NORGES HANDELSHØYSKOLE
Bergen, spring 2012

# Are we more rational when it matters more?

*A study of decision-making under risk among university students in Dar Es Salaam*

**Siv-Elisabeth Skjelbred**

**Supervisor: Bertil Tungodden**

Masters thesis in Economic analysis

NORGES HANDELSHØYSKOLE

# Preface

The data used in this thesis comes from an experiment that was conducted by Alexander Cappelen, Shackar Kariv, Erik Sørensen and Bertil Tungodden as a part of their research on a comparative analysis of rationality between students at University of Dar Es Salaam, Tanzania and University of California, Berkeley. I was responsible for carrying out the experiment in Tanzania and it was conducted in February 2012. The CCEI calculations are conducted by Shachar Kariv as a part of their research. I will thank the four researchers for allowing me to participate in their research and use their data in my thesis.

I would like to thank Bertil Tungodden for being my supervisor and for good guidance throughout the process of writing this master thesis. Also, a big thank you to: Lyamai Juda, Ngowi Tumainiel and Jesse Frank Kiwelu, for great research assistance in Tanzania.

I would also like to thank the other master thesis writers for The Choice Lab for giving motivation and being good discussion partners along the way. A special thanks to Cecilie Rasmussen, late research coordinator at NHH, for excellent work during the conduction of this experiment and for being an inspiration.

Bergen 12.06.2012

Siv-Elisabeth Skjelbred

# Abstract

This thesis explores whether there is a difference in the rationality of the decisions at high- and low-stake levels. The rationality of the decisions are measured in two dimensions, the quality of individual decisions, measured by first order stochastic dominance (FOSD), and the quality of a set of decisions, or consistency, measured by the Critical Cost Efficiency Index (CCEI). The thesis introduces the term strict rationality for decisions that are both consistent and non-dominated.

The analysis is based on a lab experiment carried out at University of Dar Es Salaam, Tanzania, in which the participants were bachelor students at the university. The effect of high-stake is studied by a treatment and control experiment, where the treated subjects received higher incentives in a decision making setting than the control group. The lab-experiment uses a graphical tool-kit developed by Choi, Gale, Fishman and Kariv (2007a) designed to investigate the economical preferences of individuals.

The thesis discusses four theories for the effect of incentives: the standard view, the learning view, the effort view and the psychological view. The main finding of this thesis is that high-stake subjects spend on average more time on each decision, which for some subgroups, extravert subjects and low-IQ subjects, translate into more strictly rational decisions. Furthermore the thesis finds that the subjects are in general making quite rational decisions, considering the unfamiliar situation and tool-kit. Furthermore there is a significant learning effect as the subjects make less dominated decisions in the middle decisions than in the early decisions.

# Content

# 1. Introduction

Making choices is an important part of the daily lives of every individual in the world. The importance of the choices ranges from quite unimportant choices, such as what to have for dinner, to quite important choices influencing the lives of many, such as the choices made by leaders around the world. Decision-making is thus an important field to study, supported by the fact that in most economic models involving an agent, the agent's role is to make a decision.

The quality of the decision made, often referred to as rationality, is subject to debate. The term rationality is commonly referring to whether people display consistent preferences, which means that they can be considered as utility maximisers. Many researchers have investigated the consistency of the preferences (Arierly 2008, Tversky and Kahneman 1986 & 1974, Choi, Kariv, Müller and Silverman 2011), measuring the quality of a set of decisions.

This thesis will use a framework which makes it possible to look at rationality also in terms of the quality of each individual choice (Choi, Fishman, Gale and Kariv 2007a), measured by first order stochastic dominance. This thesis introduces the term strict rationality for decisions that are both non-dominated and consistent, and argues that this is a stronger test for rationality than what is commonly done in economic research.

The main aim of this thesis is to study whether people take better decisions when the decision is important. More specifically it aims to investigate whether the strict rationality in decisions under risk is affected by the importance of the decision, measured by the size of the stake. Intuitively one might expect that people take more well-founded decisions when taking important decisions; however, related literature (Kahneman 2011) suggests that some decisions are made more based on a gut-feeling than a true consideration of the alternatives.

The analysis is based on data from a lab-experiment conducted in Tanzania. It uses randomisation and treatment in order to investigate causal effects. It seeks to complement the research on rationality conducted by Choi, Kariv, Müller and Silverman (2011) as well as filling a gap in the incentive research on how incentives affect the rationality of the decisions.

The main finding of this thesis is that high-stake subjects spend on average more time on each decision and for some subgroups this translates into more strictly rational decisions. Furthermore the thesis finds that the subjects are in general largely consistent with utility maximisation even with low incentives. The thesis discusses the results and compares them with the indications from four related theories: the standard view, the effort view, the learning view and the psychological view.

The remainder of the thesis is organised as follows. Section 2 contains a review of the literature on decision-making and section 3 introduces strict rationality and the distinction between good choices and consistent choices. Section 4 describes related theory of how stakes are expected to affect the rationality of the decisions. Section 5 discusses how one can identify causal effects, while section 6 describes the experimental design. Section 7 motivates the empirical strategy, section 8 contains the analysis and section 9 contains a discussion of the results. Section 10 is a short summary of the results.

## 2. Rationality: Litterature review

In the following section the theoretical foundation for rationality is outlined. The thesis focuses on the study of rationality with regards to the stakes, in situations where an agent chooses between different combinations of two goods which can be illustrated in a two-dimensional diagram. The two goods can be called X and Y and the amounts held of each good corresponds to the value of the axes. An example is illustrated in figure 1, where an agent is faced with alternative A, which contains α units of Y and φ units of X, and alternative B, which contains β units of Y and γ units of X.

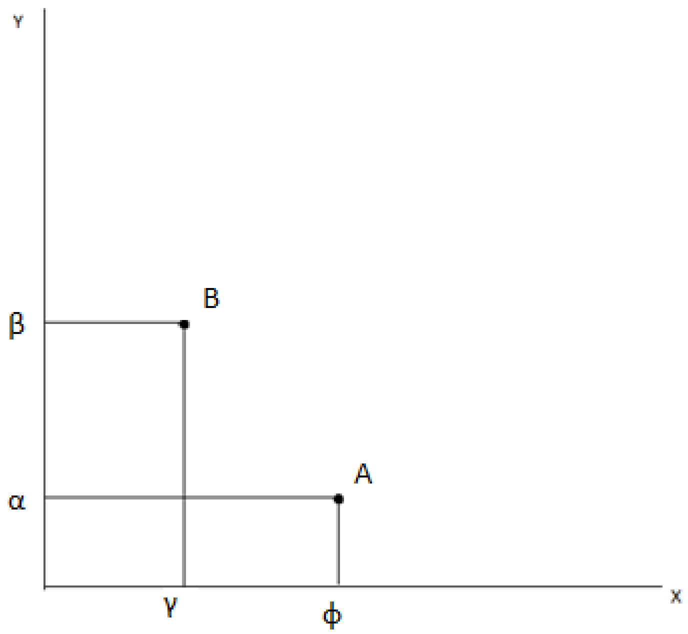


*Figure 1 Illustration of alternatives*

In addition, the focus is decisions under uncertainty, which means that X and Y is considered as states of the world, which each occur with an objective probability[1] of ½. This means that the agent is faced with alternative risky prospects or ‘lotteries’ and the outcome of the decision will depend upon the state of the world (Savage 1954). In the example it means that if the agent chooses alternative A he will receive α units of Y or φ units of X each with probability ½.

[1] Objective probability is the frequency at which an event tend to occur (Knight 1921)

## 2.1 Standard definition of rationality

In economics there are two statements which are consistent with the definition of rationality: (1) The agent has consistent preferences, which means that the preferences can be illustrated with indifference curves and utility functions (2) The agent is able to make optimal and consistent decisions given these preferences. This thesis focuses on the first dimension and assumes that agents are able to make optimal decisions given their preferences.

Consistency of preferences only requires that choices correspond and does not include an evaluation the quality of the preferences (Sen 1993). An argument for not evaluating the quality of the preferences is that there may be preferences that can rationalise any choice. Analogous to this is that it cannot be argued with certainty that an agent is making a bad decision if he prefers one apple to five oranges, as he may have a strong preference for apples which would rationalise the decision. This means that as long as the agent will consistently prefer one apple to five oranges, he is rational by the standard definition.

## 2.2 Conditions for rationality

The two main conditions for rational behaviour are transitive and complete preferences. (Pindyck and Rubinfeld 1995)

***Transitive preferences****: If an agent prefers alternative A to alternative B, and alternative B to C, then he must also prefer A to C.*[2]

***Complete preferences:*** *The consumer can compare and rank all alternatives. In other words, for any two alternatives A and B, a consumer will prefer A to B, B to A or will be indifferent between the two.*[3]

If an agent does not have transitive preferences, he has cyclical preferences, which means that he prefers A to B to C to A. This is inconsistent as the agent will never be content with holding an alternative, and will always be willing to pay to switch to one of the other

[2] [3] Pindyck and Rubinfeld 1995

alternatives. The completeness condition ensures that all alternatives are evaluated and ranked.

With complete and transitive preferences, and a finite number of alternatives, it is possible to represent the consumer's preferences graphically with the use of indifference curves or utility functions (Von Neumann and Morgenstern 1944). An indifference curve shows all combinations of alternatives which provide the same level of satisfaction to the agent, meaning that the agent is indifferent between all combinations on the curve. The agent will have an indifference curve for every possible level of satisfaction and a set of indifference curves is called an indifference map.

A utility[4] function is obtained by assigning a value to each level of satisfaction, to represent the preferences. A utility function is an ordinal ranking of preferences as alternative A has a higher value than alternative B if A is preferred to B. There are many types of utility functions (Von Neumann and Morgenstern 1944, Közengi and Rabin 2006, Kahneman and Tversky 1979, Gul 1991) which could represent different types of choices. The focus of this thesis is whether there is *any* utility function which can represent the choices of the agent and not the form of this function.

In addition to transitive and complete preferences, when the alternatives presented are combinations of two goods the condition that all goods are desirable is introduced. This assumption makes it possible to use the revealed preference approach to study consistency.

***Desirable goods****: The consumer always prefers more of any good to less.*[5]

## 2.3 Revealed preferences

The revealed preference approach identifies the preferences of an agent based on the choices he makes. Suppose an individual, facing the budget constraint B from figure 2, chooses alternative X. Then he indirectly says that he prefers X to any other alternative on the B line including Y; X is revealed preferred to Y. Furthermore, if the agent when presented with the

[4] Utility is an ordinal ranking which measures the level of satisfaction that a consumer gets from consuming a good or undertaking an activity.

[5] Pindyck and Rubinfeld 1995

budget line A chooses alternative Y, then Y is revealed preferred to Z. Then the agent has indirectly revealed that he prefers X to Z, as X is preferred to Y which is preferred to Z.

The indifference curve going through X must lie beneath the dark-grey square as any point in the square is preferred to X by giving more of both goods. In addition, the indifference curve must lie above the light-grey area, as alternative X is preferred to all alternatives in that area. By adding budget constraints and decisions, the area in which the indifference curve must lie in is reduced, which makes it possible to make a good prediction of the shape of the indifference curve.

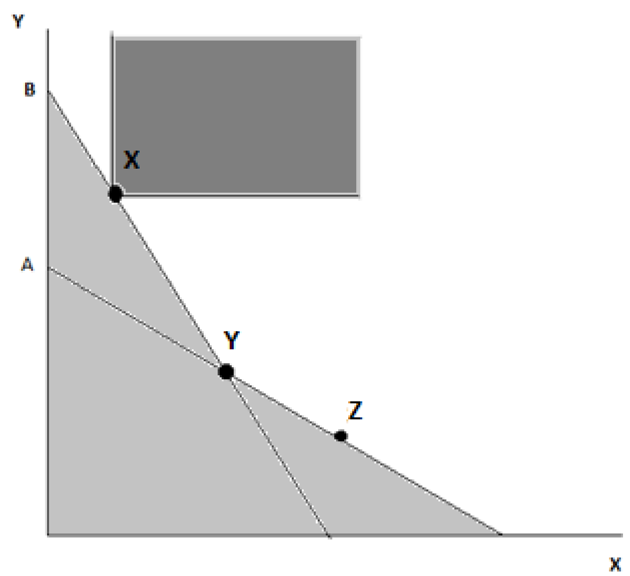


*Figure 2 – Explanation of revealed preferences (Pindyck and Rubinfeld 1995)*

The revealed preference approach can, combined with Generalised Axiom of Revealed Preferences (GARP), be used to identify whether there is a utility function which rationalises the choices the agent has made (Afriat 1967).

