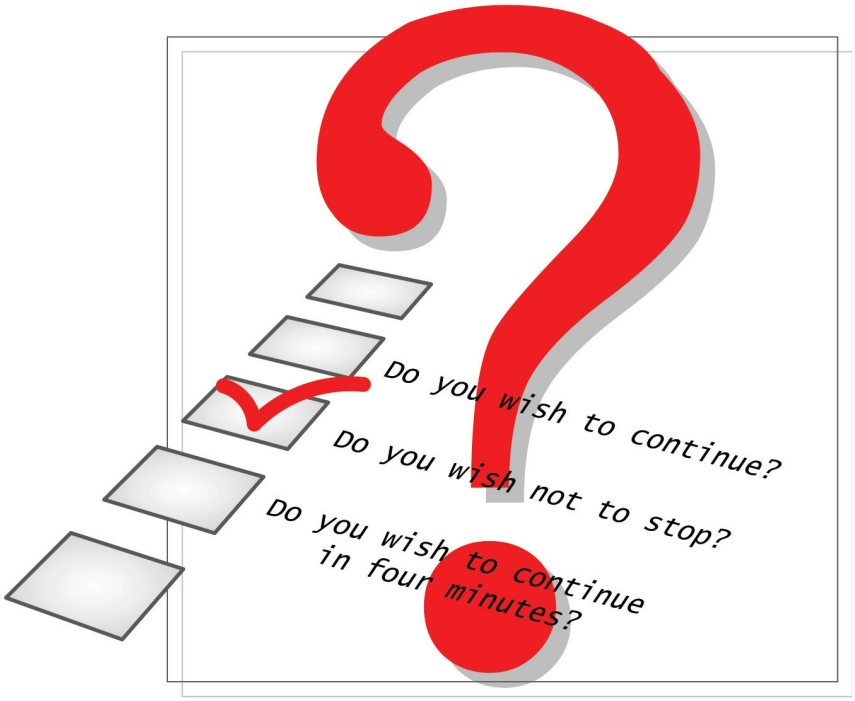
Decision Making in Use Contexts

Can User Decision Behaviour Be Influenced by Means of Sunk Time, Framing, and Default?



Produkt- og Designpsykologi
Group 1075, spring 2011
10th semester, Aalborg University

AALBORG UNIVERSITET

Det Ingeniør- Natur- og Sundheds-videnskabelige fakultet
Produkt- og Designpsykologi
Frederik Bajers vej 7
Phone: 96 35 86 00
E-mail: webinfo@es.aau.dk
http://es.aau.dk

Title:
Decision Making in Use Contexts - Can User Decision Behaviour Be Influenced by Means of Sunk Time, Framing, and Default?

Project period:
PDP10, Spring 2011

Project group:
1075

Participants:
Sigrid Boesen
Krestina Petersen Warming

Supervisor:
Johan Trettvik

Copies: 4
Number of pages: 206
Appendices hereof: 63
Attachment: 1 DVD
Finished 31.05.2011

Abstract:

This project is founded upon psychological research which suggests that humans make "irrational" decisions given certain circumstances. The scope is to examine if these observed decision patterns also apply in use situations. Three subjects within the field of decision making are examined; sunk time-, framing-, and default theory. Within each of the areas a series of use related experiments are set up in order to examine the subjects from different perspectives. The results from the sunk time experiments suggest that a sunk time effect is present in use situations, in this case mediated by waiting times. The framing effect is examined in terms of slide bar reference framing and colour framing however, no significant effects are found. The default effect is tested in system setup scenarios where test subjects are able to adjust- or keep default settings; the results imply that for unfamiliar scenarios the effect is not apparent, whereas the effect is emphasised in familiar scenarios. Despite not all results showing clear tendencies, it is emphasised that decision behaviour in use situations can be influenced by design means; hence such parameters can be used to affect users' behaviour and thus need to be accounted for in design for human beings.

Preface

This report is the result of the 10th semester project, written by group 1075 studying Engineering Psychology at Aalborg University under the department of Electronic Systems. The scope of the project is to examine the psychological implications of decision making, and apply findings in user interaction experiments and analyse the results with the purpose of examining whether or not human decision making is predictable in user interfaces.

All references in this paper are written according to the Harvard referencing method. An Appendix and Attachments are appended, together with a DVD containing an electronic edition of the report, together with supporting material such as developed test interfaces.

The project group would like to thank everyone who participated in the experiments.

