psychological research are typically small. This picture of non-significant effects did not hold for the task, *Recall last-4 digits,* that mainly require concentration and was casted as the simple task. For *Recall last-4 digits,* I observed, a significant negative effect from the incentive. This finding is contrary to the hypothesis of a positive effect from incentives for simple tasks. This finding is quite surprising as it contradicts not only the relative price effect but also the fact it was the task that would be expected to yield the lowest level of intrinsic interest suggesting a very limited crowding out effect. However, it is much in line with Ariely et al. (2009) when testing Recall last-3 digits for three incentive levels who found a: ”declining performance as a function of incentives”(12).

The relative price effect predicts a positive effect of the incentive on performance. However, previous research from Ariely et al. (2009), challenges the assumption that increases in motivation necessarily lead to improvements in performance:

> In eight of the nine tasks we examined across the three experiments, higher incentives led to worse performance. In fact, we were surprised by the robustness of this effect; we had expected some of the six tasks included in the first experiment to respond in a positive monotonic fashion to level of incentive. (20)

The only task that provided a positive relation was a key pressing task that: “requires only pure physical effort” (Ariely et al., 2009, 14). This is a quite radical finding, as most tasks require more than pure physical effort.

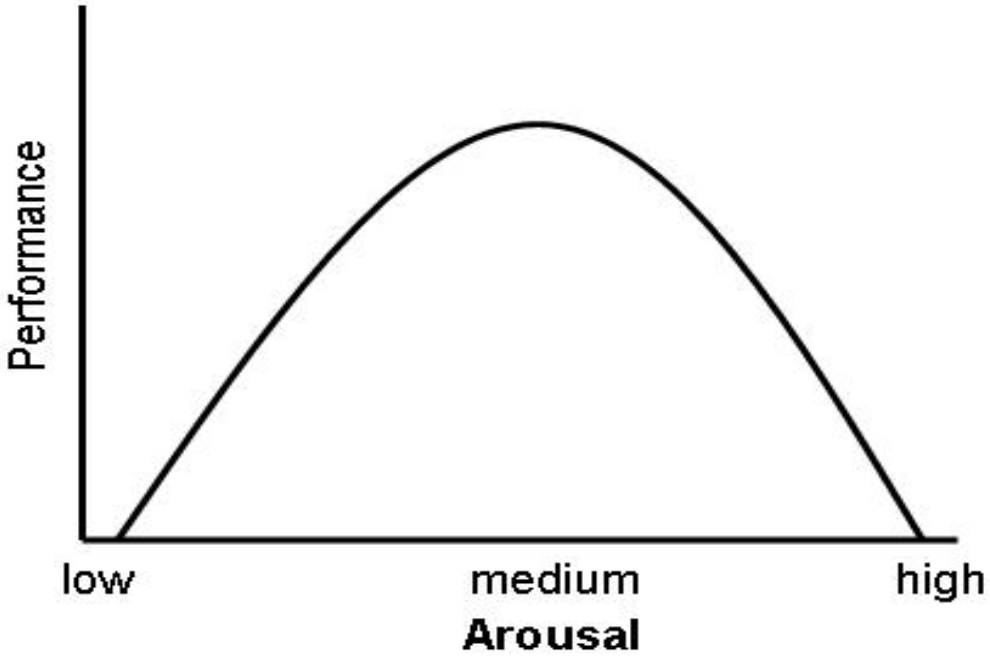
One of the nine tasks Ariely et al. (2009) performed was Recall last-3 digits. A task that is very similar to *Recall last-4 digits*. Ariely et al. (2009) found a negative effect from incentivising. This is

much in line with the results obtained from performing *Recall last-4 digits*. Prior to their experiment, Ariely et al. (2009) anticipated that the effect of large incentives would not affect the tasks that mainly involved concentration. Ariely et al. (2009) did not find such an effect. Ariely et al. (2009) explained their findings for the effect of incentives by refering to:

> The idea that excessive incentives could undermine task performance is embodied in the "Yerkes-Dodson law" [Yerkes and Dodson (1908)], which posits that there is an optimal level of arousal for executing tasks, and that departures from this level in either direction lead to a decrement in performance. (3)

In order to understand the relation between motivation and performance Ariely et al. (2009) refers to the Yerkes-Dodson law, depicted in Figure 3., as the main source of explanation for their results.

Figure 3. Depiction of the Yerkes-Dodson law.



The figure shows a relation where there is an optimal level of arousal for performing a task. Any deviation from this optimal level yields lower performance. This relation allows for individual differences in optimal levels and different optimal levels for different tasks. These differences would simply push the curve in either direction along the arousal axis, with the same general relation. A possible crowding out effect would dampen the overall level of arousal from introducing the incentive. The Porter- Lawler model assumed an additive effect of intrinsic and extrinsic rewards and further provided the assumption that increased effort does not automatically improve

performance. Because, individuals might not possess the necessary abilities needed to achieve high levels of performance. Previous research suggest that increased motivation can have a negative effect on performance due to a general focus of attention where increased motivation tends to narrow individuals' focus of attention (Easterbrook, 1959) . Attentional focus can further be detrimental for tasks that involve insight or creativity where participants need to make unusual connections between elements (McGraw & McCullers, 1979). These notions support the relation depicted in the Yerkes-Dodson law where too high levels of arousal get self-defeating. The Yerkes-Dodson law provides an interpretation of the results for the two incentive conditions in the current experiment. The task results for *Recall last-4 digits*, according to this perspective, provides a picture of a movement along the arousal axis. This movement, lead to a decrement in performance for *Recall last-4 digits* or as (Ariely et. al, 2009) put it: "It now appears that beyond some threshold level, raising incentives may increase motivation to supra-optimal levels and result in perverse effects on performance"(20). In this perspective, the non-significant results on the two complex tasks can be interpreted in three ways. Firstly, the reward was not perceived as significant, in spite of the intentions of the researcher. Implying that the bonus did not influence the motivation or the performance of the participants significantly. Secondly, that the introduction of the extrinsic reward only offset the drop in intrinsic motivation, for the complex tasks, on average. Leaving the overall motivation and performance unaltered. Thirdly, that the added motivation, that followed the incentive, dominated the crowding out effect, to such an extent, that it lead to an unaltered effect on performance. From one supra-optimal level to another supra-optimal level, on each side of the optimal peak level. Leaving the performance unaltered. The last explanation is supported by Ariely et al. (2009) who notes that: "Tasks that involve only effort are likely to benefit from increased incentives, while for tasks that include a cognitive component, there seems to be a level of incentive beyond which further increases can have detrimental effects on performance"(16).
There are several other potential reasons for why the experiment did not provide support, for the hypotheses of the performance in relation to the complexity of the task, mainly connected to the validity of the research. The fact that there was a large uncertainty connected to the "isolated" effect of the incentive, as there was not recorded a variable that explains or indicate the abilities of the individual subjects. This would have provided a better baseline for detecting potential effects of the incentive dependent on this indicator variable. The potential effect of the bonus might for this reason have "drowned in noise". This implies that there is a considerable potential of a type I error, which is the incorrect rejection of a true null hypothesis, for the negative effect of the incentive on

the complex tasks. The fact that, the simple task yielded the opposite result of the hypothesis is either a quite surprising finding or simply a result of the choices made by the researcher resulting in a type I error. It cannot be determined with certainty, that the researcher did not subconsciously change the pace when reading the digit sequences for the rewarded group. Perhaps, as a result of the rewards being paid out of the researchers own pocket, in spite of the fact that measures were made to avoid pace differences. This potentially harmful effect could have been avoided by recording the sequences. An alternative explanation for the significant difference in performance for *Recall last-4 digits* is that it is relatively easier to make inferences about the possibility for reaching the defined performance levels for this task. When participants realise that they are not able to meet the performance thresholds, their performance would suffer from the bonuses becoming unobtainable. This leads to lower scores on average for this task in the incentivised condition.

One could argue that the burden of proof is on providing a bonus, as the relative price effect would predict a positive relation to incentivising. Furthermore, providing a bonus, where there previously was none, come with additional costs to the principal. From this perspective the findings of a non-significant effect for the complex tasks and a significant negative effect on the simple task provides no argument for incentivising for these specific tasks. The opposite expectation, represented by Ariely et al. (2009), that incentives seriously harms the performance, for the tasks included in their experiment, is a far more radical finding than what the current thesis found.