***Generalised Axiom of Revealed Preferences:*** *If an alternative X is revealed preferred to Y, then Y is never strictly directly revealed preferred to X, that is, X is never strictly within the budget set when Y is chosen.*

Preferences that are consistent with GARP are also consistent with maximising a utility function. A utility function rationalises the choice X, if at any given price, the chosen alternative has utility that is at least as high as any other available alternative. Figure 3 illustrates an example of a violation of GARP, given that the agent chooses alternative X on budget line A and alternative Y on budget line B. Making these two decisions violates GARP as X is strictly within the budget set when Y is chosen. Furthermore it will give inconsistent preferences, which can be illustrated by introducing the two possible alternatives Q and Z, on budget line A and B respectively. Given the presence of alternative Q and Z the following preference relations can be represented:

Y is preferred to Z (By decision)
Z is preferred to X (By more is better)
X is preferred to Q (By decision)
Q is preferred to Y (By more is better)

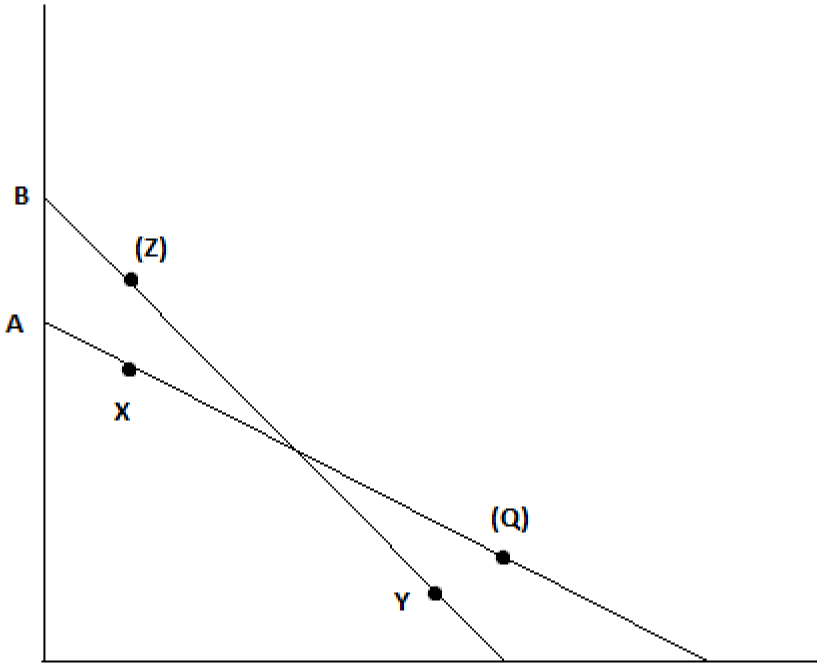


*Figure 3 – Illustration of violation of GARP (Pindyck and Rubinfeld 1995)*

Given the four preference relations the agent has inconsistent which means that Y is preferred to X and X is preferred to Y and the agent violates GARP.

## 2.4 Measuring consistency

GARP offers an exact test for whether the data is consistent with maximising a utility function, as the data either satisfy GARP or not. To quantify the extent of the GARP violations a variety of indices have been introduced (Varian 1991, Houtman and Maks 1985, Afriat 1972). The main index is the Critical Cost Efficiency Index (CCEI) (Afriat 1972) and is the focus of this thesis.[6] CCEI provides a summary statistic of overall consistency with GARP.

***Critical Cost Efficiency Index (CCEI)*** *measures the minimum amount by which each budget constraint must be adjusted in order to remove all violations of GARP (Choi, Kariv, Müller and Silverman, 2011).*

Figure 4 illustrates adjustments for decisions made on budget line B and budget line D. Choosing alternative X and Y at budget constraint B and D respectively, violates GARP as X is within the budget set when Y is chosen. The CCEI is a test of how much the budget constraint must hypothetically be shifted down in order to remove the GARP violation. This means that the agent's budget is reduced to a situation in which X is not strictly within the budget set when alternative Y is chosen.

To illustrate, a shift in the budget constraint from B to A is necessary for the choice of alternative X to not violate GARP. The new budget line A is slightly below alternative Y, making Y unaffordable to the agent when X is chosen. This means that the violation is removed as the agent could not have chosen Y with this hypothetical budget line.

With two decisions, it is only necessary to shift one of the budget constraints to remove the GARP violation, meaning that the adjustment is either shifting from B to A or D to C. CCEI measures the minimum amount to which the budget constraint must be adjusted, which means that the lowest possible shift is chosen. In the illustration the shift from B to A is the lowest and CCEI is calculated as A/B, giving a number between zero and one, where one means that the data is completely consistent with GARP and no adjustments are necessary.

[6] Choi, Kariv, Müller and Silverman (2011) give reason to believe that the different methods give reasonably the same results in measuring the extent to which the data displays utility maximising preferences

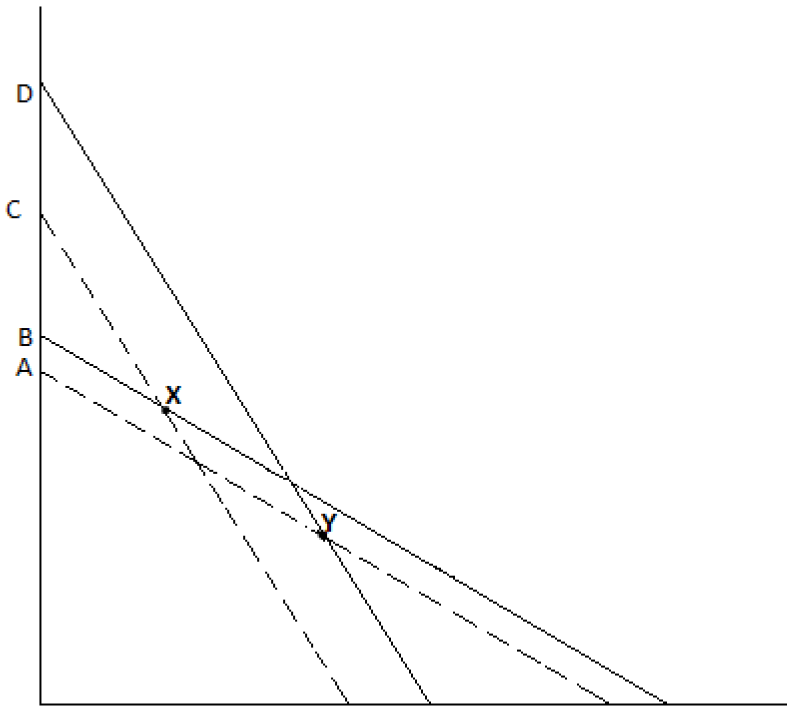


*Figure 4 – Illustration of how CCEI is calculated (Choi, Fishman, Gale and Kariv 2007a)*

The difference between CCEI and one can be interpret as the upper fraction of income that an individual is wasting by making inconsistent choices, because the CCEI measures the fraction of the budget constraint which must be removed in order to remove all violations of GARP (Choi, Kariv, Müller and Silverman 2011). If CCEI is below one, some of the budget is removed to eliminate the GARP violations and it may be argued that the agent does not exploit the full budget he receives. He is wasting the opportunity to make choices at higher budget constraints, indirectly leaving money left at the table.

# 3. Introducing strict rationality

As outlined in the literature overview rational behaviour has traditionally been synonymous with consistent preferences. However, displaying consistent preferences may be considered as weak rationality, as strictly rational decision should have two elements: the quality of each independent decision and the consistency of the set of decisions. In most situations economists refrain from evaluating the quality of the individual decisions; however, presenting alternatives as allocations between two goods is a situation in which it is uncontroversial to argue that agents can make bad decisions, as they may make first order stochastically dominated choices. The term rational will in the continuation of this thesis thus be associated with decisions that are both non-dominated and consistent. The term quality of the decisions will be used when discussing the separate elements of strict rationality.

***Strictly rational decisions*** *= non-dominated choices + consistent preferences*

## 3.1 The elements of strict rationality

The first requirement of making strictly rational decisions is making non-dominated choices. Figure 4 illustrates a decision problem an agent can be faced with, including a 45 degree line called the security line. In the illustration the axes x and y can be seen as representing two equally likely states of the world. The axes x and y can be considered as accounts in which the agent can allocate money to, and depending on the state of the world he will receive either the money in account x or the money in account y. As x and y constitute the same good, money, and the accounts are equally likely to be selected, the two goods should be equally valuable to the agent, indirectly ruling out that the agent can have preferences for achieving money through one of the accounts.

The budget line represents possible allocations between x and y, which means that it represents different lotteries the agent can choose between, which is analogous to choosing how much risk you want in a financial portfolio. The point A is the certainty equivalent in which the agent allocates equally much to the accounts x and y, eliminating the risk. In the illustration the x asset is the cheaper asset as the decision-maker has to give up more than one unit of x to acquire one additional unit of y.

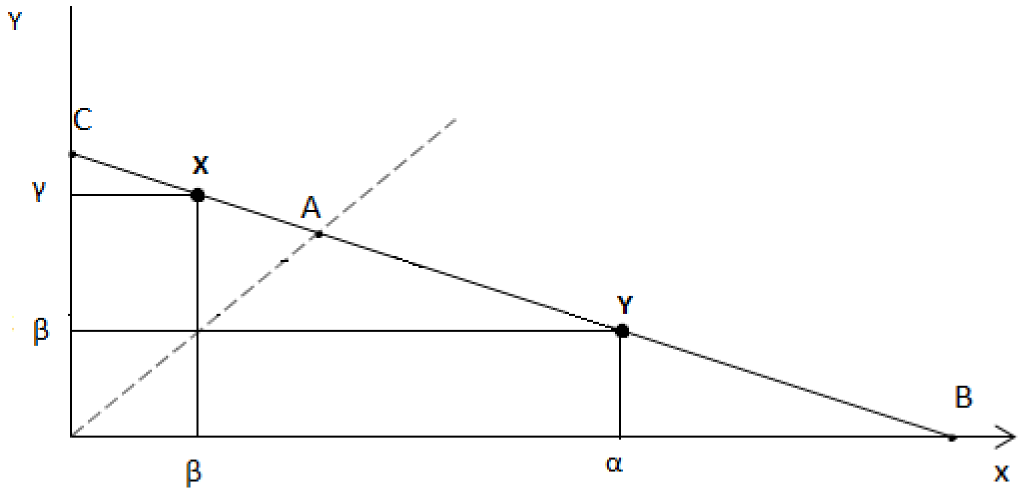


*Figure 5 – Illustration of a decision problem an agent can be faced with and a dominated choice (Choi, Fishman, Gale and Kariv 2007a)*

By definition first-order stochastic dominance is:

***First-order stochastic dominance****: A first order stochastically dominated alternative is an alternative for which it exists alternatives that give higher expected pay-off for the same risk, or the same expected pay-off for lower risk. (Mas-Colell, Whinston and Green 1995)*

A strictly rational individual would always take choices between A and B in this decision problem as any choices to the left of A are first order stochastically dominated, and would constitute a bad decision in this setting. All the alternatives to the left of A are dominated by possible alternatives to the right of A which give higher expected pay-off for the same risk.

To give an example, the allocation X is first order stochastically dominated by Y. The expected pay-off of X is given by: $E(X) = \frac{1}{2}\gamma + \frac{1}{2}\beta$ while the expected pay-off of Y is given by: $E(Y) = \frac{1}{2}\alpha + \frac{1}{2}\beta$. As $\alpha > \gamma$ the expected pay-off of Y is higher than X and they have the same risk as both alternatives will pay out minimally $\beta$ with the same probability. Choosing X would thus not be considered as a part of strictly rational decisions as long as the accounts are equally probable to be selected and the person does not have preferences for winning through one of the accounts. First order stochastic dominance can be measured by:

$$FOSD = \frac{C + E}{\left(1 - \frac{C}{EM} * CM\right) + C}$$

Where C = allocation to cheaper asset, E = allocation to most expensive asset, CM = maximal possible allocation to the cheapest asset, EM = maximal possible allocation to the most expensive asset. To make all non-dominated decisions valued equally the measure should be forced to be no larger than one, meaning that all allocations to the right of A should have an FOSD score of one. This is done by setting FOSD to one, if C is larger than or equal to E.

The FOSD-measure is thus bound between zero and one, and the closer it is to zero, the further to the shorter side of the 45 degree line has the agent made the choice, or in figure 5, the closer to C. Measures close to one can be evaluated as a trembling hand where the intention was to choose allocation A from figure 5, but the agent was not accurate in enforcing the decision.

The second measure in strict rationality is consistency, measured by CCEI as explained in section 2.4. Previous test of consistency only collected a few decisions from each subject, which in addition to using binary choices designed to compare the predictive abilities of competing preference theories (Camerer 1995, Starmer 2000), made it harder to truly measure consistency.  Presenting decisions in a graphical manner, makes it possible to make many decisions in a short period of time on a variety of budget lines, which makes it possible to reveal preferences and measure consistency to a larger extent than traditional binary choices.