Department of Electronic Systems
Aalborg University, 31th of May 2011

Krestina Petersen Warming

Sigrid Boesen

Contents

CHAPTER 1

Introduction

People make decisions all the time, but some decisions appear to be more irrational than others. For example, when paying for a buffet we are more likely to over-eat than if we just order an a la carte meal. If full, then why continue to eat? Sometimes we continue performing the same action over and over again while expecting a different outcome (which by Einstein is defined as madness!). We accept things for free we do not even like or need, and buy things just because they are on sale.
If we act and decide differently based on external influences such as discounted products and free merchandise, then how are we influenced when it comes to making decisions in use situations? Are we also influenced by external factors in user interfaces which alter our behaviour when deciding upon an action? Imagine if external influences such as the colour of a button or a default setting make us act differently than we would have if the colour and default attributes had been different. If external frames affect our decisions this could prove fatal in the setting of a nuclear power plant control room, a cockpit, car etc. as one "irrational" decision could potentially lead to disaster.
On a smaller scale; user interfaces which do not account for human decision making behaviour might as a consequence appear less user-friendly, since accounting for decision making behaviour holds the potential of being able to guide and to some extent manipulate user behaviour.

Much research suggests that humans do make decisions according to behavioural patterns ([Tversky and Kahneman, 1974], [Slovic and Tversky, 1974], [Tversky and Kahneman, 1979], etc.), which i.a. enables marketing practitioners to optimise their sales strategies and advertisement. If the observed behavioural decision patterns translate to a HCI domain this knowledge could be advantageous in relation to designing user interfaces, as behaviour could be accounted for to a further degree. A product designed without considerations of decision behaviour might unintentionally influence the user to make "irrational" decisions. By examining if users' decision making can be influenced in use situations a basis for avoiding influencing the user unintentionally can be made. Furthermore the results can be used to intentionally influence the users. The question to be resolved through this project is if in fact users can be influenced by the product they are interacting with to make "irrational" decisions.
The findings already made concerning decision making incongruence are mainly based on imaginary experimental setups, why it would be somewhat redundant to apply them directly in user interfaces if in fact decisions are made differently in a use context. Therefore, this project seeks to examine if the patterns already observed in classical decision making research, such as framing-, default-, and sunk time effects, seem to appear in HCI-based settings as well.

As the subject of decision making covers multiple interesting areas a limited set of three main areas are chosen for further examination in the current project. This however, has resulted in a total of two preliminary and nine main experiments including a total of 671 experimental observations, why the structure of the report can appear complex, thus an overview of the report will be provided to understand the underlying logic of the structure. The focus of the project will first be presented in a thesis statement, where after a general overview of the decision making foundation used in this project is provided. From this overview the three main areas of interest are derived, and will be examined theoretically in the problem analysis. From the theoretical analysis the subjects are then explored and

applied in experimental use settings, one subject at a time. Finally, the report will end with a discussion, conclusion and future works, where the implications of the findings are treated in relation to further applicability.

CHAPTER **2**

Thesis Statement

Decisions are made every day continuously why the subject of decision making has become widely examined in order to understand which processes and perhaps patterns underlie human behaviour ([Savage, 1954], [Tversky and Kahneman, 1974], [Slovic and Tversky, 1974], [Tversky and Kahneman, 1979], [Goldstein and Gigerenzer, 2002], etc.). Opposing views have been raised over the years concerning whether humans are rational or irrational decision makers. Especially within the field of economics normative models put forward by Neumann and Morgenstern (1947) and altered by Savage (1954) constitutes a basis for the conception that rational behaviour can be modeled and that human's act accordingly. The challenge of uncovering whether humans are rational decision makers or not has ever since been taken on by multiple researches with roots in psychology, with results suggesting that humans are not rational decision makers seen from a normative perspective ([Tversky and Kahneman, 1974], [Slovic and Tversky, 1974], [Tversky and Kahneman, 1979], [Johnson and Goldstein, 2003], [Hsee, 1998], etc.).

Normative decision making can be classified as the "ought to do", which Savage's axioms seek to express (Savage, 1954), whereas later research seeks to examine the "is doing"; creating descriptive theories of decision making. However, the descriptive revelations in classical decision making research have mainly been found on the basis of imaginary experimental setups, such as in extensive studies made by i.a. Arkes and Blumer (1985) on Sunk Cost and by Tversky and Kahneman (1981) concerning Framing.

It is with a descriptive aim this project will seek to examine decision behaviour in a less imaginary user interaction context. The context is based on situations where a user is faced with a decision in the interaction with an interface, such as a computer or a machine. If there are implications that humans act irrational in such use contexts, the findings may potentially be used to adjust for such incongruence's in expected behaviour; optimising how user interfaces can be constructed to best fit or best control use. A restricted set of decision phenomena will be examined in the present project, since the subject spans a wide range of phenomena which is too much to cover given the frames of the project. Furthermore, some of the observed phenomena are of more relevance than others in terms of user interaction situations. Thus, the initial problem states:

Can observed decision patterns be observed in user interaction scenarios, thus be used actively to predict and guide user behaviour?

The thesis problem will be sought answered through literature studies found in the following sections, and by practical implementation of the relevant findings. First, a general overview of the decision making subject is presented after which the term rationality is discussed, to form a basis for understanding the approach taken throughout the remaining of the project.