Some limitations should be kept in mind when interpreting the results obtained in this experiment. The participants in the experiment were students of a business school and it cannot be determined if the results generalise to the general population. The experiment tested three specific tasks and the results are specific to these tasks, without being able to determine if the findings generalise to other tasks. There was no measurement of motivational levels implying that all inferences of motivational levels is based on the insights provided by other researchers. Further, it cannot be said with certainty that the results of the current experiment is not in fact particular and not general or as (Eisenberger & Cameron, 1996) point out, that group differences are actually random variations from a true difference that falls close to zero. Furthermore, it is possible that the results mainly reflects the choices of the researcher as proposed by Kunz and Pfaff (2002). However, ignoring previous research with references to its limitations might be a way of missing potentially important insights regarding the effect of providing incentives. The current paper does not provide an

argument for providing incentives for the three tasks that were tested. This result is supported by Ariely et al. (2009) who found that:

> For many tasks, introducing incentives where there previously were none or raising small incentives on the margin is likely to have a positive impact on performance. Our experiment suggests, however, that one cannot assume that introducing or raising incentives always improves performance. (20)

12. Conclusion

This thesis provides several indications for the effect of monetary incentives on three specific tasks in an experiment with students. The hypothesis was that providing incentives would have a positive effect for simple tasks and a negative effect for complex tasks. No such relation was found for the complex tasks, as the performance was not significantly different for the two conditions. One of these complex tasks was the *Candle problem*. For the *Candle problem*, there was no indication of the negative effect Glucksberg (1962) found, much in with Ramm et al. (2013) who did not find a significant detrimental effect of incentives for this specific task. The *Figure test*, the second complex task, did not provide any significant effect on performance from the incentive either. These findings might be considered surprising, given the relative price effect much of the economic theories are founded on, but does not provide support for an actual detrimental effect of incentives for the two complex tasks. However, in this experiment the simple task, *Recall last-4 digits*, provided a significantly negative effect of the incentive, which was a finding that was surprising, as it was the opposite effect of what was expected. Further, it was the task that, from a theoretical perspective, was the least likely to produce a negative effect of the incentive. It cannot be determined if this finding is in fact a result of the choices made by the researcher. However, it can be noted that Ariely et al. (2009) found the same relation, for a very similar task. These findings indicate that one cannot just assume a positive or negative effect on performance from providing incentives. The participants in the experiment were students and it cannot be determined if the results generalise to the general population. The experiment tested three specific tasks and the results are specific to these tasks, without being able to determine if the findings generalise to tasks in an actual work setting.

13. References

Bibliography

Agresti, A., & Franklin, C. (2009). *Statistics: The Art and Science of Learning from Data* (2nd ed.). New York: Pearson Education

Alchian, A. A., & Demsetz, H. (1972). Production, Information Costs, and Economic Organization. *The American Economic Review, 62*(5), 777-795.

Amabile, T. M., DeJong, W., & Lepper, M. R. (1976). Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology, 34*(1), 92-98.

Ariely, D., U. Gneezy, G. Loewenstein, & N. Mazar. (2003). *Large stakes and big mistakes*. Federal Reserve Bank of Boston Working Paper 05-11.

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

Aronson, E., Wilson T. D., & Brewer M. B. (1998). Experimentation in social psychology. In: Gilbert, D. T., Fiske, S. T., & Lindzey, G. (Eds.), *The handbook of social psychology*, 4th ed.,Vol. 1 pp. (99-142). Boston: McGraw-Hill.

Bandura, A. (1977). *Social Learning Theory.* Oxford: Prentice-Hall.

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46*(3), 610-620.

Beer, M. & Katz, N. (2003). Do Incentives Work? The Perception of A Worldwide Sample of Senior Executives. *Human Resource Planning, 26*(3): 30–44.

Behling, O., Labovitz, G., & Kosmo R. (1968). The Herzberg Controversy: A Critical Reappraisal. *The Academy of Management Journal, 11*(1), 99-108.

Bertrand, M., & Mullainathan, S. (2001). Are CEOS Rewarded for Luck? The Ones without Pincipals Are. *Quarterly Journal of Economics*, 116(3), 901-932.

Bregn, K. (2004). Betydningen af fairness, hensigt og tab: Implikationer for løn som ledelsesredskab i den offentlige sektor. *Samfundsøkonomen, 6.*

Besanko, D., Dranove, D., Shanley M., & Schaefer S. (2007). *Economics of Strategy*. New York: John Willey & Sons.

Camerer, C. F., & Malmendier, U. (2007). Behavioral Economics of Organizations. In: Diamond, P., & Vartiainen, H. (eds.), *Behavioral Economics and its Application*, pp. (235-290). New Jersey: Princeton University Press.

Cameron, J., & Pierce W. D. (1994). Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Review of Educational Research, 64*(3), 363-423.

Cameron, W. B. (1963). *Informal Sociology: A Casual Introduction to Sociological Thinking.* New York: Random House.

Cerasoli, C. P., Nicklin, J. M., & Ford M. T. (2014). Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis. *Psychological Bulletin, 140*(4), 980-1008.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Los Angeles: Sage Publications.

DeCharms, R. (1968). *Personal Causation: The internal affective determinants of behavior*. New York: Academic Press.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105-115.

Deci, E. L. (1975). *Intrinsic Motivation*. New York: Plenum Publishing.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin, 125*(6), 627-668.

Deci, E. L., & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. *Advances in experimental social psychology, 13*, 39-80.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum Publishing.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.

DiNapoli, T. P. (2015, March 11th). *Wall Street Bonuses Edge Up in 2014.* Retrieved on October 7th 2016 from: http://www.osc.state.ny.us/press/releases/mar15/031115.htm

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*(3), 183-201.

Edmans, A., & Gabaix, X. (2009). Is CEO pay really inefficient? A survey of new optimal contracting theories. *European Financial Management, 15*(3), 486-496

Eisenberger, R., & Cameron, J. (1996). The detrimental effects of reward: Myth or reality?, *American Psychologist, 51*(11), 1153-1166.

Fama, E. F., & Jensen, M. C. (1983). Separation of Ownership and Control. *Journal of Law and Economics*, *26*(2), 301-325.

Field, A., & Hole, G. (2003). *How to design and report experiments*. London: Sage Publications.

Frey, B. S. (1992). *Economics as a science of human behavior: towards a new social science paradigm*. Boston: Kluwer Academic Publishers.

Frey, B. S. (1997). *Not just for the money: an economic theory of personal motivation*. Cheltenham: Edward Elgar Publishing.

Frey, B. S., & Jegen, R. (2000). *Motivation Crowding Theory: A Survey of Empirical Evidence*. CESifo Working paper series no 245.

Gagné, M., & Deci, E. L. (2005). Self-Determination Theory and Work Motivation. *Journal of Organizational Behavior*, *26*(4), 331-362.

Glucksberg, S. (1962). The influence of strength of drive on functional fixedness and perceptual recognition. *Journal of Experimental Psychology*, *63*, 36–41.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, *8*(4), 597-607.

Guba, E. G. (1990). The Alternative Paradigm Dialog. In: Guba, E. G. (Ed.), *The Paradigm Dialog*, pp. (17-30). Newbury Park: Sage Publications.

Hawkins, F. H. (1993). *Human Factors in Flight.* Charlottesville: University of Virginia Press.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New Jersey: Lawrence Erlbaum Associates Inc.

Herzberg, F., Mausner, B., & Snyderman, B. B. (1959). *The Motivation to work*. New York: John Wiley & Sons.

Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

Holmberg, S., & Schmitt, M. (2014, Fall). The Overpaid CEO. *Democracy Journal*, *34*, 60. Retrieved on September 12th from: http://democracyjournal.org/magazine/34 /the-overpaid-ceo/

Holmström, B., & Milgrom, P. (1994). The Firm as an Incentive System. *American Economic Review*, *84*(4), 972-991.