CCEI provides overall consistency with the data as it is a summary measure of how much each budget constraint must be relaxed in order to remove all GARP violations. This means that the CCEI score does not give any information about the size of each violation or the number of violations. Furthermore, it means that an agent that has made one large violation and an agent that have made many small violations may achieve the same CCEI score. The Houtman-Maks index[7] overcomes this problem partly as it counts the minimum number of decisions that must be eliminated in order for the data to be consistent with GARP. However, it does not give information about the extent to which each of the eliminated decisions violates GARP. Given this the focus of the analysis will be on the CCEI score.

---

[7] Houtman-Maks index is very computational as it controls all possible combinations of decisions that may not violate GARP

Choi, Fishman, Gale and Kariv (2007a) found 0.95 to be a reasonably measure for very consistent decisions. They conducted an experiment where 25 000 hypothetical subjects chose randomly on the budget lines presented, and none of the subjects obtained a CCEI score above 0.95. This means that it is unlikely that a subject could obtain a CCEI score above 0.95 without making conscious decision. Furthermore, only 12 out of the 25 000 random choosing subjects had a score above 0.9, which means that also scores above 0.9 can be evaluated as quite consistent and not compatible with random decision.

It is important to keep in mind that dominance and consistency are fundamentally different. An agent can for instance make many dominated choices while still revealing consistent preferences. An example of this is if the agent always allocates everything to x. He will be making dominated choices whenever y is cheaper than x, but his preferences will be consistent as he will always prefer x. On the other hand, the subject can make only non-dominated decisions in an inconsistent manner. An example is if he chooses alternative A in figure 5 (page 11) when first met with the alternatives, but chooses alternative B when met with the same alternatives later.

## 3.2 Measuring strict rationality

To measure strict rationality the elements of dominance and consistency must be combined into one univariate measure. A threshold of 0.95 on the CCEI score is argued to be a measure describing the agent as strictly consistent and the same threshold can be imposed on FOSD. A possible requirement is that a strictly rational agent must achieve both a CCEI score above 0.95 and a FOSD score above 0.95, which will be a strict test for rationality. An alternative is a measure that scales strict rationality, which can be obtained by equally weighing the CCEI and FOSD scores.

Another alternative measure, which is already exists in literature (Choi, Kariv, Müller and Silverman 2011), can be used if the assumption that agents must display symmetric preferences is imposed. For each of the decisions made, a hypothetical mirrored budget constraint can be constructed and the symmetric decisions of the agent can be imposed. For example, as illustrated in figure 6, if the agent chooses alternative α when faced with the budget line from A to B, the mirror budget line is B' to A' and the symmetric decision is alternative α'.

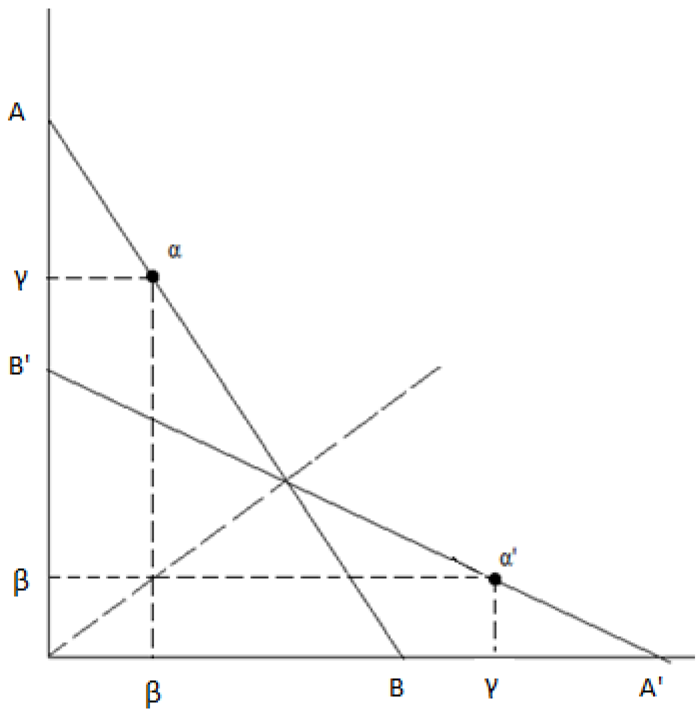


*Figure 6 – Illustration of mirror budget line (Choi, Kariv, Müller and Silverman 2011)*

With the mirrored budget lines, the number of observations is doubled, as there are 50 real observations and 50 hypothetical mirrored decisions. The total of the 100 decisions can be referred to as the combined data set and the CCEI score for the combined data set is referred to as CCEI-combined. The CCEI-combined tests both the dominance of the decisions and the consistency of the preferences because all dominated decisions will have a mirrored decision which is strictly within the budget set of the real decision, which means that the decisions violate GARP. For example, a consistent, but dominated strategy, is always allocating everything to x, in which the agent would choose the intercept B on the original budget constraint. The mirrored decision will be the intercept B'. This will lead to a GARP violation as B is strictly within the budget set when B' is chosen, and the dominated decisions requires a CCEI adjustment. The CCEI-combined is a stronger test of GARP as it includes dominated decisions and it will always be lower than the CCEI for the real observations. The CCEI-combined thus measures the subject's ability to make trade-offs in both a consistent and a non-dominated manner.

# 4. Does stakes matter?

The main aim of this thesis is to explore whether the size of the stakes, or more generally the size of the incentives, matter for the level of rationality. To my knowledge, there is no existing literature on incentives and rationality. However, the literature on incentives and performance is related as will be argued under section 4.2 and 4.3. This chapter summarise and discuss related literature on what the effect of stakes is expected to be on the rationality of decisions. Related literature on performance concludes that the effect of incentives is mixed (Camerer and Hoghart 1999) and that the effect of incentives may depend on the situation.

## 4.1 Standard view

Standard economic theory assumes a fully rational human being, denominated as Homo Economicus, in theories and models for decision-making. Homo Economicus is assumed to make rational and logical decisions that provide him with the greatest satisfaction he can achieve. In other words the standard view assumes that agents are rational by nature and will always make good decisions. Given this theory the size of the stake will not affect the rationality of the decision as people will be rational at any level of incentives.

New research (Choi, Kariv, Müller and Silverman 2011, Ameriks, Caplin and Leahy 2003) however, argues that there is heterogeneity in the ability to make rational decisions and that this ability depends on factors like sex, age and education. They introduce the term low-quality decision-makers which are people who make choices they may not have taken if they had the skills or the knowledge to make more rational decisions. This new research does not imply that there will be an effect of incentives, but that it may be heterogeneity in the measures of rationality, as it will depend on individual characteristics.

***Standard view:*** *People are rational by nature and will make non-dominated and consistent decisions regardless of the stakes.*

***Alternative view:*** *People are as rational as they can be by nature, but the extent of rationality depends on individual characteristics which are not influenced by the size of the stakes.*

## 4.2 Effort

An alternative view on rationality is that agents have the potential to be strictly rational; however, being rational is a process which requires effort and thus is costly for the agent (Binmore 1992). Economists generally assume that agents do not work for free, and that they will work harder, more persistently and more effectively when given incentives (Camerer and Hogarth 1999). However, this statement rest on the assumption that incentives increases motivation and effort. Some research supports this statement: incentives does make subjects work harder and give more correct answers (Libby and Lipe 1992) and raising the incentives from zero to a low level has a positive effect on performance (Camerer and Hoghart 1999). On the other hand, crowding-out theory (Frey and Jegen 2001) show that incentives may undermine intrinsic motivation[8] which could reduce performance and there may be psychological mechanisms interacting with the incentive levels.

For an agent to make a rational trade-off between risk and expected pay-off he must identify all dimensions relevant to the decision and weigh every alternative against each other based on his preferences, which would require effort and cognitive[9] ability (Eysenck 2009). Making rational trade-offs in itself without incentives does not constitute an interesting task, as the agent may consider the decisions to be unimportant as he will evaluate hypothetical lotteries. Decision-making is thus not expected to be subject to intrinsic motivation, which means that effort should increase with incentives. If increased effort increases performance, and there is no other factors influencing the rationality of the decision on the margin, the incentives will have a positive effect on rationality. However, if making rational decision is considered as a task which is very easy to do or very hard to improve on, the marginal monetary return to effort will be low, and the subject may not increase effort with higher incentives (Samuelson and Bazerman 1985 & Kahneman, Peavler and Onuska 1968).

It can be studied whether higher measures of rationality in high-stake decisions are due to increased effort, by using time spent on the decisions as a proxy for effort. Time is an

[8] Intrinsic motivation refers ot motivation that is driven by an interest in the task itself, and not due to external pressure.

[9] Cognition is a mental process and usually refers to whether an individual is capable of processing information and applying knowledge to solve the problem.

imperfect measure as it only describes the duration of the decision process and not the intensity of effort put into the decision.

***Effort view:*** *Rational decisions require the elicitation of effort, and effort increases with incentives.*

## 4.3 Learning

Learning theory is another possible mechanism which may influence the rationality measures. As the experimental setting is often unusual for the subject, they may need time to adjust to the setting in order to make rational decisions. Binmore (1992, p 51) argues that an experimentalist should be cautious about making predictions based on the observed data unless the following criteria are satisfied: (1) the task at hand is reasonably simple, (2) the incentives are adequate and (3) there is sufficient opportunity for trial and error. The argument is thus that subjects may not make rational decisions in unfamiliar circumstances if they do not have the time to learn the setting and understand how the decisions should be made.

In general the learning effect can be studied by presenting an agent with a set of decision. If the learning effect is present the agent's last decisions will be better than the first decisions. This means that the learning effect can be studied by controlling whether there is a significant difference between the rationality of the first decisions and the last decisions in a set of decision. Combining this with the incentive theory, an agent that is faced with higher stakes may have a higher incentive to learn as he may lose more by making bad decisions.

***Learning view****: Rational decisions are dependent upon the familiarity of the situation and the possibility for learning, and the incentive for learning may be influenced by the size of the stakes.*

## 4.4 Psychological mechanisms

The effort theory rests on the following assumptions: (1) incentives increases motivation and effort (2) the increased effort results in increased performance.  However, it can be psychological barriers which affects whether effort increases performance. Making decisions that is evaluated by researchers may feel like being in a performance setting. In a performance setting a person may experience feelings like excitement, stress, pressure and self-consciousness. Excitement increases the awareness and helps focus the attention. Excitement positively affects performance up to a point, and after this point increasing excitement reduces performance as the excitement is too high (Yerkes and Dodson 1908).

Another effect of the performance situations is if the task at hand is very important, which may lead to the sensation of pressure on performance. Too much pressure can lead to performance decrements[10], for example it is shown that basketball players miss more free throws in a game than during training (Dandy, Brewer and Tottman 2001). One explanation for performance decrements due to pressure is increased self-consciousness which shifts the behaviour from practiced automatic skills to a consciously controlled process which in some situations are less effective (Baumeister 1984).

Linking this to decision making, it is argued that a person has two different internal processes of thinking, often referred to as system 1 and system 2 (Kahneman 2011). System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control, while system 2 allocates attention to the effortful mental activities that demand it, such as complex calculations. Decisions based on system 1 can be considered as intuition based judgements and decisions based on system 2 can be considered as deliberate consideration of alternatives which requires effort. If rational decisions are best made by using the intuition based automatic skill (system 1), increasing the incentives may increase the consciousness, and reduce the performance. However, if rationality is dependent upon conscious consideration of alternatives, increased incentives may motivate the agent to make the effort necessary to use system 2 in the decision. The latter case is most likely as system 1 is considered to be more prone to errors of biases and heuristics (Kahneman 2011).

The presence of psychological mechanisms as stress and excitement indicate that the effect of incentives depends upon the situation, the task at hand and the abilities used in the task. Arierly, Gneezy, Loewenstein and Mazar (2009) found performance decrements associated with too large incentives on tasks requiring creativity, concentration and motor skills. Furthermore, they found that performance drops on task that require cognitive thinking, such as calculations, but increase on tasks that only require effort, such as pushing two buttons

[10] Performance decrement due to pressure is often referred to as "choking under pressure", which is inferior performance despite individual striving and situational demands for superior performance (Baumeister 1984)

alternately. As decision making require some cognitive ability the measures of rationality can be reduced with higher stakes.

***Performance view:*** *Rational decisions may be disrupted by increased pressure caused by too large incentives*

***Process of thinking:*** *Rational decisions require mental activity which can be stimulated by increased incentives*

# 5. The importance of stakes: Identifying causality

A fundamental challenge when studying the impact of stakes on rationality is identifying the causal[11] effect of the stakes. One possibility is to observe the choices people make in low-stake situations and high-stake situations and compare the rationality of the decisions in the two situations. For instance, research in finance (Shefrin and Statman 1985, Odean 1998) describes that small stake investors, like household investors, tends to realise more gains than losses; furthermore, the stocks they sell do better after the sale than the stocks they keep. This may indicate that decisions on a small scale are not consistent with maximising profit, which in turn could mean that people are less rational on small scale decisions.