CHAPTER **3**

Theoretical Basis

The content of this chapter will provide a brief overview of the branch of the decision making research which is used as a basis for forming the present project. A short overview of the development within the field will be presented, which highlights opposing views in terms of decision making behaviour. Based on the introductory overview a discussion concerning the definitions of rational and irrational behaviour is made, which terminates with the view adopted on these definitions in the present project. Finally, as the subject of decision making spans widely a selection of sub subjects to examine further is made.

3.1 Decision Making

This section will first provide a brief overview of the evolution within the psychological field of decision making ranging to areas of economy and behaviour as well. Decision making is of great importance in many areas i.a. in the economic area. Decision patterns enable economists to model and predict behaviour of consumers and the market, why decision making models were mainly constructed based on a monetary basis. Following the overview a discussion of judgment and decision making concludes the section.

Brief Overview of Decision Making Research

One of the first known theories of individual choice in economics is Expected Utility Theory (EU) proposed by Daniel Bernoulli (1738). He proposed the model in response to the perception that people are willing to pay the same to enter a gamble as the expected value of the gamble. He illustrated his point by the St. Petersburg Paradox where the rules are simple; the contestant pay a fixed fee to enter the game, where a coin is flipped. For each head the coin provides in a row the contestant wins 2^n, where n is the number of succeeding heads. The question then is how much a reasonable individual should be prepared to pay to enter the gamble. [Bernoulli, 1738]
If the economic assessment, that people are willing to pay the same to enter a gamble as the expected value of a win, is true, then the expected value should equal the fee. By performing simple probability analysis the average expected value can be calculated. The first coin toss provides a probability of 1/2 to win 1 dollar, the second toss provides the probability of 1/4 to win 2 dollars, third toss 1/8 likelihood to win 4 dollars etc. Adding all these probabilities together provide an expected value of:

$$\sum_{k=1}^{\infty} \frac{1}{2} = \infty$$

The logic is then that whatever the fee is to enter the gamble one should take the gamble, since the expected win for the contestant diverges to an infinite amount of money (in the idealised form). However, Bernoulli proposed that entering such a gamble the utility of the expected win matters as to how much one would invest to enter the gamble. The utility of winning 1000 coins for a poor man is greater than for a rich man, based on this they should not pay the same to enter the game as their expected utility differs. Bernoulli did not test his assumption that an individual are not prepared to pay an infinite sum of money to enter the gamble, in his original paper. Nonetheless, it was not until 1947 that the idea of EU

was revived, when Neumann and Morgenstern applied it to game theory [Starmer, 2000]. Their analysis of Expected Utility Theory was applied to games of chance with computable and known probabilities. They showed that a set of appealing axioms could be derived from the Expected Utility hypothesis, which provided grounds for interpreting EU normatively (constituting how individuals ought to choose).

Savage (1954) added an extension to the work by Neumann and Morgenstern (1947), and presented it as Subjective Expected Utility theory (SEU) of choice under uncertainty [Savage, 1954]. The extension of EU allowed for subjective (personal) probabilities in situations were no mathematical probabilities are available. For example, if an individual is faced with the choice of either going to a park to play football with some friends or finishing an assignment, all parameters must be ascribed a value according to normative theory in order for the individual to make a rational decision. However, the value of playing football depends on the weather, and so does the value linked to writing an assignment. Furthermore other factors may play a role, such as who is playing football etc. All of these parameters cannot be quantified in general as they depend on individual perceptions of the different circumstances; hence the subjective aspect is added.

Normative Axioms

The normative axioms established based on the work of Neumann and Morgenstern (1947) and Savage (1954) define how we ought to make decisions, by maximising expected utility. Therefore, the conception was that if the axioms are violated the decision will not maximise expected utility, and will as a consequence be non-normative and thereby irrational. The main normative axioms are described below.

1. Transitivity
If A is preferred to B and B is preferred to C, then A should be preferred to C (choice ordering)

2. Comparability (or Completeness)
If presented to a choice of A and B the answer will always be one of (i) Prefer A to B (ii) Prefer B to A (iii) Indifferent between A and B

3. Dominance
An option (the dominant) which is at least as good as, and better, in one or more aspects than the remaining options (the dominated) should always be preferred.

4. Independence
The preference for one choice should not be changed by an outcome independent for the choice. For example if you have the choice between buying a red car or a blue car of the same type, and you prefer the red car, the preference should not be altered by the fact that you find out that your neighbour just bought a green car.

5. Invariance
The preference of an option should be independent of the description of the option.

Despite the axioms appearing straightforward they were questioned by multiple researchers within different fields, who questioned whether humans do act according to a set of mathematically described axioms claiming that if we are rational decision makers we should always follow these axioms. The present project is based on the observations made in the following.