Kelleher, J. B. (2013, November 1st). *90 Percent of Employers Tie Workers' Pay to Company Performance*. Retrieved on October 3rd 2016 from: http://www.huffingtonpost.com/2013/09/01 /workers-pay-company-performance_n_3851972.html

Kenny, C. (2014, April 28th). *Why Factory Jobs Are Shrinking Everywhere*. Retrieved on September 13th 2016 from: http://www.bloomberg.com/news/articles/2014-04-28/why-factory-jobs-are-shrinking-everywhere

Kerr, S. (1975). On the Folly of Rewarding A, While Hoping for B. *The Academy of Management Journal, 18*(4), 769-783.

Kirk, J. & Miller, M. L. (1986). *Reliability and validity in qualitative research.* Beverley Hills: Sage Publications.

Koestner, R., Ryan, R. M., Bernieri, F., & Holt, K. (1984). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using Cognitive Evaluation Theory. *Journal of Personality and Social Psychology, 52*(3), 233-248.

Kohn, A. (1993a). *Punished by rewards*. Boston: Houghton-Mifflin.

Kohn, A. (1993b). Why Incentive Plans Cannot Work. *Harvard Business Review, 71*(September-October), 54-63.

Kohn, A. (1996). By All Available Means: Cameron and Pierce's Defense of Extrinsic Motivators. *Review of Educational Research, 66*(1), 1-4.

Kunz, A. H., & Pfaff, D. (2002). Agency theory, performance evaluation, and the hypothetical construct of intrinsic motivation. *Accounting, Organizations and Society, 27*(3), 275-295.

Langer, E. J., & Imber, L. M. (1979). When practice makes imperfect: Debilitating effect of overlearning. *Journal of Personality and Social Psychology, 37*(11), 2014-2024.

Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review, 90*(5), 1346-1361.

Lepper, M. R., & Greene, D. (1975). Turning play into work: effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology, 31*(3), 479-486.

Lepper, M. R., & Greene, D. (1978). *The Hidden Cost of Reward: New Perspectives on Psychology of Human Motivation.* New York: Erlbaum.

Lepper, M. R., Greene, D. & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28*(1), 129-137.

Lepper, M. R., Keavney, M. & Drake, M. (1996). Intrinsic motivation and extrinsic rewards: A commentary on Cameron and Pierce's metanalysis. *Review of Educational Research, 66(1)*, 5-32.

Locke, E. A., & Latham, G. P. (1990). Core findings. In: Locke, E. A., & Latham, G. P. (Eds.), *A Theory of goal setting and task performance*, pp. (27-62). New Jersey: Prentice Hall

Maslow, A. (1943). A theory of human motivation. *Psychological Review, 50*, 370–396.

McGraw, K. O., & McCullers, J. C. (1979). Evidence of a Detrimental Effect of Extrinsic Incentives on Breaking a Mental set. *Journal of Experimental Social Psychology, 15*, 285-294.

McSpadden, K. (2015, March 11th). *The Average bonus in Wall Street was $172,860 in 2014.* Retrieved on October 7th 2016 from: http://time.com/3741667/wall-street-bonuses-wages/

Mook, D. G. (1983). In Defence of External Validity. *American Psychologist, 38*(4), 379-387.

Mossholder, K. W. (1980). Effects of externally mediated goal setting on intrinsic motivation: A laboratory experiment. *Journal of Applied Psychology, 65*(2), 202-210.

Myers, D. G. (1996). *Social Psychology*. New York: McGraw-Hill.

Oettingen, G., & Gollwitzer, P. M. (2001). Goal setting and goal striving. In: Tesser, A., & Schwarz, N. (Eds.), *Blackwell handbook of social psychology: Intraindividual processes,* Vol.1, pp. (329-347). Oxford: Blackwell.

Overskeid, G. (2002). Psychological hedonism and the nature of motivation: Bertrand Russell's anhedonic desires. *Philosophical Psychology, 15*, 77-93

Pink, D. H. (2010). *Drive: The surprising truth about what motivates us*. Edinburgh: Canongate

Porter, L. W., & Lawler, E. E. (1968). *Managerial attitudes and performance*, Homewood: Irwin-Dorsey Press.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature, 37*(1), 7-63.

Raith, M. (2008). Specific Knowledge and Performance Measurement. *The RAND Journal of Economics*, *39*(4), 1059-1079.

Ramm, J., Tjøtta, S., & Torsvik, G. (2013). *Incentives and creativity in groups*. CESifo Working paper series no 4374.

Rebitzer, J. B., & Taylor, L. J. (2010). *Extrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches to Agency and Labor Markets.* IZA Discussion Paper No. 5058

Rummel, A., & Feinberg, R. (1988). Cognitive evaluation theory: A meta-analytic review of the literature. *Social Behavior and Personality,16*(2), 147-164.

Ryan, R. M. (1995). Psychological needs and facilitation of integrative process. *Journal of Personality*, *63*(3), 397-427.

Ryan, R. M., & Deci, E. L. (1996). When paradigms clash: Comments on Cameron and Pierce's claim that rewards do not undermine intrinsic motivation. *Review of Education Research*, *66*(1), 33-38.

Sansone, C. & Harackiewicz, J. M. (2000). *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*. San Diego: Academic Press.

Skinner, B. F. (1984). Selection by consequences. *Behavioral and Brain Sciences, 7*(4), 477-481.

Sloan, A. (2016, February 14th). *Beware big talk about the tax code*. Retrieved on September 6th 2016 from: https://www.washingtonpost.com/business/beware-big-talk-about-the-tax-code /2016/02/11/6ccab0fe-cf35-11e5-88cd-753e80cd29ad_story.html.

Smith, W. E. (1975). *The effect of anticipated vs. unanticipated social reward and subsequent intrinsic motivation*. New York: Cornell University. Dissertation.

Tang, S.-H., & Hall, V. C. (1995). The overjustification effect: A meta-analysis. *Applied Cognitive Psychology*, *9*(5), 365-404.

Taylor, F. W. (1947). *Principles of Scientific Management.* New York: Harper.

The Georgia Straight. (2013, October 4th). *Robert Reich explains how globalization, technology and natural resources are reshaping global economy*. Retrieved on September 9th 2016 from: http://www.straight.com/news/498931/robert-reich-explains-how-globalization-technology-and-natural-resources-are-reshaping-global-economy.

Turén, T. (2008). *Videnskabsteori for begyndere.* København: Rosinante Forlag a/s.

University of California Television. (2015). *The Great Divide with Joseph Stiglitz and Robert Reich* [Video]. Berkeley: UCTV.

Vroom, V. H. (1964). *Work and Motivation*. New York: Wiley

Watson, T. (1994). Linking employee motivation and satisfaction to the bottom line. *CMA Magazine*, *68*(3), 4.

Weitzman, M. L. (1980). The "Ratchet Principle" and Performance Incentives. *The bell journal of Economics*, *11*(1), 302-308.

Wiersma, U. J. (1992). The effect of extrinsic rewards in intrinsic motivation: A meta-analysis. *Journal of Occupational and Organizational Psychology*, *65*(2), 101-114.

Winter, J. C. F. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assement, Research & Evaluation, 18*(10), 1-12

Zuckerman, M., Porac, J., Lathin, D., Smith, R., & Deci, E. L. (1978). On the importance of self-determination for intrinsically motivated behavior. *Personality and Social Psychology Bulletin*, *4*(3), 443-446.

14. Appendices

Appendix 1. Task presentations

The tasks on the following pages, as they were presented to participants on each there sheet of A4 and available to all were pencils and empty squared paper when performing the *Figure test*. The digit sequences for *Recall last-4 digits* was presented verbally.

The *Candle problem*

“Fix a lit candle to the wall (Bulletin board) in a way so it will not drip onto the table below it”.