A regression of this data would give a positive coefficient in front of x as the size of the stakes would be positively correlated with measures of rationality. However, this does not necessarily mean that high stakes causes more rational decisions. Many different causal interpretations can be consistent with the same data. The results may be a result of a selection bias, meaning for instance that agents who know that they makes some bad decisions or feel that they do not have the insight necessary to make good decisions, choose to be small stake investors, and agents who is confident in their decision making choose high-stake investments. Another possible selection bias is that the investors who choose to be small-stake investors are more loss-averse[12] than the high-stake investors, and try to avoid the sensation of loss by not realising the losses.

The possibility of selection bias makes it harder to identify causal effects of high-stakes as the regressions may result in significant results based on the type of people selecting themselves to high-stake decisions and low-stake decisions. In other words, the two groups may have heterogeneity in other variables than the size of the stake, making it harder to identify the difference in stakes as the causal effect of the difference in the rationality of the decision. If the other variables are observable, such as age and education level, the difference in these factors can be controlled for by adding covariates to the regression. If there is heterogeneity in unobservable variables or variables for which there is no observations, an

[11] Causality is the relationship between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first.

[12] People prefer to avoid losses, and evaluate losses more than equal sized gains.

omitted variable bias[13] may occur, and the estimates may be inconsistent and biased, thus not valid.

The omitted variable problem can be eliminated by replacing the variable of interest with an instrumental variable[14]. The instrument should not be correlated with the unobserved characteristics and it should be correlated with the variable that is explored, or in other words it should be external[15] and exogenous[16] (Heckman 2000). The instrumental variable approach is a way to identify causality, which seeks to ensure that the factor of interest is the one causing the observed difference in the data. Randomised controlled trials can be viewed as the ultimate instrumental variable as it fulfil both criteria for a good instrument.

## 5.1 Randomised controlled trials

Randomised controlled trials are a method for holding all other factors fixed and exploring the causal effect of a variable. Randomised controlled trials have been used for a long time in the natural sciences due to a requirement of controlling what would happen to a patient if he did not receive medicine, to identify the real effect of the medicine. As a patient cannot both receive treatment and not receive treatment randomised controlled trials have been used.

Randomised controlled trials are studies in which participants are allocated at random to receive an intervention, called treatment, or not to receive intervention, called control. Random allocation means that all participants have the same chance of being assigned to each of the two groups. The randomisation into treatment and control will, in theory, ensure that the two groups are identical, on average, on all other observable and unobservable variables than the treatment difference. Given that the randomisation is complete[17], any significant difference between the mean responses of the two groups can be attributed to the

---

[13] An omitted variable bias occurs when a variable is omitted from the regression which should have been in the model

[14] An instrumental variable in a variable used instead of the variable of interest which should be uncorrelated with the error term in the equation and correlated with the variable of interest.

[15] An external variable is a variable that is determined outside the model

[16] An exogenous variable satisfies the orthogonality condition, it is statistically independent of the other factors in the model

[17] The two groups are identical on average on all other observable and unobservable characteristics than the treatment

treatment and not some other unidentified factor. The instrumental variable is thus the assignment to treatment, which perfectly predict the variable of interest, namely the effect of the size of the stake (Stolberg, Norman and Trop, 2004).

To ensure statistical significant result, it is important to consider the sample size. If the expected effect of the treatment is relatively small the sample must be larger to ensure a significant effect than if the expected effect it large. Sample sizes are difficult to determine as too small samples may not give significant results, and too large samples may be more costly than necessary. Medium sized samples are thus often used.

Let me provide a theoretical presentation which proves that randomisation and treatment identifies causality (Angrist and Pischke 2008, p 122). The example is related to analysing the measures of rationality if the treatment constitutes of getting higher incentives than the control group, which will be referred to as high-stake incentives. Thus the subjects may be allocated to the treatment, denoted by $T_i$=1, which means that they receive high-stake incentives, or control, denoted by $T_i$=0, which means that they receive low-stake incentives. The outcomes are denoted as $Y_{i_0}$ and $Y_{i1}$, where $Y_{i0}$ is the outcome if the subject is not treated, that is if he receives low-stake incentives, and $Y_{i1}$ is the outcome if the subject is treated, that is, if he receives high-stake incentives. In randomised experiments the averages of participants receiving treatment and not receiving treatment is obtained, and the difference is calculated with equation (1).

$$(1)\quad E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 0).$$

By simple manipulation equation (2) can be obtained:

$$(2)\quad E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1) + E(Y_{i0}|T_i = 1) - E(Y_{i0}|T_i = 0),$$

where $E(Y_{i0}|T_i = 1)$ may be interpret as the outcome the treated subject would have had if he was not treated. In this case it means the decisions a high-stake subject would have made in a low-stake situation. The sum of the two last terms is the selection bias which is the difference between the outcome of the treated subjects if they did not receive the treatment and the outcome of the non-treated subjects. The selection bias is thus the difference between the rationality measure of the decisions high-stake subjects would have made in a low-stake situation and the decisions low-stake subjects make. If the randomisation is complete the treated subjects should have had the same outcome if they did not receive treatment as the non-treated subjects. That is: the high-stake subjects would have made equally rational decisions as low-stake subjects as the low-stake subjects did. Given this equation (3) is given:

$$(3)\ E(Y_{i0}|T_i = 1) - E(Y_{i0}|T_i = 0) = 0,$$

which would make equation (2) into:

$$(4)\ E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1).$$

The rules of expectation allow us to rewrite equation (4) into equation (5).

$$(5)\ E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1) = E(y_{i1} - y_{i0}|T = 1)$$

This means that as long as the randomisation is perfect, meaning that there is no selection bias, the observed difference between the averages of the two groups can be interpret as the difference in outcome caused by the treatment.

# 6. Experimental design

## 6.1 Background

To study the effect of incentives, an experiment with two incentive level was conducted. The lowest incentive level was equal to the typical pay-off in experiments, because it is interesting to study whether the incentives researchers typically provide is sufficient for rational decisions. The high incentive level was chosen to be quite high, in order to compare it to important decisions. To be able to give truly high-stake incentives within a reasonable budget, the experiment was conducted in a developing country. Tanzania was chosen because The Choice Lab has experience with running experiments in Tanzania and had necessary local assistance.

Even though this thesis focuses on the role of the stakes, the experiment had another main purpose. The experiment was conducted to compare the rationality of the decisions in a developing country and a developed country. Thus the same experiment was also conducted at University of California, Berkeley. The results from UC Berkeley will be used as an indicator of what the results could have been with standard incentives, at a school where the subjects are more used to experiments and computers.

## 6.2 Sample

The experiment was conducted with 220 students from the University of Dar Es Salaam, Tanzania. The subjects were bachelor students, mainly recruited from social science department and department of information and communication technologies. 33 % of the subjects were female. The subjects ranged from 19 to 32 years old, but the majority of the group, 90 %, was below 25 years.

The comparison experiment at University of California Berkeley was conducted with 47 subjects.

## 6.3 Treatment

The subjects were randomly allocated to two groups, one group with high-stake exchange rates and one group with low-stake exchange rate for the points earned in the experiment. Subjects allocated to receive high-exchange rates will be denominated as high-stake subjects, and subjects allocated to receive low exchange rates will be denominated as low-stake subjects. The results in Table 7-1 indicate that the randomisation was successful on observable variables as there are no significant differences between the two groups on background variables. The treatment was given by randomly drawing a desk number to be seated at, where the desks to the left in the room were allocated to low-stake treatment and desks to the right in the room were allocated to high-stakes treatment. Splitting high- and low-stakes physically in the room minimised the risk that the subjects became aware of the treatment difference, as they would be sitting next to people with the same exchange rate as themselves. This is an important factor as the subjects in the high- and low-stake treatment could have reacted differently if they knew that they had different incentives for the task than others.

The experiment was identical for all subjects in all respects but the exchange rate. The subjects in the high-stake treatment had a conversion rate of 1 token[18] to 1 000 Tanzanian Shilling (TZS), while the low-stake treatment had a conversion rate of 1 token to 100 TZS. It was ensured that the subjects understood the conversion rate and the meaning of tokens by requiring that some test-questions were completed before the experiment continued.

A short follow-up survey that documented living standard among students was conducted in order to achieve a better picture of how large the size of the stakes was for the subjects. One of the expenditure questions was to specify all expenses from yesterday, which gives an illustration of the size of the stakes compared to expenses on a typical day. This was also one of the advantages of conducting the experiment in Tanzania as they typically buy food and pay for transport on every day basis.

[18] Token was the value-label of the axes

**Table 6-1 Verification of randomisation**

| | (1)<br>Low-stake | (2)<br>High-stake | (3)<br>Difference |
|---|---|---|---|
| Female | 0.308 | 0.364 | -0.056 |
| | (0.045) | (0.046) | (0.065) |
| Age | 23.154 | 23.345 | -0.192 |
| | (0.181) | (0.202) | (0.272) |
| Secondary school mother | 0.404 | 0.364 | 0.0302 |
| | (0.048) | (0.046) | (0.067) |
| Secondary school father | 0.567 | 0.555 | 0.013 |
| | (0.049) | (0.048) | (0.068) |
| Mother to university | 0.125 | 0.091 | 0.034 |
| | (0.032) | (0.027) | (0.042) |
| Father to university | 0.308 | 0.309 | -0.001 |
| | (0.045) | (0.044) | (0.063) |
| IQ | 12.837 | 12.573 | 0.264 |
| | (0.583) | (0.523) | (0.781) |
| Extraversion | (0.310) | (0.266) | 0.043 |
| | (0.034) | (0.029) | (0.044) |
| Agreeableness | 0.761 | (0.780) | -0.019 |
| | (0.030) | (0.025) | (0.039) |
| Neuroticism | -0.303 | -0.350 | -0.047 |
| | (0.035) | (0.034) | (0.048) |
| Outgoing | 0.396 | (0.404) | -0.008 |
| | (0.030) | (0.029) | (0.041) |
| Conscientiousness | 0.835 | 0.834 | 0.001 |
| | (0.032) | (0.027) | (0.042) |

*Note: (1) and (2) represents the averages of the control group (low-stake) and the treatment group (high-stake) respectively and standard errors in parenthesis. (3) presents the differences between low-stake and high-stake and standard errors in parenthesis. *p<0.10, **p<0.05, ***p<0.01.*

The expected pay-off was 75 tokens, which means that a subject could expect to receive 7 500 TSZ in the low-stakes treatment and 75 000 TSZ in the high-stakes treatment. The average reported expenses yesterday was about 11 000 TZS, meaning that the low-stake expected payoff was about 70% of the average expenses of a typical day.

An illustration of the size of the high-stake is the yearly costs. The tuition fee for students and University of Dar Es Salaam ranges from 1 000 000 TSZ to 1 500 000 TSZ, depending on the study programme, which means that the high-stake expected pay off minimum constitute 5 % of the yearly tuition fees.

The comparison experiment at UC Berkeley had an exchange rate of 0.6 USD to each token, which in exchange rate corresponds to 1000 TSZ. However, as prices in Tanzania are lower than the prices in the US, the purchasing power[19] of the 1 000 TZS received in USA is lower

[19] Purchasing power is the number of goods you can buy for one unit of a currency

than in Tanzania. The purchase power rate of Tanzanian Shilling to US dollar is 0.4[20], which means that, 400 TSZ shilling could buy the same amount of goods in Tanzania as 1000 TSZ could buy in the USA. Given this the incentives at UC Berkeley was closer to low-stakes than high-stakes in terms of purchase power.

## 6.4 The experiment

The experiment contained two parts. The first part was not incentivised and included a standard IQ-test, a Big-Five personality analysis and some background questions such as age and sex. This first part was conducted before the subjects were introduced to exchange rates and the incentivised part, in order to make the information unaffected by the treatment, which is a requirement for using it as covariates.

In the second part the subjects were presented with decision problems in the graphical interface from Choi, Kariv, Müller and Silverman (2011). This part was incentivised and payoffs were calculated in tokens, which where the value label of the axis. The subjects were instructed that the tokens would be converted into Tanzanian Shillings by the conversion rate handed out on paper.

The subjects were presented with 50 of decisions under risk as illustrated in figure 7. Each decision problem was presented as a choice from a two-dimensional line, a budget line. The subjects were instructed to choose an allocation on the budget line, which would allocate tokens between two accounts labelled x and y, which corresponds to the axis. Furthermore, they were told that these two accounts represented two equally likely states of the world and the actual payoff of a particular choice was determined by the tokens allocated to the account that was drawn. They were instructed that the decision could be viewed as a decision of which lottery they were willing to accept. No feedback was given about their decision.

The choices were made using the computer mouse to move the pointer on the computer screen to the desired point and clicking at this point. Each decision problem started with the computer selecting a budget line randomly from a set of budget lines that intersects with at least one of the axis at ten or more tokens and with no intercept exceeding 200 tokens. The

[20] World Bank (2010) http://databank.worldbank.org/data/Views/Reports/TableView.aspx

budget lines were thus independent of each other and independent of the budget lines of the other participants.