Violation of Normative Axioms

Violation of the transitivity axiom was i.a. found by Tversky (1969) who asked test subjects to express which college applicant they preferred. The test subjects were presented with pairs of college applicants, e.g. A and B, then B and C etc., and were asked to state who they preferred of the two when weighing intelligence more than the two other parameters (see figure 3.1). The results showed that test subjects typically preferred A to B, B to C, C to D and D to E. However, when presented with the pair of A and E test subjects typically preferred E over A and were thereby violating transitivity.[Tversky, 1969]

Applicant	Intelligence	Emotional Stability	Social Facility
A	69	84	75
B	72	78	65
C	75	72	55
D	78	66	45
E	81	60	35

Figure 3.1: *Table of college applicant ratings based on three parameters shown to test subjects.*

As to explain this intransitivity pattern Tversky suggests that it is a way to render decision making less cognitively demanding. Instead of grasping all information he suggests that we use "decision rules". In the specific case of the college applicant assessment task, each pair is compared in terms of the parameters in turn, instead of gathering all parameter information to make an overall comparison. Furthermore, if two parameter ratings are similar the parameter is disregarded and other more distinguished parameters are used to base the decision upon.

Shafir (1993) showed how choosing one of two options is not the same as rejecting the other option, violating the comparability axiom. He made two descriptions of two holiday resorts (as below) and presented them to one test group, where the task was to choose which resort one would prefer. Another test group was shown the same descriptions, but was told that they should imagine holding reservations for both resorts, but now had to cancel one.

Resort A:

- Average weather
- Average beaches
- Medium quality hotel
- Medium-temperature water
- Average nightlife

Resort B:

- Lots of sunshine
- Gorgeous beaches and coral reefs
- Ultra-modern hotel
- Very cold water
- Very strong winds
- No nightlife

In the first test group 67% preferred resort A and 33% resort B. However, in the second test group where one had to be cancelled, 52% cancelled resort A and 48% resort B. When having to choose an option we tend to focus on the positive aspects, whereas having to deselect an option we are looking for negative features, hence A offers several good reason

compared to B for being selected, but also compared to B, A offers several good reasons for being rejected. Thereby, we do not have one clear preference for an option which then eliminates the other options; it is dependent on the circumstances.[Shafir, 1993]

The axiom exposed to most critique and following challenges is the Independence axiom. In 1953 Allais made a set of choices called the Allais Paradox, which are presented in figure 3.2. The idea is to present two lotteries (e.g. situation 1), where one of lotteries A and B can be chosen in which one of 100 tickets is drawn. The paradox occurs when situation 1 is first presented and a choice is made, followed by situation 2 where another choice needs to be made. Most people choose option A in the first situation, but option D in the second situation. This might appear obvious looking at the two situations together, however, the violation of the Independence axiom is easily revealed by placing a hand over the last column, which reveals that the two situations are identical.

		Lottery Ticket Numbers (1-100)		
		1	2-11	12-100
Situation 1	Choice A	$ 1,000,000	$ 1,000,000	$ 1,000,000
	Choice B	$ 0	$ 5,000,000	$ 1,000,000
Situation 2	Choice C	$ 1,000,000	$ 1,000,000	$ 0
	Choice D	$ 0	$ 5,000,000	$ 0

Figure 3.2: *The Allais paradox presented as a choice between different lotteries.*

According to the axiom of Independence, the information in the last column should not influence the choice, since it is identical for A and B in situation 1 ($ 1,000,000) and for C and D in situation 2 ($ 0). Therefore, preferring A to B should likewise lead to a preference of C over D, given the axiom. [Allais, 1953]

With regards to the Invariance axiom Tversky and Kahneman (1981) made a famous study called the Asian disease problem. From their research they showed how the description of a set of choices (framing) can affect the decisions made. Two test groups were told that the U.S. was preparing for an outbreak of the unusual Asian disease, which was expected to kill 600 people. One test group was presented with options A and B, whereas the other test group was presented with options C and D.

A. *If program A is adopted, 200 people will be saved.*
B. *If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds chance that no people will be saved.*

C. *If program C is adopted, 400 people will die.*
D. *If program D is adopted, there is a one-third probability nobody will die and a two-thirds chance that 600 people will die.*

For the first setup, 72% of the test subjects chose A and 28% chose B. Saving 200 lives for certain is perceived as more attractive than a one-third chance of saving 600 lives. In the second setup, only 22% of the test subjects chose option C while 78% chose option D despite them being equivalent to option A and B, respectively. These results contradict the Invariance axiom which stated that preference of an option should be independent of the description of the option. [Tversky and Kahneman, 1981]

Tversky et al (1988) found an effect they called the "prominence effect" which shows that attributes weigh more heavily in a choice task than in a matching task. They made an

experiment where test subjects were told that a transport ministry in a country where 600 people died from traffic each year considered two programs for dealing with the problem. On test group were given the first of the two following scenarios, whereas the other group were presented with the second scenario.