Tools:

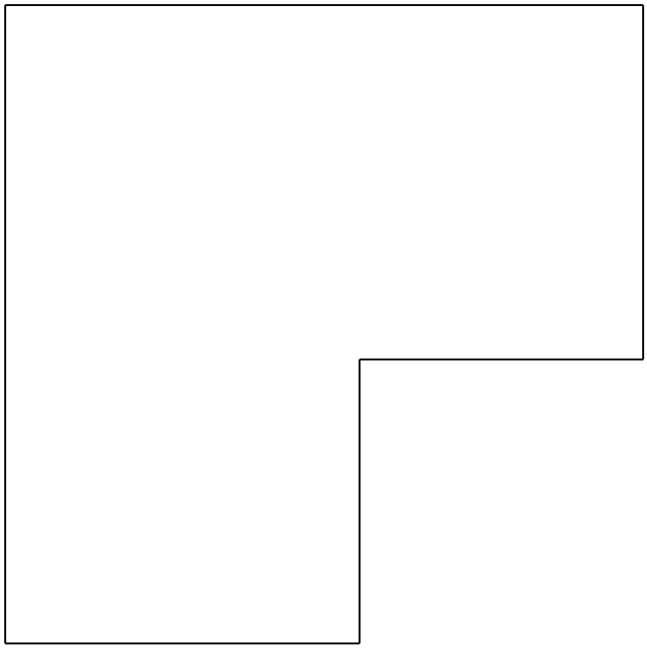
A small box with *thumbtacks*

A box of *matches*

A *candle*

Figure test

“Show how you can divide this figure into four equal parts that has the same size and shape, and has the shape of the figure”.



Recall last-4 digits

“I will read a sequence of digits in a quick pace and stop at an unannounced point, and ask you to provide the last four digits before the stop”.

Sequence of digits:

542310239490340349596**3486**

774493020459**4098** 3894798734874093**4873**

49838740983094834**3472** 7191984754**7492**

3712717304765**8593**

870383398274**8236**

467593274847**0932**

77289**3892**

8398216366**4738**

9292002846**8902**

2624434536112232**4536**

474937255**0209**

8355257282389**2385**

Appendix 2. Data output from SAS JMP – in the order presented in the thesis

Overall model for *Candle problem.*

Response Candle Problem

Effect Summary

Source	LogWorth	PValue
Sex	0,386	0,41069
Years of studying	0,198	0,63361
age	0,148	0,71085

Remove Add Edit ☐ FDR

Lack Of Fit

Summary of Fit

RSquare	0,021157
RSquare Adj	-0,06041
Root Mean Square Error	137,1356
Mean of Response	229,825
Observations (or Sum Wgts)	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	14633,12	4877,7	0,2594
Error	36	677022,65	18806,2	Prob > F
C. Total	39	691655,78		0,8542

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	145,44405	274,489	0,53	0,5995
Sex[F]	18,688513	22,45245	0,83	0,4107
age	5,320184	14,23816	0,37	0,7109
Years of studying	-11,63753	24,20782	-0,48	0,6336

Overall model for *Figure test.*

Response Figure Test

Effect Summary

Source	LogWorth	PValue
age	0,307	0,49365
Years of studying	0,052	0,88635
Sex	0,026	0,94278

Remove Add Edit ☐ FDR

Lack Of Fit

Summary of Fit

RSquare	0,023164
RSquare Adj	-0,05824
Root Mean Square Error	166,6064
Mean of Response	278,625
Observations (or Sum Wgts)	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	23696,2	7898,7	0,2846
Error	36	999277,2	27757,7	Prob > F
C. Total	39	1022973,4		0,8362

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15,055665	333,4774	0,05	0,9642
Sex[F]	1,971471	27,27754	0,07	0,9428
age	11,962611	17,29798	0,69	0,4936
Years of studying	-4,233487	29,41014	-0,14	0,8863

Overall model for *Recall last-4 digits.*

Response Recall last four digits

Effect Summary

Source	LogWorth	PValue
age	0,632	0,23337
Years of studying	0,174	0,66951
Sex	0,045	0,90094

Remove Add Edit FDR

Lack Of Fit

Summary of Fit

RSquare	0,058042
RSquare Adj	-0,02045
Root Mean Square Error	3,186254
Mean of Response	8,5
Observations (or Sum Wgts)	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	22,52027	7,5068	0,7394
Error	36	365,47973	10,1522	Prob > F
C. Total	39	388,00000		0,5355

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17,013254	6,377567	2,67	0,0114*
Sex[F]	0,0653939	0,521668	0,13	0,9009
age	-0,400978	0,330814	-1,21	0,2334
Years of studying	0,2420453	0,562452	0,43	0,6695

Pairwise correlations for scores for the three tasks.

Multivariate

Correlations

	Candle Problem	Figure Test	Recall last four digits
Candle Problem	1,0000	0,1999	-0,2172
Figure Test	0,1999	1,0000	-0,4181
Recall last four digits	-0,2172	-0,4181	1,0000

Correlation Probability

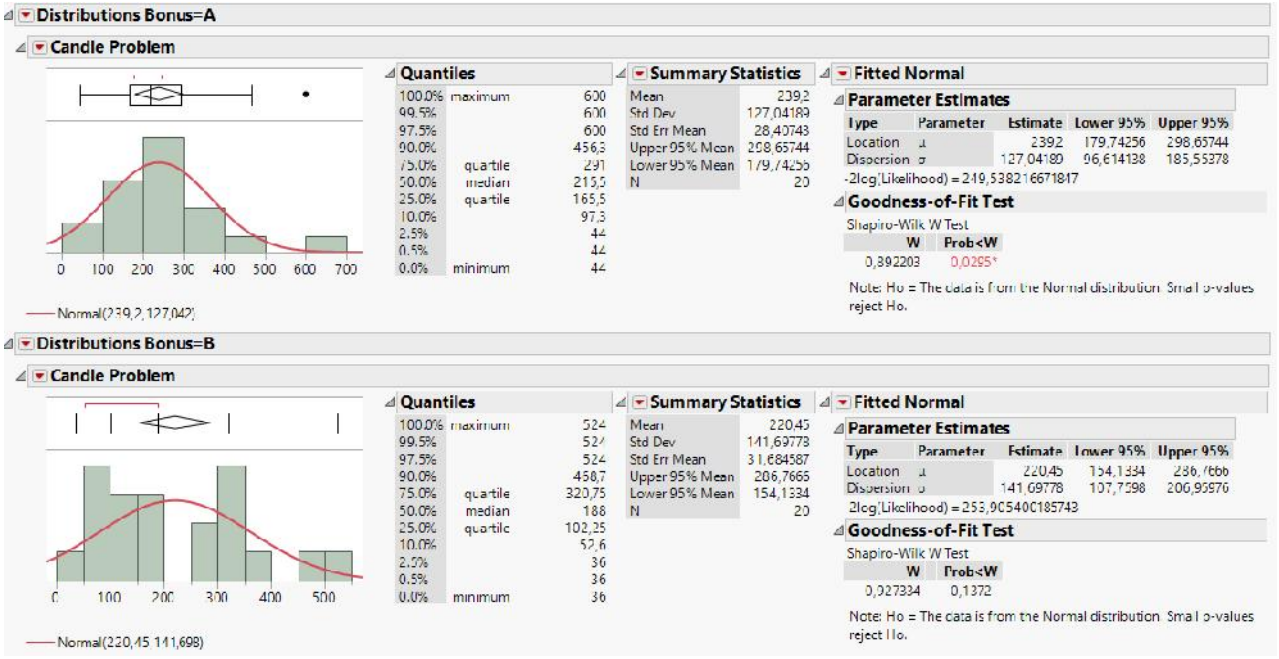
	Candle Problem	Figure Test	Recall last four digits
Candle Problem	<,0001	0,2163	0,1783
Figure Test	0,2163	<,0001	0,0073
Recall last four digits	0,1783	0,0073	<,0001