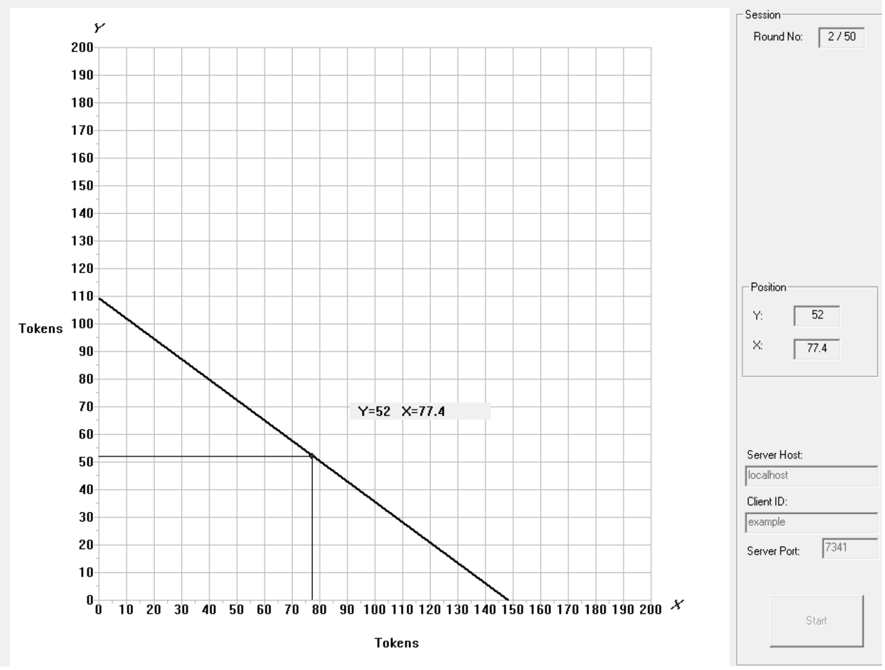


*Figure 7 A screen shot of the framework and a possible decision problem presented the participants.*

As the slope of the budget line varied with the intercepts of x and y, the relative prices of x and y, and the income the subjects could distribute varied. The slope of the budget line measures the relative prices, and y is relatively cheaper than x if the slope-coefficient is larger than one. In the example in figure 7, x is relatively cheaper than y and the subject has a larger budget to distribute than if the budget line was closer to origin. The variation in prices and income and the large number of decisions is essential for testing the consistency of the preferences.

## 6.5 Payment

To ensure that the subjects evaluated each decision independently and did not have risk-strategies over the set of decisions, the subjects were instructed that one out of the 50 decisions would randomly be drawn to be a real decision. With only one of the decisions chosen, the subjects should make each decision as it was real without considering the allocations they made in the previous decisions.

## 6.6 Missing data and data exclusion

Conducting an experiment that requires computers and stable network connection in a developing country can be a challenging task. This unfortunately led to some data being lost while conducting the experiment. For four subjects (104, 110, 111 and 121) the data from the decision part of the program is missing as a virus infected their computers. One subject (112) had to be dropped because there was no information from part one of the experiment. One subject (108) was dropped as the information about sex and age was not registered due to a technical error. The analysis does thus consist of data from 214 subjects.

When analysing treatment one should be careful with excluding observations as even though the initial sample was randomised and thus comparable, it is not necessary the case that the exclusions are random, which would make the sample biased. However, the virus infections and the technical problem affected the subjects randomly, meaning that the sample is not expected to be biased as a result of these eliminations.

# 7. Empirical strategy

Given the experimental design the following regression can be used to estimate the effect of treatment on the measures of rationality:

$$Y = \alpha + \beta T + \gamma X + e,$$

where Y is the outcome variable of interest, T is a dummy for assigned to high-stake, where T=1 if assigned to high-stake and T=0 if assigned to low-stake, X is background variables and e is the error term. The coefficient $\beta$ measures the change in Y with respect to T, holding all other factors fixed, or in other words the effect of high-stake treatment on the measure of rationality. Background variables are included in the regression in order to control for the effect of other variables on the measures of rationality. In addition, background variables are often included to avoid the omitted variable bias, as discussed in section 5, and to ensure a ceteris paribus analysis, holding all other factors fixed.

The experiment contains three measures of rationality: first order stochastic dominance (FOSD), CCEI and the CCEI-combined. One regression is thus FOSD on a dummy for treatment status and including the covariates age, sex, IQ, big-five personality traits and education level of parents. The coefficient in front of each of the background variables illustrate the effect that variable has on the rationality measure. The coefficient in front of the treatment variable measures the effect on the measure of rationality of being allocated to high-stake treatment. A positive coefficient would mean that the high-stake subjects take more rational decisions than the low-stake subjects, and a negative coefficient would mean that high-stake subjects take less rational decisions than the low-stake subjects. The standard errors of the coefficients can be used to construct a confidence interval[21] in which the coefficient is expected be.

The coefficient and its standard error are used to test the hypothesis that the variable has an effect on the measures of rationality. The null hypothesis is that the coefficient is zero, which for the treatment variable means that high-stakes does not impact the measure of rationality. The standard hypothesis test is the t-test which assumes that the dependent variables have

[21] A confidence interval is an interval in which the coefficient with a decided probability is expected to be in

normal distribution with the same variance in each group. These conditions are typically met in medium sized samples. If the regression includes background variables the null hypothesis is that once the background variables are accounted for, the treatment does not have an additional effect on the measure of rationality. The t-statistic is calculated as below:

$$t - statistic\ = \frac{coefficient}{standard\ error\ of\ coefficient}$$

When the expected effect, whether the coefficient is positive or negative is not well determined by theory, a two-sided test should be conducted. A two-sided test means that the alternative hypothesis is that the coefficient is different from zero, with no requirement in which direction. The effect of treatment is not theoretically clear as it depends on which of the discussed mechanisms that affects rationality, thus a two-sided test is conducted in this thesis. The absolute value of the t-statistic is compared to a critical value from a t-distribution table, determined by the significance level and the degrees of freedom. If the absolute level of the t-statistic is higher than the critical value, the null hypothesis can be rejected.

Alternatively the t-statistic can be used to find the smallest significance level at which the null hypothesis can be rejected, called the p-value. If the p-value is 0.1 there is a 10 % chance that the null hypothesis is rejected when it is true, and the coefficient is said to be significant at a 10 % level. Analogously it can be said that the coefficient is different from zero with 90 % certainty, or that zero is not in the 90 % confidence interval of the coefficient.

One of the possible measures for strict rationality is a dummy variable[22] which takes the value 1 if the subjects have both FOSD and CCEI score above 0.95, and zero if this criterion is not met. A regression where the dependent variable is a dummy is called a linear probability model, and the coefficient in front of the treatment variable is interpret as the impact being allocated to high-stake has on the probability that a subject is strictly rational.

If the treatment does not have the same effect on all subgroups in the sample, the average effect can hide interesting details which can be found by a control for heterogeneous effects.

---

[22] A dummy variable is one that takes the value zero or one to indicate the absence or presence of some categorical effect that may be expected to shift the outcome.

For instance, there is a possibility that the effect of incentives is unequal for men and women as research observe that women and men respond differently to changes in context (Gneezy and Croson 2009). Heterogeneous effects may result in non-significant average effects if the effect of one sub-group is not large enough to raise the average for the full sample, or if two subgroups have effects that neutralise each other.

The interaction effect approach enables a control for heterogeneous effects. An interaction effect occurs when the effect of the treatment variable with respect to the dependent variable, depends on the magnitude of another variable (Woolridge 2009). An example is that the effect of the incentives may depend upon the sex of the subject. The interaction effect can be used in combination with dummy variables. For example a dummy variable for female and a dummy variable for being allocated to high-stake treatment would give the regression below:

$$y = \alpha + \beta F + \gamma\, T + \phi(T * F)$$

Where T is the dummy for treatment and F is the dummy for female. In this regression the coefficient $\beta$ is measuring the difference between men and women in the control group, $\gamma$ is measuring the effect of treatment on men, $\phi$ is measuring the difference between the treatment effect on men and women and $\gamma$ + $\phi$ measures the treatment effect on women, and must be tested jointly to test significance. The dummy variable approach includes an assumption that all other factors influence the two subsamples equally. The alternative is to run separate regressions for the two subsamples when controlling the treatment effect, but this makes it difficult to estimate and test the treatment difference between the subsamples.

Some experiments are criticised as they do not measure what would happen in a real setting, either as a result of the unfamiliar circumstances or as an experimenter demand effect. The experimenter demand effect occurs if the subject acts accordingly to what they think the experimenter's purpose is (Rosenthal and Rosnow, 2009). As the subjects are unaware of the purpose of the experiment and it is hard to interpret it from the task, the experimenter demand effect is expected to be low. Making decisions in an experiment may be different than making decisions in the real world even though the decisions made in this experiment replicate how an investor chooses a financial portfolio. However, the effect of being in an experiment is expected to affect both treatment groups equally, which means that it would only affect the level of the scores and not the difference between the two treatment groups. This means that higher or lower measures of rationality may be observed than what would occur outside the experiment, but as long as both groups are affected equally the observed treatment effect will be the same as what would be observed without these effects.

# 8. Analysis

The analysis begins with a study of each of the two elements of strict rationality: dominance and consistency and a discussion of the implications these results have on the learning hypothesis. The analysis continues with a study of strict rationality, which is completed with a study of effort and heterogeneous effects.

## 8.1 Dominance

The analysis of dominance is conducted by studying the average FOSD score for each subject, which is a measure of the quality of the decisions on an individual basis. The average FOSD score has two elements: the individual FOSD scores and the number of decisions. Thus dominance will be studied in more detail by studying the lowest observed FOSD score for each individual, as a measure of the largest mistake of each individual, and the number of non-dominated decisions. The study of the elements of dominance can shed light on the average scores, as the average score for two individuals could be the same if one subject made one considerably dominated decision and another made many slightly dominated decisions.

The average FOSD for the full sample is 0.976, indicating that the subjects make reasonably good decision on average.[23] Table 8-1 reports the average FOSD and the distribution of FOSD in the two treatment groups. The mean FOSD of the lows-stake group is slightly lower than the FOSD of the high-stake group and the percentile distribution of the FOSD score for high-stake subjects is mostly to the right of the one for low-stake subjects. However, the difference is small and not statistically significant, see table 8-2, which means that the two groups have close to similar quality of the decisions based on the average measure of first order stochastic dominance.

[23] If the subjects took only non-dominated decisions the FOSD score would have been one, indicating that they made some bad decisions.

**Table 8-1: First-order stocastic dominance distribution**

| | | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | 5 % | 10 % | 25 % | 50 % | 75 % | 90 % | 95 % |
| Low-stake | 0.975 | 0.919 | 0.947 | 0.967 | 0.982 | 0.993 | 0.998 | 1.000 |
| High-stake | 0.977 | 0.920 | 0.942 | 0.966 | 0.989 | 0.996 | 0.999 | 1.000 |

*Note: The table reports the average FOSD and the distribution of FOSD in the two treatment groups*

**Table 8-2 First order stochastic dominance regression**

| | (1) No covariates | (2) With covariates |
|---|---|---|
| High-stake | 0.0034 | 0.0037 |
| | (0.0038) | (0.0038) |
| Constant | $0.97^{***}$ | $0.97^{***}$ |
| | (0.0027) | (0.025) |
| *N* | 214 | 213 |

*Note: The table reports regressions of FOSD on treatment status, controlling for covariates. FOSD is a measure bound between 0 and 1 and calculated as described in section 7.1. High-stakes is a dummy variable for being allocated to high-stakes treatment. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. None of the covariates were statistically significant. Standard errors in parenthesis; $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.*

The lowest observed FOSD score for each individual can be viewed as a measure of the size of the largest individual mistake and can be used to study the extent of the dominance. A large portion of the subjects, 179 out of 214, make dominated decisions as they have a lowest observed FOSD-score below 0.95. The averages, reported in table 8-3, indicate that the mistakes done by the low-stake subjects are larger than the mistakes done by high-stake subjects, as the average lowest observed FOSD is 0.648 for low-stake subjects and 0.726 for high-stake subjects. Furthermore, the percentile distribution of the high-stake subjects is always to the right of the percentile distribution of the low-stake subjects. The results indicate that the high-stake subjects are less inclined to make large mistakes than the low-stake subjects. However, the difference is not statistically significant, reported in table 8-4, which means that it cannot be excluded that the observed difference is caused by random variance. The covariate of IQ is significant and indicates that subjects with higher IQ make smaller mistakes than subjects with lower IQ. This will be discussed further under the analysis of heterogeneity.