Option A: *Reduce to 570 casualties at the cost of $12 million*
Option B: *Reduce to 500 casualties at the cost of $55 million*

Option C: *Reduce to 570 casualties at the cost of $12 million*
Option D: *Reduce to 500 casualties at the cost of $?*

In the first scenario 67% preferred option B, "only" saving 70 lives more at an extra cost of $43 million. However, in the second scenario where they were asked how much they would pay to save 70 extra lives, only 4% of the second test group provided a number above $55 million, with the rest of the suggestions way below. From a "rational" perspective 67% of the test subjects should have presented a value above $55 million. It is argued that in regards to making choices people select the superior option in terms of the most important attribute (in this case lives saved), a qualitative reasoning. When it comes to matching it is more cognitively demanding and entails a more quantitative assessment, which cannot be performed without paying attention to values of both attributes (lives and cost) and their relative importance. [Tversky, Sattath and Slovic, 1988]

Other examples of effects where anomalies in choice appear and people act "irrational" in relation to the normative theory are:

The compromise effect, where people's choice change from one original option to another original option by adding an extra alternative option [Kivetz, Netzer and Srinivasan, 2004]. This effect violates the independence notion.

The too much choice effect, where people are impaired in decision making when exposed to multiple options, as well as their performance based on a choice between many options correlates with number of choices [Iyengar and Lepper, 2000]. In relation to the assumption of us being rational decision makers who seek to maximise our utility more choices should allow us to make an even better decision, as more options increase the likelihood that an options matching our preferences will appear. However, more options in relation to fewer make us more likely to decide against choosing.

Affective forecasting concerns decisions we make based on a prediction founded on a (mis) perception, which is known as the focusing illusion. It has been found that we overestimate future happiness and pain based on which we judge differently than we otherwise would have. For example research by [Blumenthal, 2009] suggests that jurors who are exposed to statements of pain victim are more likely to death sentence the offender. It seems that we are not able to account for later adaption, unless made aware of it.

Another effect which defies the normative logic is that of choosing according to a default. In choices where a default is present we seem more prone just to go with the already chosen, which in turn showcases preference reversals, which defies i.a. transitivity. [Johnson and Goldstein, 2003]

The endowment effect is also a representative for the irrational choices we make, as this effect manifest itself in a higher valuing of possessed items than items where a property right as not been established [Kahneman, Knetsch and Thaler, 1991]. This effect is also linked to the default effect.

Emotions also act as generators for irrational behaviour, for example regret. This particular feeling can affect whether a choice is made or not, and be a motivator for choosing certain options. In a vaccination setup made by [Ritov and Baron, 2002] test subjects' decision was affected by whether they were told the result of a risky choice they had to make or not - whether they would get the chance to regret or not. Regret was also shown to function as a motivation for choosing more expensive products when test subjects were asked to consider regret [Simonson and Tversky, 1992].

The mentioned anomalies in choice, which defies rational decision making according to the normative theory, are interesting as they appear almost systematic. If the same type of decision patterns can be found in use cases, then it will enable UI designers to design more directly for a certain wanted behaviour in choice.
The anomalies in choice observed this far contradicts SEU, why a new model for decision making under risk was developed.

Prospect Theory

The normative model of expected utility dictates how we should make decisions, but does not seem to describe how we actually do make decisions. Kahneman and Tversky (1979) put forward such a descriptive model for decision making under risk, called Prospect Theory. This theory explains many of the observed phenomena that SEU cannot account for. Opposite SEU, Prospect Theory does not dictate ideal decision making, but is a mean to describe the observed discrepancies with SEU.

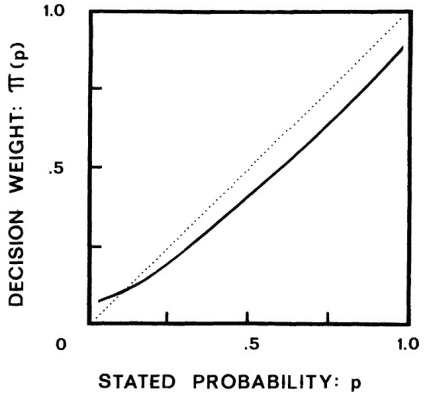


Figure 3.3: *The weighting function of Prospect Theory [Tversky and Kahneman, 1979].*

Prospect theory assumes two phases of the choice process. The first is the editing phase where the decision problem is recognised and insignificant components might be discarded. In this phase a reference point is as well established in order to enable decision options to be classified as either "losses" or "gains". Furthermore, probabilities supplied in this phase are distorted by the individual, see figure 3.3. The graph shows that low probabilities are over-weighted and moderate and high probabilities are under-weighted. This means that posed with a probability of 1% it will be weighted to have disproportionately more influence than 0% (which lies on the dotted line). A probability of 99% would contrarily be weighted to have disproportionately less influence than 100% (which also lies on the dotted line). The weighting function then accounts for what was observed in the Allais paradox concerning

over-weighting small probabilities.
In the second phase the identified prospects are evaluated by the attitude towards risk in terms of gains and losses.

Prospect Theory suggests that the evaluation of a given set of decision outcomes in terms of gains and losses is based on a neutral point of reference. In figure 3.4 the value function of Prospect Theory is shown. The horisontal axis represents an objective measure (e.g. monetary), with gains on the right side of origo and losses on the left side. The vertical axis represents the value of the psychological gain or loss. The function on the right side is steep right at the beginning, but with a diminishing slope, which means that the difference between \$0 and \$10 is greater than that between \$100 and \$110, psychologically. Likewise, the slope on the left side diminishes in a similar way.