Scatterplot Matrix

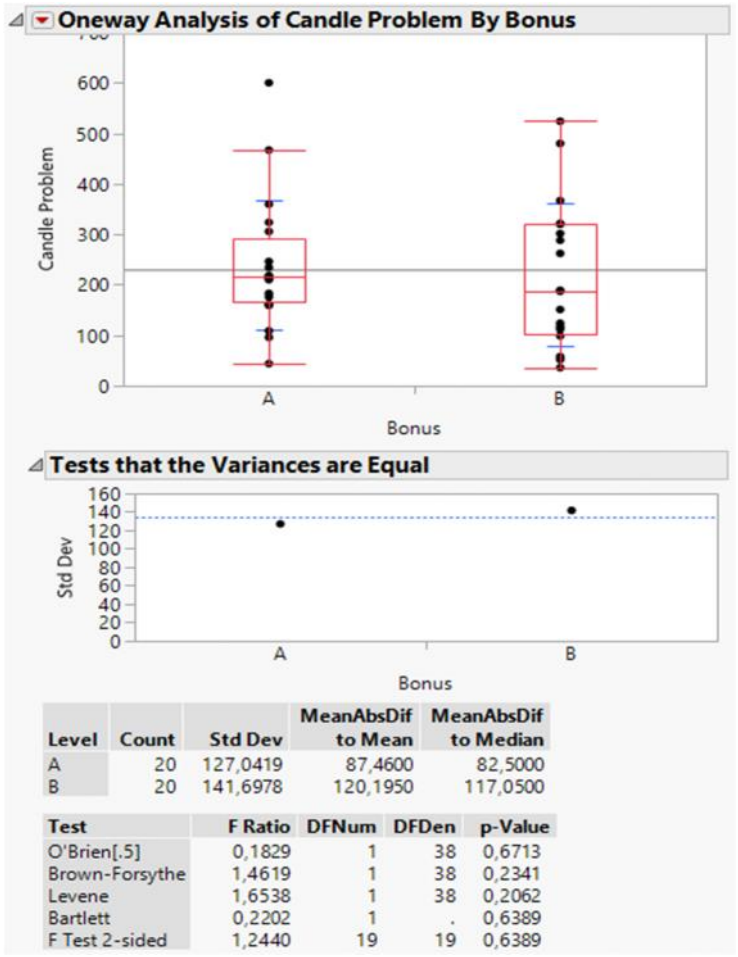
Pairwise Correlations

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	-,8 -,6 -,4 -,2 0 ,2 ,4 ,6 ,8
Figure Test	Candle Problem	0,1999	40	-0,1191	0,4814	0,2163	
Recall last four digits	Candle Problem	-0,2172	40	-0,4952	0,1012	0,1783	
Recall last four digits	Figure Test	-0,4181	40	-0,6455	-0,1225	0,0073*	

Test of normality assumption for *Candle problem.*



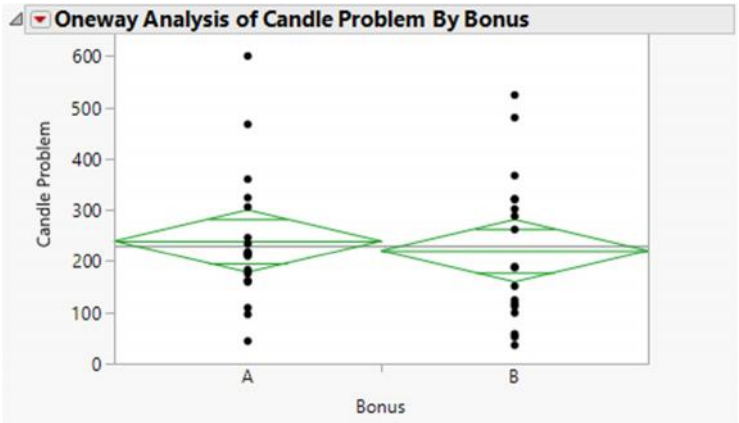
Test of equality of variance assumption for *Candle problem.*



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
A	20	127,0419	87,4600	82,5000
B	20	141,6978	120,1950	117,0500

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	0,1829	1	38	0,6713
Brown-Forsythe	1,4619	1	38	0,2341
Levene	1,6538	1	38	0,2062
Bartlett	0,2202	1	.	0,6389
F Test 2-sided	1,2440	19	19	0,6389

Independent-two-samples t-test for *Candle problem* by bonus.



Oneway Anova

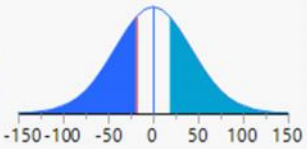
Summary of Fit

t Test

B-A

Assuming equal variances

Difference	-18,75	t Ratio	-0,44061
Std Err Dif	42,55	DF	38
Upper CL Dif	67,40	Prob > |t|	0,6620
Lower CL Dif	-104,90	Prob > t	0,6690
Confidence	0,95	Prob < t	0,3310



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	3515,63	3515,6	0,1941	0,6620
Error	38	688140,15	18109,0		
C. Total	39	691655,78			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	20	239,200	30,091	178,28	300,12
B	20	220,450	30,091	159,53	281,37

Std Error uses a pooled estimate of error variance

Wilcoxon rank sums test for *Candle Problem.*

Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
A	20	429,000	410,000	21,4500	0,500
B	20	391,000	410,000	19,5500	-0,500

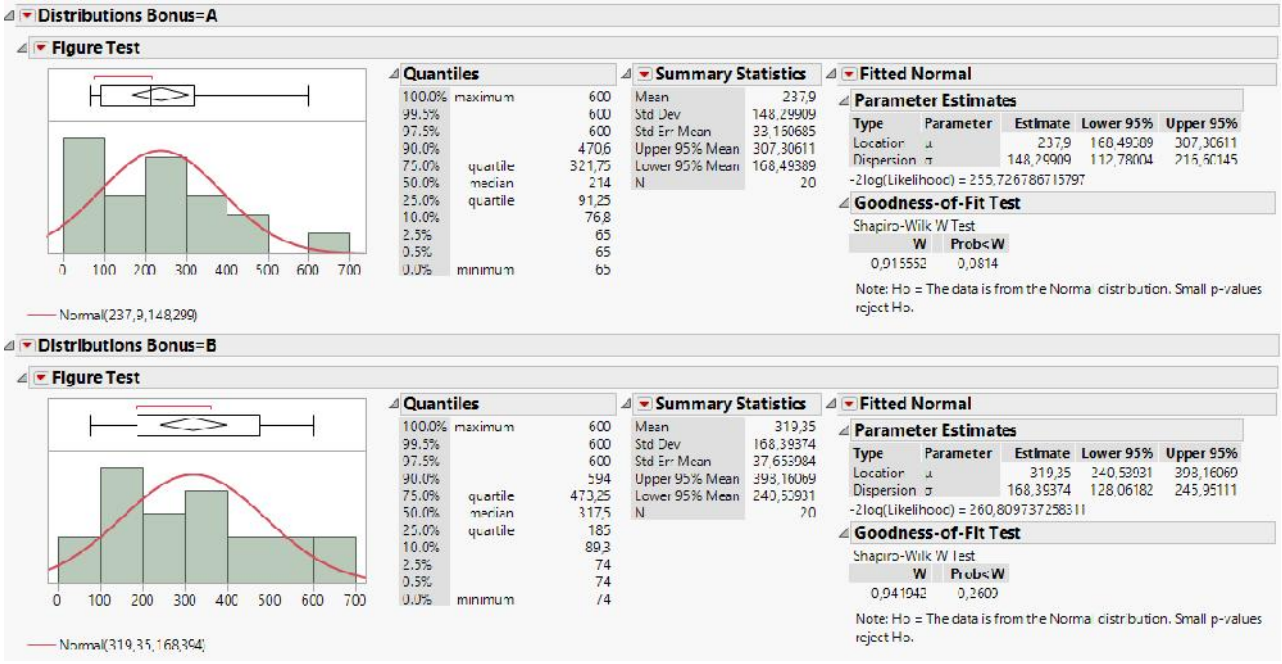
2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
391	-0,50045	0,6168