**Table 8-3, lowest observed FOSD for each subject**

| | | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | 5 % | 10 % | 25 % | 50 % | 75 % | 90 % | 95 % |
| Low-stake | 0.684 | 0.183 | 0.275 | 0.535 | 0.722 | 0.892 | 0.978 | 0.994 |
| High-stake | 0.726 | 0.240 | 0.316 | 0.574 | 0.801 | 0.931 | 0.981 | 0.996 |

*Note: The table reports the average and the percentile distribution for the lowest FOSD observation for each subject.*

**Table 8-4 Regression of lowest observed FOSD**

| | (1)<br>Minimum observed FOSD<br>No-covariates | (2)<br>Minimum observed FOSD<br>With covariates |
|---|---|---|
| High-stake | 0.041 | 0.048 |
| | (0.034) | (0.034) |
| Female | | -0.0091 |
| | | (0.037) |
| Age | | -0.0061 |
| | | (0.0080) |
| IQ | | 0.0052* |
| | | (0.0030) |
| Extraversion | | 0.064 |
| | | (0.059) |
| Neuroticism | | 0.064 |
| | | (0.049) |
| Agreeableness | | 0.086 |
| | | (0.072) |
| Openness | | -0.054 |
| | | (0.073) |
| Conscientiousness | | 0.089 |
| | | (0.071) |
| Mother to university | | -0.042 |
| | | (0.067) |
| Father to university | | 0.044 |
| | | (0.048) |
| Mother to secondary school | | -0.0030 |
| | | (0.044) |
| Father to secondary school | | 0.019 |
| | | (0.046) |
| Constant | 0.68*** | 0.63*** |
| | (0.024) | (0.20) |
| *N* | 214 | 214 |

*Note: The table reports regressions of the lowest observed FOSD score for each subject on treatment status, controlling for covariates. High-stakes is a dummy variable for being allocated to receive high-stakes treatment. Standard errors in parentheses* * *$p < 0.1$,* ** *$p < 0.05$,* *** *$p < 0.01$*

To further analyse the extent of dominance, the number of non-dominated decisions, decisions with FOSD-score of one, is studied. A large number of dominated decisions may be a sign of lack of concentration or effort, and high-stake subjects may have a larger incentive to uphold concentration as it is more costly for them to make mistakes. However, this hypothesis is not supported by the data as the two treatment groups take on average about the same number of non-dominated decisions, reported in table 8-5.

The result is thus that receiving high-stake incentives did not give a statistically significant effect on the individual quality of the decisions, neither measured by average FOSD, the size of the largest mistake nor the number of non-dominated decisions.

**Table 8-5, number of non-dominated**

| | | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | 5 % | 10 % | 25 % | 50 % | 75 % | 90 % | 95 % |
| Low-stake | 36.02 | 26.0 | 28.0 | 31.0 | 36.0 | 40.5 | 45.0 | 47.0 |
| High-stake | 36.63 | 27.0 | 29.0 | 32.0 | 36.0 | 40.0 | 47.0 | 49.0 |

*Note: The table reports the average and the percentile distribution for the number of non-dominated decisions, that is decisions with FOSD=1, for each subject.*

## 8.2 Consistency

The consistency of the decisions is the second element of strict rationality and is measured by CCEI. A CCEI score above 0.95 is considered as quite consistent and not compatible with random decisions.[24] The subjects are displaying preferences that are largely consistent with utility maximisation as the average CCEI score for the full sample is 0.86 and 86 of the subjects have a CCEI score above 0.95.

The average CCEI score for the low-stake subjects is 0.854 and the average for the high-stake subjects is 0.868. Furthermore, the percentile distribution of the high-stake subjects is, at every level, to the right of the distribution of the low-stake subjects. The same is illustrated by the histogram in figure 8, where it can be observed that a larger fraction of the high-stake subjects are having CCEI scores in the upper range. However, this observed difference is not statistically significant, reported in table 8-7, which means that the observed difference may be caused by random variance.

**Table 8-6, CCEI distribution**

| | | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | 5 % | 10 % | 25 % | 50 % | 75 % | 90 % | 95 % |
| Low-stake | 0.854 | 0.648 | 0.687 | 0.782 | 0.888 | 0.972 | 1.000 | 1.000 |
| High-stake | 0.868 | 0.585 | 0.658 | 0.755 | 0.936 | 0.990 | 1.000 | 1.000 |

*Note: The table reports the percentile distribution and the average CCEI score for high- and low-stake subjects.*

[24] See section 3.1 or Choi, Fishman, Gale and Kariv (2007a) for discussion

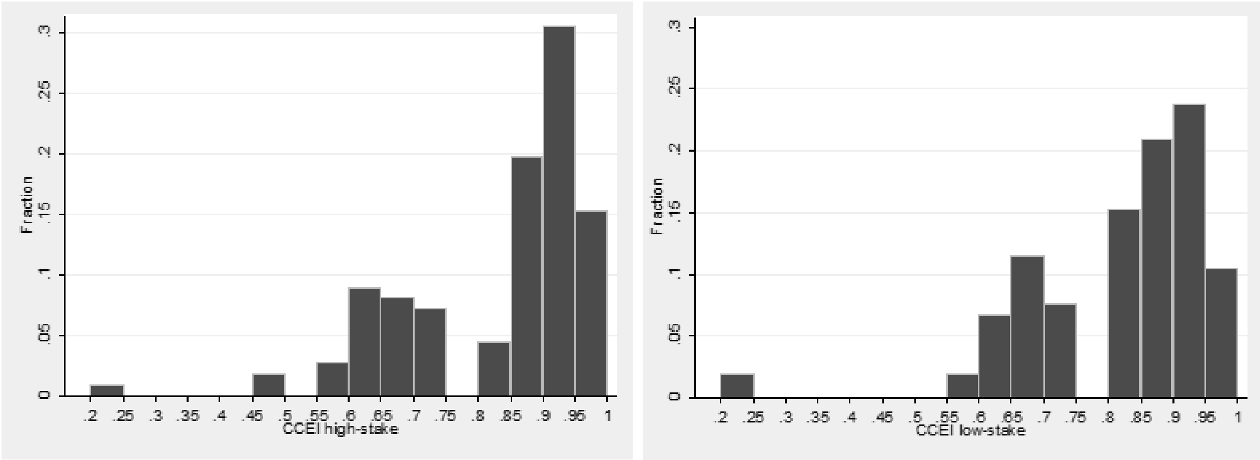


*Figure 8 – Distribution of observed CCEI scores, with high-stake distribution to the left and low-stake distribution to the right.*

**Table 8-7 Critical Cost Efficiency Index**

| | (1)<br>CCEI<br>No covariates | (2)<br>CCEI<br>With covariates |
|---|---|---|
| High-stake | 0.014 | 0.014 |
| | (0.020) | (0.020) |
| Constant | $0.85^{***}$ | $0.85^{***}$ |
| | (0.014) | (0.14) |
| *N* | 214 | 214 |

*Note: The table reports regressions of CCEI on treatment status. CCEI is a measure of how much the budget lines must be shifted in order to make the preferences consistent with GARP. CCEI is bound between 0 and 1, where the closer it is to 1 the less of a perturbation must be made in order to remove all GARP violations. High-stakes is a dummy variable for being allocated to receive high-stake treatment. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parenthesis; $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.*

## 8.3 Learning

A possible explanation for why the larger incentives are not translating into less dominated or more consistent decisions may be the learning effect. Given that subjects need time to adjust to the setting, the quality of the decisions made by the subjects may be equal in the first rounds, which would reduce the difference between the two groups and could cause statistically insignificant results. Due to the learning process, the effect of incentives may only occur in the later rounds; furthermore, the high-stake subjects may have a larger incentive to learn as their mistakes are more costly in absolute terms.

The learning hypothesis can be tested by comparing the FOSD scores for the first ten decisions and the mid ten decisions (Binmore 1992). The mid ten decisions are chosen because the subjects may get tired of the experiment which could reduce the quality of the decisions in the latest rounds. Since the decision problems are randomised across individuals any observed difference between the first and the mid decisions can be allocated to being early and late in the decision setting. Early and middle decisions can thus be studied as a treatment effect, where early decisions is the control where the subjects do not have time to learn and middle decisions is the treatment where subjects have been given time to learn the decision making setting. This means that any significant difference between the first and the middle decisions can be allocated to being early and late in the experiment.

Taking the average FOSD for the first ten decisions and middle ten decisions separate give the results in table 8-8. As the learning hypothesis states that the mid ten decisions should be better than the first ten decisions a one sided t-test for the difference is conducted. Both low-stake subjects and high-stake subjects take significantly less dominated decisions in the middle decisions than in the early decisions indicating that there may be a learning effect. The treatment effect, on the other hand, is neither significant for the first decisions, nor for the middle decisions.

**Table 8-8 Dominated choices by time of decision**

| | (1)<br>Low-stake | (2)<br>High-stake | (3)<br>Difference |
|---|---|---|---|
| FOSD first 10 decisions | 0.969 | 0.974 | -0.005 |
| | (0.004) | (0.003) | (0.005) |
| FOSD middle 10 decisions | 0.975 | 0.980 | -0.004 |
| | (0.004) | (0.003) | (0.005) |
| Difference | -0.006 | -0.005 | |
| | (0.003)* | (0.004*) | |
| *N* | 104 | 110 | |

*Note: (1) and (2) reports the averages FOSD scores of the first 10 decisions and the mid 10 decisions, as well as the difference between these, grouped by treatment status. (3) presents the difference between low-stake and high stake and standard errors in parenthesis; *p<0.1, **p<0.05 ***p<0.01*

## 8.4 Strict rationality

CCEI and FOSD measures the two separate parts of strict rationality and a measure of strict rationality should include both these measures. A criterion for strict rationality can be that both CCEI and FOSD should be above 0.95, which would be met by 51 of the subjects in the high-stake treatment and 35 of the subjects in the low-stake treatment. The linear probability model in table 8-9, show that it is 13 % more likely that a subject has both FOSD and CCEI scores above 0.95 if the subject is in the high-stake treatment. This result indicates that even though the difference between the FOSD scores and the CCEI scores are not statistically significant separately, a larger number of subjects in the high-stake treatment are *both* consistent *and* taking non-dominated choices. However, the results are not robust to a test where strict rationality is measured continuously with equal weights on CCEI and FOSD.

**Table 8-9 – Regression with dummy for strictly rational on high-stake**

| | (1) | (2) |
|---|---|---|
| | Strictly rational | Strictly rational |
| High-stake | 0.13* | 0.13* |
| | (0.067) | (0.068) |
| Constant | 0.34*** | 0.14 |
| | (0.048) | (0.14) |
| *N* | 214 | 214 |

*Note: The table reports a regression of strictly rational on treatment status, where regression 2 includes covariates. Strictly rational is a dummy for having both CCEI and FOSD score above 0.95 and high-stake is a dummy for being allocated to high-stake treatment. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parentheses; *p<0.1, **p<0.05 ***p<0.01*

A second control for strict rationality is the CCEI-combined, which is the CCEI score for the combined data set, with the 50 real observations and the 50 mirrored decisions. By mirroring all the decisions a measure of dominance is included in the CCEI score, as a decision that was previously dominated will in combination with the mirrored decision give a violation of GARP as explained in section 3.2.

The treatment effect of high-stake on the CCEI-combined is close to significant, with a p-value of 0.12 in the regression with covariates. Furthermore the coefficient is positive indicating that high-stake subjects take more strictly rational decisions than low-stake subjects. From the adjusted R-squared [25] it can be noted that adding the covariates does not increase the explanatory power of the model, meaning that there may be other covariates which could improve the model and reduce the standard errors.

[25] In linear regressions the $R^2$ is the sample correlation coefficient between the outcomes and their predicted values or in other words how much the variables included in the regression model explain the dependent variable. The adjusted $R^2$ adjust for the fact that adding covariates will always increase $R^2$, and the adjusted $R^2$ will only increase if the new term improves the model more than would be expected by chance. (Woolridge 2009)

**Table 8-10 – Regression of treatment on CCEI score for combined dataset**

| | (1)<br>CCEI for combined dataset<br>No covariates | (2)<br>CCEI for combined dataset<br>With covariates |
|---|---|---|
| High-stake | 0.047 | 0.049 |
| | (0.031) | (0.031) |
| Constant | $0.73^{***}$ | $0.78^{***}$ |
| | (0.023) | (0.183) |
| *N* | 214 | 214 |
| *Adjusted - $R^2$* | 0.006 | 0.005 |

*Note: The table reports regressions of CCEI for the combined dataset on treatment status, controlling for covariates. CCEI is a measure of the extent to which the budget constraints must be relaxed in order to remove all violations of GARP in the decision set, bound between 0 and 1. High-stake is a dummy for being assigned to receive high-stake treatment. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parentheses $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$*

## 8.5 Effort

There is some evidence that the high-stake subjects are more strictly rational than the low-stake subjects. A further control is whether the high-stake subjects have higher effort than the low-stake subjects, which would increase performance unless there are psychological mechanisms that may lead to performance decrements. In analysing effort, time spent on each decision can be used as a proxy. The causal effect of time cannot be identified with the data available, which means that it cannot be controlled whether taking more time on each decision causes more rational decision. A regression of time on CCEI would give a significant and positive coefficient; however, this can be a result of selection bias and does not necessarily mean that using more time causes the decisions to be more consistent.