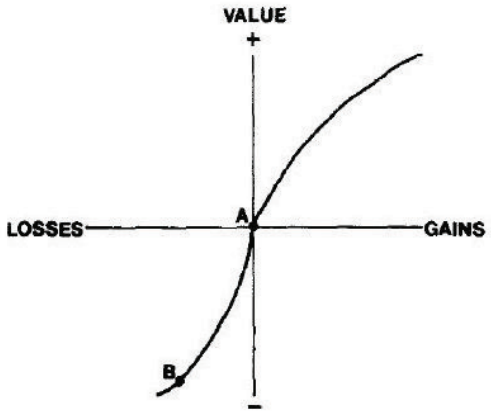


Figure 3.4: *Value function of Prospect Theory [Tversky and Kahneman, 1979].*

Depending on which position on the curve one estimates oneself to be, the attitude towards risk varies. Being on the right side of the curve makes one risk averse, since risking a gain of e.g. \$10 in order to win another \$10 is not proportional in psychological value - losing the \$10 will loom larger than gaining another \$10. Opposite, being on the left side of the curve makes one risk seeking, since a sure loss is guaranteed unless a gamble with a chance to gain an amount is taken, despite the chance of losing even more. Furthermore, the curve for losses is steeper than for gains, which models how people feel losses more than gains of equivalent value. [Tversky and Kahneman, 1979]

Prospect Theory can account for a broad range of observed anomalies in choice in experimental settings as well as in observatory field studies. For example, the Asian disease problem setup by Tversky and Kahneman (1981) in which the framing of the choice options acted as a critical influencer on the choices made by test subjects (see section 3.1). When the options were framed in terms of losses test subjects became more willing to choose a risky option, whereas gain framed test subjects exhibited a risk averse behaviour and chose the certain option. Relating to the value function, test subjects who were loss framed were provided with the reference point that nobody has died yet, so they have everything to win. The gain framed test subjects were in contrary provided with a reference point of what will happen if nothing is done; 600 casualties, why a certain rescue of 200 lives seems more attractive than the gamble of all 600 lives.

Heuristics

Kahneman and Tversky (1974) claim that the underlying processes of human judgment do not comply with the normative theories. They propose that due to human judgment being overconfident, inducing illusory correlations, being influenced by random anchors, ignoring base-rates and misinterpreting randomness, we use rules of thumb instead of underlying normative processes in judgment and decision situations.

The rule of thumb thesis encompasses that we seem to simplify the decision we are faced with and apply a fitting heuristic. The logic is that by reducing the complexity of a decision task it is quicker and less cognitively demanding to judge the premises and make a decision. Multiple heuristics have been identified i.a. the recognition heuristic, which was shown to violate normative theories by Goldstein and Gigerenzer (2002).
Goldstein and Gigerenzer (2002) argue that under certain conditions the recognition heuristic actually leads to a less-is-more effect, where knowing less about a subject improves one's judgment. The heuristic is most successfully applied in situations where only one parameter from a set of parameters is recognised, and in general where knowledge is limited. Furthermore, it is attributed by a binary all-or-none distinction, which means that either the object is recognised or not.
In order to demonstrate how the recognition heuristic can be utilised successfully given limited knowledge, Goldstein and Gigerenzer (2002) set up a series of experiments and simulations where large cities in different countries were the object. They showed that American students were no better at choosing which American cities are larger from pairs of American cities, than German cities from pairs of German cities, despite not knowing much about Germany. The students mainly chose the German cities they were able to recognise, whereby they had a high success rate.
As argued in the study, the recognition heuristic is a clever tool for us to use, since cities (or other attributes) we are able to recognise must also be the significant ones, and in this case thereby the larger ones. In relation to a normative point of view, the more information available to us, the more rational decisions we are able to make. However, this simple heuristic proves the opposite.[Goldstein and Gigerenzer, 2002]

Other established heuristics are i.a. the availability heuristic which is based on recall, the familiarity heuristic based on experience, the representativeness heuristic which relies to some extent on stereotypes often shown with description experiments ([Tversky and Kahneman, 1973]), the anchor and adjust heuristic where people make estimates based on an initial value, etc. Common to all of these different heuristics is that they are simple rules we seem to apply in given decision scenarios in order to diminish the problem at hand to a cognitively manageable problem. Furthermore, they do not necessarily follow the normative logic.

Judgment and Decision Making

When deciding a preceding analysis or judgment forms the ground for making a decision. The judgment is formed by the surrounding circumstances which all affects the judgment actively, such as the context in which the decision is made. Social contexts can make us decide differently from what we would do if alone as the context might provide more possibilities towards actions, or because it is judged that it is easier to do what the rest of the group does etc. The context in which a decision is made is also often affected by previous events and prospective events, which is taken into account in the judgment process. Also affecting judgments and thereby decisions are the experiences, gender, age, country of origin, religion beliefs etc. of each individual, which in turn make judgment and decisions an integrated process based on different premises for each individual.