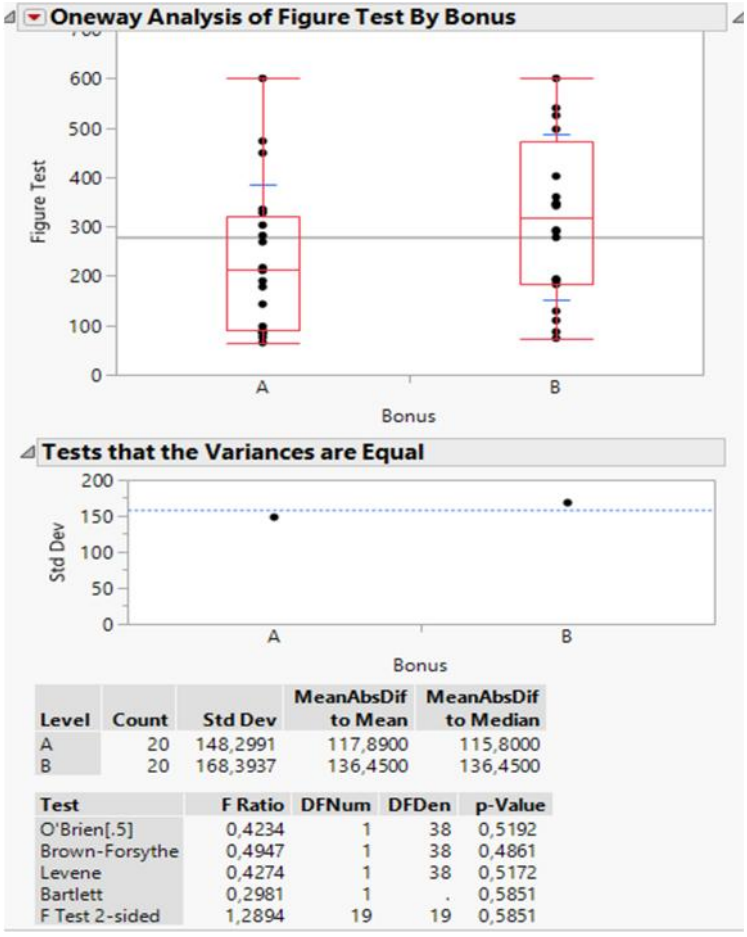
1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
0,2642	1	0,6073

Test of normality assumption for *Figure test.*



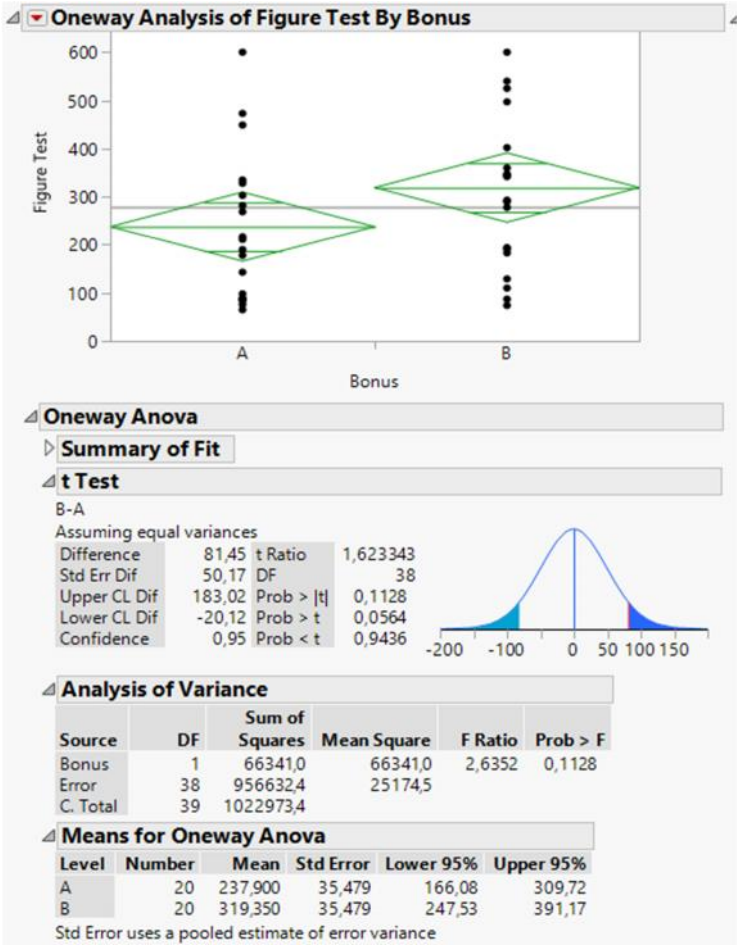
Test of equality of variance assumption for *Figure test.*



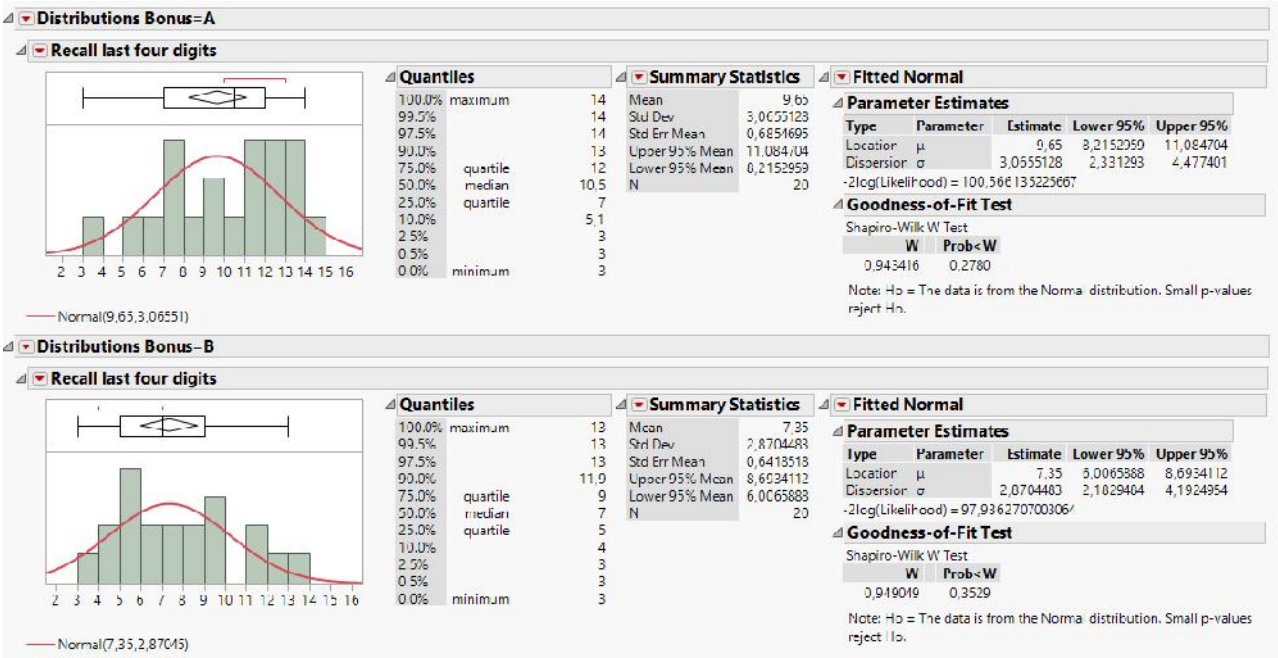
Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
A	20	148,2991	117,8900	115,8000
B	20	168,3937	136,4500	136,4500

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	0,4234	1	38	0,5192
Brown-Forsythe	0,4947	1	38	0,4861
Levene	0,4274	1	38	0,5172
Bartlett	0,2981	1	.	0,5851
F Test 2-sided	1,2894	19	19	0,5851

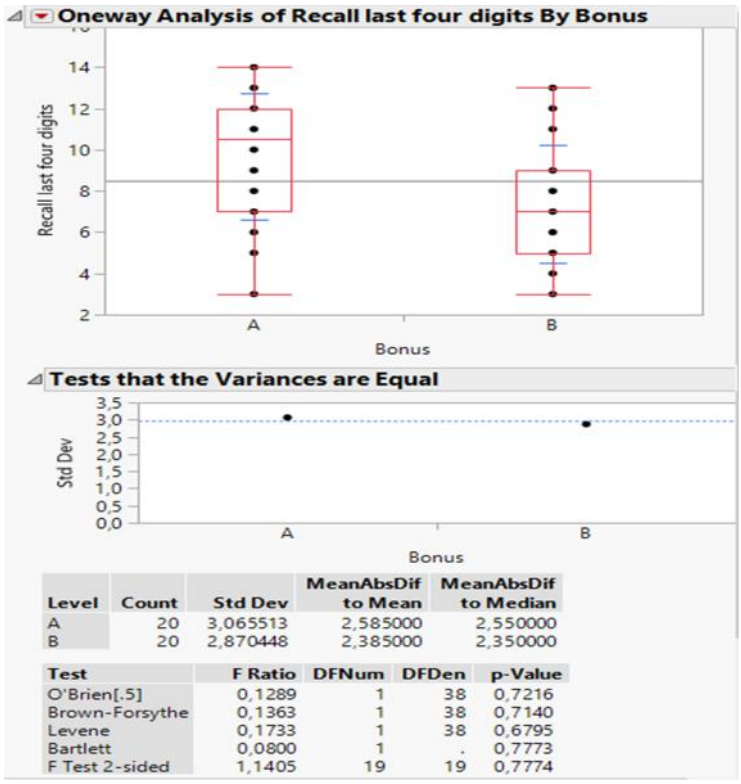
Independent-two-samples t-test for *Figure test* by bonus.