However, it can be identified causally whether subjects exposed to higher stakes spend more time on the decisions. High-stake subjects are taking 2.36 more seconds on each decision than low-stake individuals, significant at a 10 % level. On average, low-stake subjects take 19.7 seconds, and high-stake subjects take 22.0 seconds per decision, reported in table 8-11. This provides suggestive evidence that high-stakes induces more effort, even though this does not translate into more rational decisions.

There are three possible explanations for why higher effort does not translate into statistically significant more rational decisions. A first possible explanation is that effort is not the only thing affecting the measures of rationality on the margin, which is the case if decision making is subject to psychological mechanisms like stress which can reduce performance. A second possible explanation is that increased effort only marginally affects the rationality of the decision. This would be the case if decision making is a task that is either hard to improve on or easy to do. For example an irrational subject, may not have the abilities necessary to make rational decisions, which would mean that he could only slightly increase his rationality for any level of incentives. On the other hand, a subject that is rational may consider the task at hand reasonably simple, requiring minimum effort, which would make him rational at any level of incentives. However, this second explanation is not necessarily consistent with the high-stake subjects spending more time on making the decisions, and this theory will be studied further under heterogeneity. A third possible explanation is that the incentives influence the subjects heterogeneously.

**Table 8-11 Time**

| | (1)<br>Time<br>No covariates | (2)<br>Time<br>With covariates |
|---|---|---|
| High-stake | 2.36$^{*}$ | 2.36$^{*}$ |
| | (1.24) | (1.26) |
| Constant | 19.7$^{***}$ | 7.55 |
| | (1.98) | (8.62) |
| *N* | 214 | 214 |

*Note: The table reports regressions of time on treatment status, where (2) controls for covariates. Time is measured by the average time spent on each decision and high-stake is a dummy variable for belonging to the high-stakes treatment. Covariates include age, sex, education level of parents (whether they attended secondary school and university), IQ and big-five personality traits. Standard errors in parentheses* $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

# 8.6 Heterogeneity

The lack of significance may reflect heterogeneous effects in the sample. Heterogeneity in the treatment effects occurs if a subsample responds differently to the treatment than another subsample. If there are heterogeneous effects of the treatment, interesting details may be hidden in the insignificant average scores. Three possible sources of heterogeneity: gender, personality and IQ, will be studied.

## 8.6.1 Gender

Gender may be a source of heterogeneity as research show that men and women react differently to changes in the context (Gneezy and Croson 2009). A possible explanation is thus that women get more stressed by high-incentives than men, giving a higher effect of high-incentives on men, and no, or even negative, effect of incentives on women.

Table 8-12 reports the regression of the CCEI-combined including an interaction effect. The high-stake dummy measures the effect of high-stake on men, which indicates that men in the high-stake treatment have a CCEI-combined which is 0.044 larger than the men in the low-stake treatment. The female coefficient measures the difference between women and men in the low-stake treatment and indicates that low-stake women have on average a CCEI-combined score which is 0.026 larger than the low-stake men. The interaction variable, high-stake*female, measures the difference of the treatment effect between women and men, which indicate that female have a higher positive effect of being allocated to high-stake treatment than men. Lastly, the sum effect female measures the total effect women have of high-stake incentives which indicate that women in the high-stake treatment have a CCEI-combined score which is 0.03 above the score of low-stake women. However, none of these coefficients are statistically significant, which means that it cannot be rejected that the observed difference is caused by random variation. The results are thus that the treatment effect is not statistically significant, neither for men nor for women and there is no statistically significant difference between the consistency of men and women.

**Table 8-12 Regression of CCEI-combined controlling for heterogeneous effect on women**

| | (1) CCEI for combined data set No covariates | (2) CCEI for combined data set With covariates |
|---|---|---|
| High-stake | 0.044 | 0.051 |
| | (0.038) | (0.039) |
| Female | 0.026 | 0.029 |
| | (0.048) | (0.050) |
| High-stake*female | 0.0039 | -0.012 |
| | (0.066) | (0.067) |
| Sum effect female | 0.030 | 0.018 |
| | (0.045) | (0.049) |
| Constant | $0.72^{***}$ | $0.56^{***}$ |
| | (0.027) | (0.069) |
| *N* | 214 | 214 |

*Note: The table reports the CCEI score for the combined data set on high-stake, female and interaction effect of high-stake and female, controlling for covariates in regression (2). The CCEI score is a measure of how much the budget constraints must be relaxed in order to satisfy GARP. High-stake is a dummy for being allocated to high-stake treatment. Female is a dummy for being female. High-stake*female is the interaction effect of being allocated to high-stake when you are a female. Sum effect female is the treatment effect on female subjects. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parentheses* $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

As a robustness check, separate regressions for men and women can be run in table 8-13, as the specification with covariates in table 8-12 assumes that the covariates are influencing men and women equally. The high-stake dummy from regression (4) in table 8-13 is the same measure as the sum effect female from regression (2) in table 8-12 and the high-stake dummy from regression (2) in table 8-13 is the same measure as the high-stake dummy in regression (2) in table 8-12. However, the coefficients of the two regressions are not similar, which indicates that the covariates influence the sexes differently. However, neither of the two methods results in statistically significant results.

**Table 8-13 Regression of CCEI-combined - by gender**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | CCEI male<br>No covariates | CCEI male<br>With covariates | CCEI female<br>No covariates | CCEI female<br>With covariates |
| High-stake | 0.044 | 0.063 | 0.048 | 0.048 |
| | (0.040) | (0.048) | (0.048) | (0.048) |
| Constant | 0.720 | 0.544 | 0.746 | 0.605 |
| | (0.028)*** | (0.085)*** | (0.036)*** | (0.109)*** |
| *N* | 142 | 142 | 72 | 72 |

*Note: The table reports regressions of CCEI on treatments status where (2) and (4) is controlling for covariates. CCEI is a measure of the extent to which the budget constraints must be relaxed in order to remove all violations of GARP in the decision set, bound between 0 and 1. High-stake is a dummy for being assigned to high-stake treatment. (1) and (2) includes male only while (3) and (4) includes female only.  . Standard errors in parentheses* $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

### 8.6.2 Personality

Extraversion is a personality trait that can be a source of heterogeneous effects as extravert and introvert subjects may respond differently to the context. Introvert subjects may feel uncomfortable in a situation where they feel controlled, making them more subject to stress as a result of high incentives. Extravert subjects, on the other hand, are considered as assertive, and may be more confident in their own ability to make the right decision and thus less inclined to be worried about making mistakes. As table 8-14 reports, there is a weakly significant effect of high-stake on extravert subjects. Extravert subjects have on average 0.110 higher CCEI-combined in high-stake treatment than in low-stake treatment. The coefficient of high-stake measures the effect of high-stakes on introvert subjects and supports a hypothesis that introvert subjects can have negative effect of high incentives, however, the coefficient is not statistically significant. The difference between extravert and introvert subjects in the control group is not significant, indicating that extravert and introvert subjects are equally consistent is general, but extravert subjects respond positively to high incentives.

**Table 8-14 Regression of CCEI-combined, controlling for heterogeneous effects of extraversion**

| | (1)<br>CCEI for combined data set<br>No covariates | (2)<br>CCEI for combined data set<br>With covariates |
|---|---|---|
| High-stake | -0.047 | -0.060 |
| | (0.080) | (0.081) |
| Extravert | -0.00018 | -0.019 |
| | (0.063) | (0.066) |
| Highs-stake*extravert | 0.11 | 0.13 |
| | (0.087) | (0.088) |
| Sum effect extravert | 0.110 | 0.108 |
| | (0.059)* | (0.061)* |
| Constant | 0.73*** | 0.79*** |
| | (0.058) | (0.22) |
| *N* | 214 | 214 |

*Note: The table reports the CCEI score for the combined data set on high-stake, extraversion and the interaction effect of high-stake and extraversion, controlling for covariates in regression (2). The CCEI score is a measure of how much the budget constraints must be relaxed in order to satisfy GARP. High-stake is a dummy for being allocated to high-stake treatment. Extraversion is a dummy for achieving a positive score on extraversion in the personality test. High-stake*extraversion is the interaction effect of being allocated to high-stake when you are an extravert person. Sum effect extravert is the treatment effect on extravert subjects. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parentheses* * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

### 8.6.3 IQ

A third possible source of heterogeneity is IQ. The subjects with high-IQ may be more used to high-stake settings, which suggest that there may be a positive interaction effect between high-IQ and high-stake. Furthermore, subjects with high-IQ may also be more used to the presentation of decisions in a graphical setting. Including a dummy variable for IQ, which is one if the subject belongs to the upper half of the IQ-scores and zero if he belongs to the lower half, give the regression in table 8-15. The coefficient of high-stake is statistically significant and measures the treatment effect on low-IQ subjects. It indicates that low-IQ subjects have a CCEI-combined which is 0.093 higher if they are allocated to high-stake treatment instead of low-stake treatment. In addition subjects with high-IQ have statistically significantly higher CCEI-combined than subjects with low-IQ in the low-stake treatment. The treatment effect is not statistically significant for high-IQ subjects.

**Table 8-15 Regression of CCEI-combined, controlling for heterogeneous effects of IQ**

| | (1) CCEI for combined data set No covariates | (2) CCEI for combined data set With covariates |
|---|---|---|
| High-stake | $0.093^{**}$ | $0.096^{**}$ |
| | (0.044) | (0.045) |
| High-IQ | $0.10^{**}$ | $0.095^{**}$ |
| | (0.044) | (0.045) |
| High-stake*High-IQ | -0.092 | -0.095 |
| | (0.061) | (0.063) |
| Sum high-IQ | 0.012 | 0.000 |
| | (0.043) | (0.045) |
| Constant | $0.68^{***}$ | $0.78^{***}$ |
| | (0.031) | (0.21) |
| *N* | 214 | 214 |

*Note: The table reports the CCEI score for the combined data set on high-stake, high-IQ and interaction effect of high-stake and high-IQ, controlling for covariates in regression (2). The CCEI score is a measure of how much the budget constraints must be relaxed in order to satisfy GARP. High-stake is a dummy for being allocated to high-stake treatment. High-IQ is a dummy for being in the upper half of the IQ distribution in the sample. High-stake*High-IQ is the interaction effect of being allocated to high-stake when you are in the high-IQ distribution. Sum effect high-IQ is the treatment effect on high-IQ subjects. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parentheses* $^{*}$ *$p < 0.1$,* $^{**}$ *$p < 0.05$,* $^{***}$ *$p < 0.01$*

Linking this to the result in the study of effort, it may be that the effort necessary for being rational is different for high-IQ subjects and low-IQ subjects. This can be studied by controlling whether there are heterogeneous effects of stakes on time spent on each decision. Table 8-16 show that there is a significant treatment effect of high-stake on the time spent on the decisions for the low-IQ subjects, which indicates that subjects with low-IQ scores spend 5.6 seconds more on each decision in the high-stake treatment than in the low-stake treatment. Subjects with high-IQ spend 4.94 seconds more on each decision than subjects with low-IQ in the low-stake treatment. The treatment effect on high-IQ subjects is not statistically significant.

The results indicate that it is not necessarily easier for high-IQ individuals to make strictly rational decisions, but they are willing to spend more time on the decisions in the low-stake treatment. The low-IQ subjects are induced to spend more effort with higher incentives, while there is no treatment effect on time for the high-IQ subjects. It may thus be the case that the high-IQ subjects take what they consider as necessary time for the decisions for low-incentives.

**Table 8-16 Time regression controlling for heterogeneous effect of IQ**

| | (1) | (2) |
|---|---|---|
| | Time | Time |
| High-stake | 5.60*** | 5.32*** |
| | (1.75) | (1.78) |
| High-IQ | 4.94*** | 4.33** |
| | (1.76) | (1.79) |
| High-stake*High-IQ | -6.42*** | -5.93** |
| | (2.45) | (2.49) |
| Sum effect high-IQ | -1.48 | -1.60 |
| | (1.71) | (1.77) |
| _cons | 17.2*** | 7.66 |
| | (1.24) | (8.48) |
| *N* | 214 | 214 |

*Note: The table reports the time on high-stake, high-IQ and an interaction effect of high-IQ on high-stake, controlling for covariates in regression (2). Time is the average time each subject spends on each decision, high-stake is a dummy for being allocated to high-stake treatment. High-stake is a dummy for being allocated to high-stake treatment. High-IQ is a dummy for being in the upper half of the IQ distribution in the sample. High-stake*High-IQ is the interaction effect of being allocated to high-stake when you are in the high-IQ distribution. Sum effect high-IQ is the treatment effect on high-IQ subjects. Covariates include sex, age, education background of parents (whether they attended secondary school and university) and big-five personality traits. Standard errors in parentheses* $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

# 9. Discussion

There is suggestive evidence that higher incentives cause more strictly rational decisions as a larger fraction of the high-stake subjects have both FOSD-score and CCEI score above 0.95. The CCEI-combined indicate that high-stake subjects take more strictly rational decisions; however, the results are not statistically significant. The results of dominance, consistency and strict rationality will first be discussed in accordance with each of the four theories: standard view, learning view, effort view and the psychological view, and will be followed by a discussion of more general explanations.