The decision making research of i.a. Tversky and Kahneman does not truly encompass the rich contextual inputs provided in real life situations, which are often taken into account during the judgment process. Some of the circumstances however, will play a role in the type of decision making experiments just reported on, for example age, individual experience, gender etc. As the contextual information provided to test subjects in classical decision making experiments is very limited, it provides very concise results which show consistent behaviour patterns. Providing more complex input information for judgment in terms of user interface frames, will test if the behaviour in choice observed apply in other more complex domains.
Further classical decision making experiments will therefore be examined in the following section, in order to apply their findings in user interaction based settings, to reveal if they are transferable or not, and in turn whether judgment and decision making can be predicted in the same way as in non-context rich decision experiments. Thus, as already established in the thesis statement in chapter 2 the issue of interest for this project is the observed decision making patterns. Hence, the underlying process involved in the judgment part is disregarded to the extent that the input which is judged and decided upon will be examined and sought replicated in other settings than those traditionally used.

Summary

Many views upon decision making have been put forward and tested during the past decade. Opposing views in terms of which type of decisions we ought to make, and which type of decisions we actually make are present, and to some degree still debated. Much suggests that the SEU theory is not good at explaining actual behaviour; however, it can still be used as a guideline for rational decision making. Instead, Prospect Theory seems to be able to explain most anomalies observed in relation to SEU. The cross field of economics and psychology has become more narrow because of the emergence of the field of behavioural economics, where behaviour is sought mapped instead of debated in terms of rationality or not. As to decision making in human computer interaction scenarios the anomalies found may apply, despite of them being observed based on less complex settings.

Provided with an overview of the field of decision making research, and what it encompasses, the following section will discuss the implications of classifying a decision as either rational or irrational as normative theory assumes. A stance will be taken for the remaining of the project concerning the definition of a rational versus an irrational decision.

3.2 Rationality - a Definition

The basis of the present project is decision making and how people seem to make irrational decisions given certain circumstances. However, this section will be used to discuss historical implications of irrationality versus rationality, what circumstances surround the rational decision, if there is such thing as an irrational decision, and what view is adopted in the present project concerning rationality.

The debate concerning decision making and whether it is rational or irrational has existed for many centuries. One of the early Greek philosophers, Plato, argued that in order to make rational decisions, one had to be able to control ones feelings and emotions [Jowett, n.d.]. By this a distinction of rational and irrational choice was made already. The distinction between rational and irrational decisions still exists, though it has been recognised that feelings and emotions play a vital role in decision making. This was i.a. found in research made by Antonio Damasio (1996) who put forward the "somatic marker hypothesis" based on his research with ventromedial patients (patients with lesions in the ventromedial prefrontal cortex in the brain, which is associated with pure emotion)[Zimmerman, 1994] [Damasio, 1996]. One of Damasio's patients had had surgery near the frontal lobe due to a small tumour. Despite the surgery his IQ, learning-, language-, attention- and memory abilities were not affected, nor his ability to make estimates and his social knowledge. Overall he seemed normally capable of all tasks he previously had been capable of, besides from one - decision making. He was able to make a decision, however a simple task of choosing between two appointment dates for example took him half an hour, because he was weighing pros and cons of each date in order for him to finally make a decision. Especially decisions involving personal or social matters were hard for him to make.
According to the dualistic view of rational and irrational decisions put forward by Plato, the patients studied by Damasio should be rational decision makers, due to the fact that they are emotionally impaired. According to the findings of Damasio, a good decision maker is not a person without feelings. He might be able to make a good decision after weighting all pros and cons, but given the circumstances and the surrounding society, it is not optimal to spend minutes maybe hours to make a decision. Many of the patients studied by Damasio could not function together with their families after their respective accidents or surgery. Furthermore it could be argued that emotions should be influencing our decision making, since decisions made might later impact our emotional state.

In a study made by Schwartz et al (2002) test subjects were divided into two groups of "maxmisers" or "satisficers", according to whether they always wanted to make the best possible decision or would be satisfied with a "good enough" decision. The findings showed that test subjects categorised as maximisers did make better decisions in an ultimatum game, but were less happy with those decisions than test subjects categorised as satisficers. The maximisers seem to be controlled by a feeling of wanting to avoid regret, which was reinforced by the test setup where they were told how much they could have gained [Schwartz, Ward, Lyubomirsky, Monteresso, White and Lehman, 2002].

This study shows that making a good decision, might not lead to a positive feeling, which in turn makes it ambiguous to categorise a decision as good or bad. Human nature and social conventions seem to play a role in the perception of a good decision - a rational decision. There might not be a need for us to always make the best possible decision, since we can be happier with a less good decision.