Test of normality assumption for *Recall last-4 digits.*



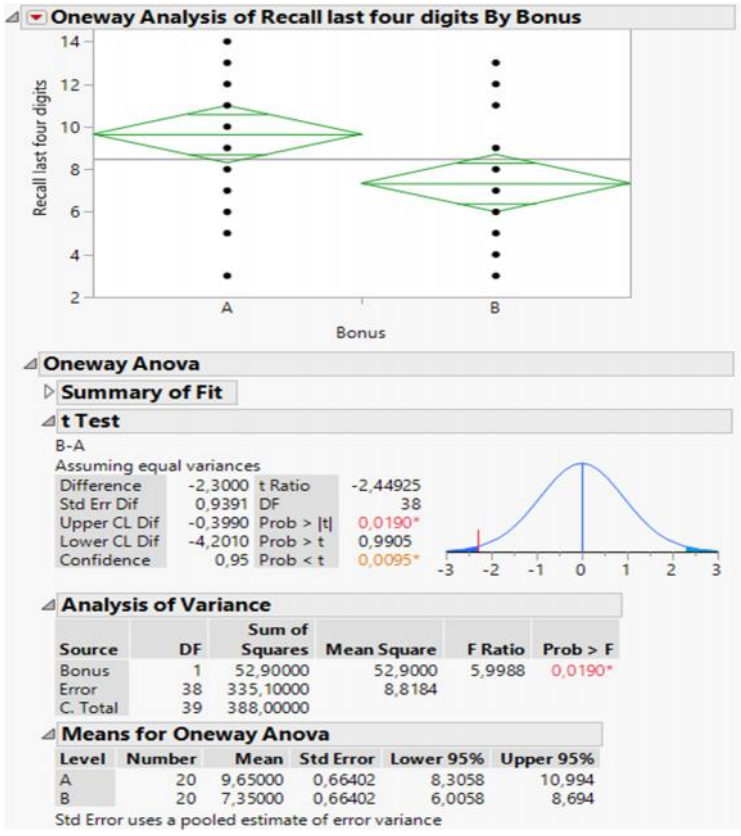
Test of equality of variance assumption for *Recall last-4 digits.*



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
A	20	3,065513	2,585000	2,550000
B	20	2,870448	2,385000	2,350000

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	0,1289	1	38	0,7216
Brown-Forsythe	0,1363	1	38	0,7140
Levene	0,1733	1	38	0,6795
Bartlett	0,0800	1	.	0,7773
F Test 2-sided	1,1405	19	19	0,7774

Independent-two-samples t-test for *Recall last-4 digits* by bonus.



Difference	-2,3000	t Ratio	-2,44925
Std Err Dif	0,9391	DF	38
Upper CL Dif	-0,3990	Prob > |t|	0,0190*
Lower CL Dif	-4,2010	Prob > t	0,9905
Confidence	0,95	Prob < t	0,0095*

Analysis of Variance

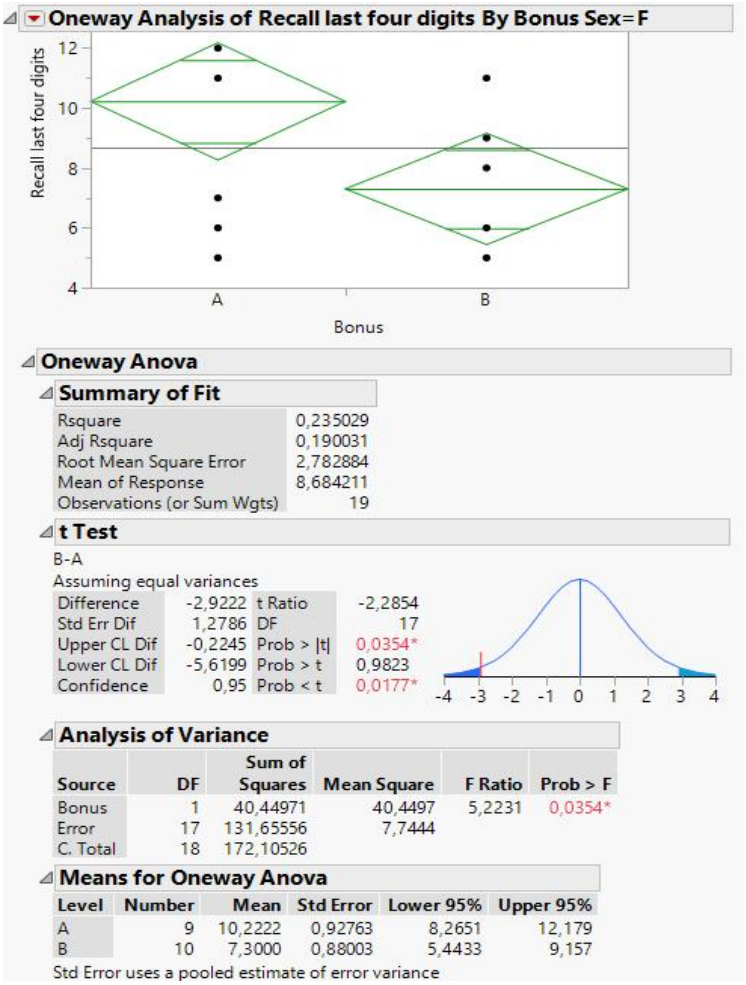
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	52,90000	52,9000	5,9988	0,0190*
Error	38	335,10000	8,8184		
C. Total	39	388,00000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	20	9,65000	0,66402	8,3058	10,994
B	20	7,35000	0,66402	6,0058	8,694

Std Error uses a pooled estimate of error variance

Independent-two-samples t-test for *Recall last-4 digits* by bonus for women.



Oneway Anova

Summary of Fit

Rsquare	0,235029
Adj Rsquare	0,190031
Root Mean Square Error	2,782884
Mean of Response	8,684211
Observations (or Sum Wgts)	19

t Test

B-A

Assuming equal variances

Difference	-2,9222	t Ratio	-2,2854
Std Err Dif	1,2786	DF	17
Upper CL Dif	-0,2245	Prob > |t|	0,0354*
Lower CL Dif	-5,6199	Prob > t	0,9823
Confidence	0,95	Prob < t	0,0177*

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	40,44971	40,4497	5,2231	0,0354*
Error	17	131,65556	7,7444		
C. Total	18	172,10526			

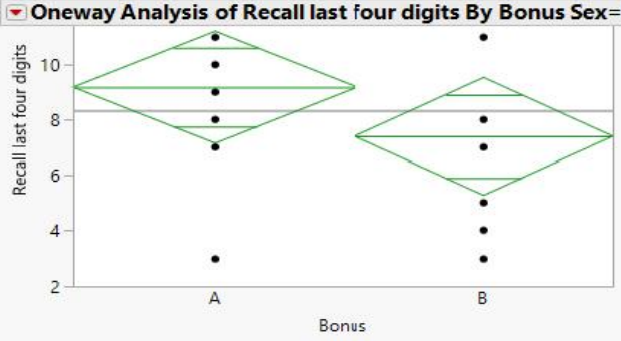
Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	9	10,2222	0,92763	8,2651	12,179
B	10	7,3000	0,88003	5,4433	9,157

Std Error uses a pooled estimate of error variance

Independent-two-samples t-test for *Recall last-4 digits* by bonus for men.