The insignificant treatment effect on dominance and consistency supports the standard view that subjects are rational by nature, and will make rational decisions regardless of the stakes. However, the subjects are not completely rational as they are making dominated and inconsistent decisions. Thus there is some support for the alternative hypothesis that rationality is subject to heterogeneity as a result of individual decision-making abilities, indicating that subjects are, by nature, as rational as they can be. However, if they were rational by nature, there would be no reason to take more time on high-stake decisions.

The significant treatment effect on time supports the theory that decision-making requires elicitation of effort and that effort increases with incentives. However, the result suggests that this increased effort does not necessarily translate into more strictly rational decisions. There are four possible reasons for why increased effort does not increase the rationality of the decision enough to achieve a statistically significant treatment effect: (1) The low-stake incentive level was high enough to induce quite rational behaviour reducing the potential difference (2) Increased effort only marginally affects the measures of rationality (3) Effort is not the only thing affecting the measure of strict rationality on the margin. (4) The incentives affected subgroups differently.

To shed light on the first argument, the average CCEI score at Berkeley was 0.934 with incentives that are considered as standard. This may indicate that low-stake incentives are high enough to induce the effort necessary for quite rational behaviour. In addition, the incentives at low-stake level may have been considered a considerable amount for the subjects, as it constituted about 70 % of the reported average daily expenses. However, if the low-incentive levels were sufficient for inducing the necessary effort, it does not explain why the high-stake subjects take more time on their decisions.

The second possible explanation is linked to the level of difficulty of making rational decisions. Decision making can be a task that is either considered as easy to do or difficult to improve on. The first will be the case if the subject does not have the abilities necessary for making rational decisions, which would make him irrational at any level of incentives and effort. This could imply that the treatment effect on time can be caused by subjects that are trying to improve their performance, but does not have the ability to make rational decisions.

The third possible explanation introduces the possibility of psychological mechanisms like stress. Stress cannot be controlled for as the experiment does not contain any measures of stress level. A hypothesis is that subjects get stressed by the high incentives, and stress reduces performance. As effort and stress work in opposite directions, the presence of stress could cause a close to zero net effect.

The fourth possible explanation is that there the different mechanisms affect subgroups differently. The data supports a heterogeneous effect due to the personality characteristic extraversion, indicating that introvert subjects are stressed as a result of the high incentives, while there is a statistically significant positive treatment effect on the extravert subjects. It may thus indicate that the psychological mechanism of stress affects some type of subjects more than others, being stronger for introvert subjects than extravert subjects.

A second source of heterogeneous effects of the treatment is differences in IQ. It seems that increased effort as a result of higher stake is caused by a large increase in effort by low-IQ subjects and this increased effort does results in more rational decisions. Low-IQ subjects have higher effort when allocated to high-stake treatment and they take more strictly rational decisions measured by CCEI-combined. High-IQ subjects, on the other hand, spend significantly longer time in the low-stake treatment than low-IQ subjects, and do not have a significant treatment effect on neither time nor strict rationality. Analysing heterogeneous effects are often subject to criticism as it may be considered as data-mining[26], which is why

[26] Data-mining is going through the data in subsets in order to find a significant effect and is built on a theory that a sufficiently determined examination of any trial will eventually lead to some subgroup for which the treatment yielded a significant effect of some sort (Deaton 2010)

the analysis is limited to three possible heterogeneous effects for which there was a preliminary hypothesis of the effect and source of heterogeneity.

The treatment effect on CCEI-combined were close to significant, and a possible explanation for why it was not significant is that that the covariates included in the regression did not increase the explanatory power of the model and did not reduce the standard error of the coefficients. Including more sociodemographic background variables, which is associated with significant differences of CCEI scores (Choi, Kariv, Müller and Silverman 2011) could have made the results significant.

# 10. Summary

The main finding of this paper is that higher incentives induce more effort in decision making, which for some subgroups translate into more strictly rational decisions. However, the increased incentives do not affect the two measures of strict rationality individually. The average effect on treatment on strict rationality measured by the CCEI-combined is not statistically significant, but a larger number of subjects in high-stake treatment take both non-dominated and consistent decisions than subjects in low-stake treatment.

There is a significant positive treatment effect for extravert subjects and subjects with low-IQ. The treatment effect on introvert subjects suggests that stress may influence their decisions as they have a negative effect of higher incentives, although it is not statistically significant. The subjects with low-IQ take more time on their decisions in high-stake treatment and this increased effort translates into more strictly rational decisions. This supports a view that increased incentives increases effort which increases strict rationality, unless other factors influence the performance.

The effect on high and low-IQ subjects do not indicate that it is easier for high-IQ subjects to make rational decisions, but that they are willing to spend more effort for low incentives than low-IQ subjects as the high-IQ subjects are spending more time and are taking more strictly rational decisions in the low-stake treatment. It does thus not support that high-IQ subjects have a better ability to make rational decisions.

The data supports a learning theory, in which the subjects take less dominated decisions in the middle decisions than in the first decisions (Binmore 1992). However, the data does not support a hypothesis that the high-stake subjects have a higher incentive to learn as the treatment effect.

For future research it could be interesting to go more into the choice pattern of each individual and study the presence of systematic errors or occasional errors which could shed more light on the difference between high- and low-stake subjects.

# References

Afriat, S. N. (1967) “The construction of a utility function from expenditure data” *Econometrica*, *6*, pp. 67-77.

Afriat S. N.  (1972) “Efficiency Estimation of production functions” *International Economic Review* 13(3). pp. 568-598.

Ameriks, J., Caplin, A. and Leahy J. (2003) “Wealth accumulation and the propensity to plan” *Quartely Journal of Economics*, *118*, pp. 1007-1047.

Angrist, J. D. and Pischke J.-S. “Mostly Harmless Econometrics: An empiricists companion” [online] Available at: http://xa.yimg.com/kq/groups/1104317/1239426945/name/mostly+harmeless+econometrics.pdf [Accessed: 25.05.2012]

Arierly, D. (2008) “Predictably irrational – the hidden forces that shape our decisions”, USA: HarperCollins

Arierly, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009) “Large stakes and big mistakes” Review of Economic Studies 76, pp. 451–469

Baron J. (2000) “Thinking and deciding” 3. edition, USA: Cambrigde University Press, pp. 53-66 and 219-274.

Baumeister, R. F. (1984) “Choking under pressure: Self-consciousness and paradoxical effects of incentives on skilful performance” *Journal of Personality and Social Psychology*¸ Vol 46, No. 3, pp. 610-620.

Binmore, K. (1992) Fun and Games, Lexington MA: Heath

Camerer, C. F. (1995) “Individual decision making” In *Handbook of experimental economics*, ed. Kagel J and Roth. A. Princeton University Press: Princeton.

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

Choi S., Fishman, R., Gale, D. and Kariv, S. (2007a) "Revealing preferences graphically: an old method gets a new tool kit" *American Economic Review, Papers & Proceedings, 97, pp.* 153-158.

Choi S., Fishman, R., Gale, D. and Kariv, S. (2007b) "Consistency and heterogeneity of individual behavior under uncertainty" *American Economic Review,* 115, pp 1020-1048.

Choi S., Kariv, S., Müller, W. and Silverman D. (2011) "Who is (more) rational? *Working paper*

Dandy, J., Brewer, N. and Tottman, R. (2001) "Self-consciousness and performance decrements within a sporting context" *Journal of Social Psychology,* CXLI, pp. 150-152.

Deaton A. (2010) "Instruments, Randomization and Learning about Development" *Journal of Economic Literature, 49,* pp. 424-455.

Eysenck, M.W. (2009) "Fundamentals of Psychology", Hove: Psychology Press

Frey, B. S. and Jegen, R. (2001) "Motivation crowding theory" *Journal of Economic Surveys,* XV, pp. 589-611

Gneezy, U. and Rustichini, A. (2000) "Pay enough or don't pay at all" *The quarterly Journal of Economics,* CXV, pp. 791-810.

Gneezy, U. and Croson, R. (2009) "Gender differences in preferences" *Journal of economic literature* 47(2), pp. 448-474.

Gul, F. (1991) "A theory of disappointment in decision making under uncertainty" *Econometrica* 59(3) pp. 667-686.

Heckman, J. J. (2000) "Causal parameters and policy analysis in economics: A twentieth century retrospective" *Quarterly Journal of Economics,* 115(1), pp. 45-97.

Houtman, M. and Maks J.A.H. (1985) "Determining all maximal data subsets consistent with revealed preference" Kwantitatieve Methoden 19, pp. 89-104

Kahneman (2011) "Thinking fast and slow" New York: Farrar, Straus and Giroux.

Kahneman D. and Tversky, A. (1979) “Prospect theory: an analysis of decisions under risk” *Econometrica* 47(2) pp. 263-291.

Kahneman, D., Peavler, W. S. and Onuska L. (1968) “Effects of verbalisation and incentive on the pupil response to mental activity” *Canadian Journal of Psychology* 22, pp. 186-196.

Knight, F.H. (1921) “Risk, uncertainty and Profit” Houghton Mifflin Company: New York.

Kőszegi B, and Rabin, M. (2006) “A model of reference-dependent preferences” *The quarterly journal of economics* Vol CXXI issue 4, pp. 1133-1165.

Libby, R. and Lipe, M.G. (1992) “Incentives, effort and the cognitive process involved in accounting-related judgments” *Journal of Accounting Research* 30, pp. 249-273

Mas-Colell, A., Whinston, M.D. and Green, J.R. (1995) “Microeconomic theory”, Oxford: Oxford University Press, chapter 6.

Odean (1998) “Are investors reluctant to realize their losses” *Journal of finance* 53 (3) pp. 1775-1798

Pindyck, R. and Rubinfeld, D. (1995) “Microeconomics” 3rd ed., Prentice-Hall Inc: New Jersey

Read, D. (2009) “Experimental test of rationality” In: Anand P., P.K. Pattanaik, C. Puppe eds. *The handbook of rational & social choice*, (2009) USA: Oxford University Press, ch. 8.

Rosenthal, R. and Rosnow, R.L. (2009) “Artefacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow’s Classic Books”, 1st edition, Oxford University Press

Samuelson, W.F. and Bazerman, M. H. (1985) “The Winners Curse in bilateral negotiation” *Research in experimental economics 3,* pp 105-137

Savage, L. J. (1954) *The foundations of statistics* New York: John Wiley.

Sen, A (1987) “Rational behaviour," *The New Palgrave: A Dictionary of Economics*, v. 3, pp. 68-76

Sen, A. (1993) “Internal Consistency of Choice, *Econometrica*, 61(3), pp. 595-521.

Shefrin, H. and Statman, M. (1985) "The disposition to sell winners too early and to ride loosers too long: Theory and evidence" *Journal of finance 40 (3)*, pp. 777-790

Stanovich, Keith E. (2009) "Rational and Irrational Thought: The thinking that IQ test miss" *Scientific American* [online] Available at: http://www.scientificamerican.com/article.cfm?id=rational-and-irrational-thought [Accessed 16. May 2012]

Starmer, C. (2000) "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature* 38(2), pp. 332-382.

Stolberg, H. O, Norman G and Trop, I. (2004) "Randomised controlled trials" *American Journal of Roentgenology* [online], Available at: http://www.ajronline.org/content/183/6/1539.full#sec-2 [Accessed 13.may.2012]

Tversky, A. and Kahneman D. (1986) "Rational Choice and the framing of decisions" *The Journal of Business"* 59, no 4, pp S251-S278

Tversky, A. and Kahneman, D. (1974), "Judgement under Uncertainty: Heuristics and Biases" *Science* 185, pp. 1124-1131.

Varian, H. R. (1991) "Goodness-of-Fit for revealed preference tests" mimeo

Von Neumann. J and Morgenstern O. (1944) "Theory of games and economic behavior", 60th Anniversary ed. USA: Princeton University Press, pp. 15-45.

Woolridge J.M, (2009) Introductory econometrics – A modern approach. 4th ed. Canada: South-Western Cengage Learning.

Yerkes, R. M and Dodson J. D. (1908) "The relationship of strength of stimulus to rapidity of habit-formation" *Journal of comparative neurology of psychology,* XVIII, pp. 459-482.