Oneway Analysis of Recall last four digits By Bonus Sex=M



Oneway Anova

Summary of Fit

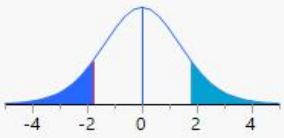
Rsquare	0,07747
Adj Rsquare	0,028916
Root Mean Square Error	3,228462
Mean of Response	8,333333
Observations (or Sum Wgts)	21

t Test

B-A

Assuming equal variances

Difference	-1,7818	t Ratio	-1,26315
Std Err Dif	1,4106	DF	19
Upper CL Dif	1,1706	Prob > |t|	0,2218
Lower CL Dif	-4,7343	Prob > t	0,8891
Confidence	0,95	Prob < t	0,1109



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	16,63030	16,6303	1,5955	0,2218
Error	19	198,03636	10,4230		
C. Total	20	214,66667			

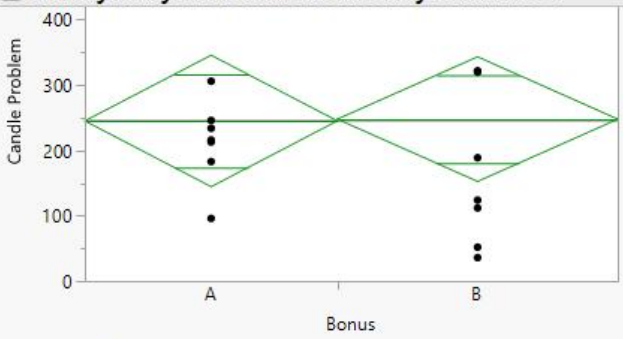
Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	11	9,18182	0,9734	7,1444	11,219
B	10	7,40000	1,0209	5,2632	9,537

Std Error uses a pooled estimate of errorvariance

Independent-two-samples t-test for *Candle problem* by bonus for women.

Oneway Analysis of Candle Problem By Bonus Sex=F



Oneway Anova

Summary of Fit

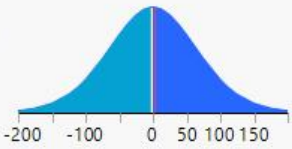
Rsquare	0,000105
Adj Rsquare	-0,05871
Root Mean Square Error	143,0219
Mean of Response	246,6842
Observations (or Sum Wgts)	19

t Test

B-A

Assuming equal variances

Difference	2,78	t Ratio	0,042271
Std Err Dif	65,71	DF	17
Upper CL Dif	141,42	Prob > |t|	0,9668
Lower CL Dif	-135,87	Prob > t	0,4834
Confidence	0,95	Prob < t	0,5166



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	36,55	36,5	0,0018	0,9668
Error	17	347739,56	20455,3		
C. Total	18	347776,11			

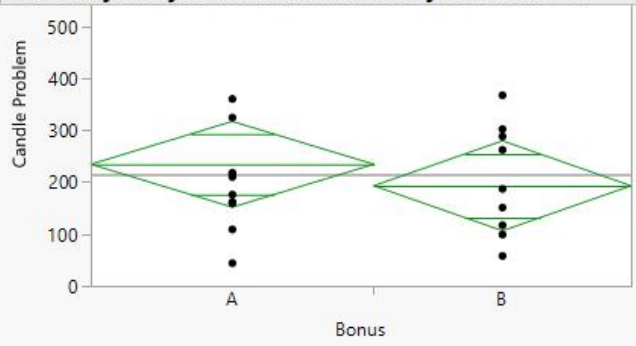
Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	9	245,222	47,674	144,64	345,81
B	10	248,000	45,228	152,58	343,42

Std Error uses a pooled estimate of error variance

Independent-two-samples t-test for *Candle problem* by bonus for men.

Oneway Analysis of Candle Problem By Bonus Sex=M



Oneway Anova

Summary of Fit

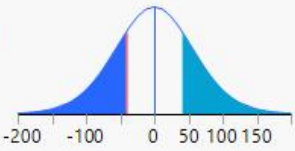
Rsquare	0,026877
Adj Rsquare	-0,02434
Root Mean Square Error	130,712
Mean of Response	214,5714
Observations (or Sum Wgts)	21

t Test

B-A

Assuming equal variances

Difference	-41,37	t Ratio	-0,72441
Std Err Dif	57,11	DF	19
Upper CL Dif	78,16	Prob > |t|	0,4776
Lower CL Dif	-160,91	Prob > t	0,7612
Confidence	0,95	Prob < t	0,2388



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	8966,06	8966,1	0,5248	0,4776
Error	19	324627,08	17085,6		
C. Total	20	333593,14			

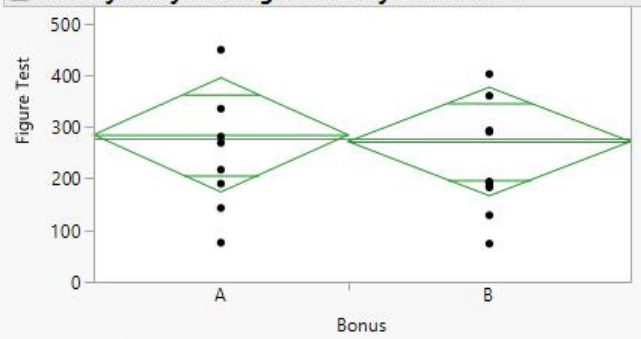
Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	11	234,273	39,411	151,78	316,76
B	10	192,900	41,335	106,39	279,41

Std Error uses a pooled estimate of error variance

Independent-two-samples t-test for *Figure test* by bonus for women.

Oneway Analysis of Figure Test By Bonus Sex=F



Oneway Anova

Summary of Fit

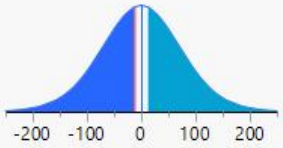
Rsquare	0,001859
Adj Rsquare	-0,05686
Root Mean Square Error	157,1052
Mean of Response	277,6842
Observations (or Sum Wgts)	19

t Test

B-A

Assuming equal variances

Difference	-12,84	t Ratio	-0,17794
Std Err Dif	72,18	DF	17
Upper CL Dif	139,45	Prob > |t|	0,8609
Lower CL Dif	-165,14	Prob > t	0,5696
Confidence	0,95	Prob < t	0,4304



Analysis of Variance

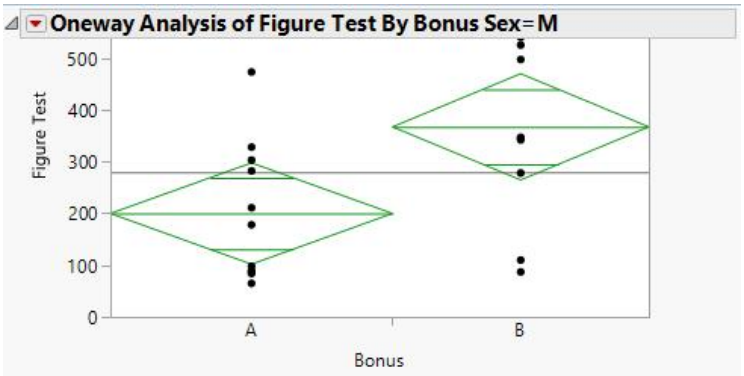
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	781,48	781,5	0,0317	0,8609
Error	17	419594,62	24682,0		
C. Total	18	420376,11			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	9	284,444	52,368	173,96	394,93
B	10	271,600	49,681	166,78	376,42

Std Error uses a pooled estimate of error variance

Independent-two-samples t-test for *Figure test* by bonus for men.



Oneway Anova

Summary of Fit

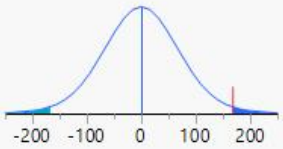
Rsquare	0,243258
Adj Rsquare	0,203429
Root Mean Square Error	154,917
Mean of Response	279,4762
Observations (or Sum Wgts)	21

t Test

B-A

Assuming equal variances

Difference	167,282	t Ratio	2,471361
Std Err Dif	67,688	DF	19
Upper CL Dif	308,955	Prob > |t|	0,0231*
Lower CL Dif	25,609	Prob > t	0,0115*
Confidence	0,95	Prob < t	0,9885



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Bonus	1	146578,70	146579	6,1076	0,0231*
Error	19	455986,54	23999		
C. Total	20	602565,24			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	11	199,818	46,709	102,05	297,58
B	10	367,100	48,989	264,56	469,64

Std Error uses a pooled estimate of error variance