In order to characterise a decision as being irrational similarly there is a need to be able to characterise a rational decision. The latter has been described by normative rules stated by Savage (1954), where the normative axioms need to be followed for a decision to maximise subjective expected utility and thereby be rational (see section 3.1). For example we need to

obey the rule of transitivity, which states that if one prefers option A to option B and option B to option C one should also prefer option A to option C. However, this particular normative rule together with the rest have all been disproved in several studies ([Ellsberg, 1961], [Tversky, 1969], [Slovic and Tversky, 1974], [Tversky and Kahneman, 1979], [Tversky and Kahneman, 1986], etc.) Results obtained from these studies make us appear as irrational beings in relation to the statistic measures put forward by Savage (1974) and economists, who state that we act rational in accordance with a predictable Bayesian prior.

If the normative rules are set up as being the "truth" as to rational decision making, decisions which do not follow these rules will as a consequence appear irrational. One opponent to this view is Gerd Gigerenzer who in his book "Rationality for mortals - How people cope with uncertainty" argues that to be is to error by rewriting the famous citation of René Descartes "I think, therefore I am" to "I think, therefore I err". With this simple rephrasing he argues that only by making "errors" humans can evolve their abilities. One of the main objections to the irrationality notion he presents is the circumstances under which this type of behaviour is observed. Instead of ecological studies where the further consequences of a choice are examined, only the immediate choice in isolation is analysed. Also, the circumstances set up for a choice in classic decision experiments i.a. set up by Kahneman and Tversky are artificial and constructed. Gigerenzer classifies these types of experiments as Type III errors; where the task is to find the right answer to the wrong question. Instead he argues that what is conventionally observed as errors is in fact what makes humans more intelligent than any mathematical prediction system. For example when we learn as children, we make errors based on bets instead of being paralysed by uncertainty. Instead of "playing safe" we make errors and learn from our mistakes in order to evolve. In this way Gigerenzer redefines irrationality from not maximising expected utility to actually gaining a larger utility. [Gigerenzer, 2008]

"Irrationality" also comes in other forms such as the form commonly denoted a *gut feeling*. This particular feeling can make humans make decisions which were not anticipated. One example is that studied by Gary Klein where a Lieutenant Commander, Michael Riley, shot down a missile not knowing what he shot down. He had spotted the missile on the radar, however, the blips made by the missile on the radar resembled that of an American aircraft. Nonetheless, he used only 90 seconds to determine by gut feeling that he had to shoot it down in order to survive, despite there being a significant risk for the blip to be an allied aircraft. In an interview Riley revealed that after the first five seconds he knew it was a missile, however, he was not capable of explaining why. Expert investigation of the blips afterwards showed no signs which should differentiate the radar blip from an allied aircraft radar blip. First upon thorough scrutiny it was found that the blip encountered a little later than it would have, had it been an aircraft [Klein, 2009]. This example shows how gut feeling can make humans act in an unreasonable manner, since for the specific situation Riley should have acknowledged the blip as an aircraft given the time slot he was operating within. The incident experienced by Riley, and this feeling referred to as a gut feeling is most likely an effect of a fine tuning of Riley's perceptual system for this specific subject, based on the experience he has gained during his training and work. By fine tuning is meant that his perceptual system has become more fine grained as a consequence of his experiences, similarly as a person becoming blind evolves a more fine tuned audio perception system. Decisions made based on subtle perceived differences are then in risk of appearing irrational to spectators whose perceptual system is not geared for the situation, or a certain "magic" quality is ascribed to the event.

This example emphasises that the rational decision versus the irrational decision is a complex matter. If the decision setup is constructed it is easy to setup rules for a correct and incorrect answer. However, applying these to a real life decision situation such as the one Riley faced implicates a far greater complexity which is not easily replicated in an experi-

mental setting. Therefore decisions are constructed and diminished in complexity in order to be measured, which in turn makes it easier to judge the choice as either rational or irrational.

Most decision experiments are constructed such as framing experiments. A well-known technique is loss and gain framing as seen in experiments made by i.a. Kahneman and Tversky (1986), where it is shown that test subjects choose the opposite when exposed to the different conditions despite the choice being the same. Gigerenzer (2008) argues that framing experiments triggers test subjects to search for additional information which can help them to guess or bet on what is meant. He mentions the example that more people are willing to receive a treatment if the doctor explain that 90% survive than if he explains that 10% dies. According to the normative theories this is considered irrational, whereas Gigerenzer argues that it is in fact intelligence that makes us decide the way we do. Whereas a computer system would look at the fact, humans tend to seek surplus information which in the constructed scenarios might be perceived as irrational, but in real life settings make us better decision makers in complex situations since we have the capacity to "read between the lines". [Tversky and Kahneman, 1986] [Gigerenzer, 2008]

In relation Ariely (2009) argues that humans do not know what they want why comparison becomes a vital component in our decision making, which is why framing experiments seem to fool us most of the times. When a reference point is given, we use that as a reference for judging the options provided to us. This has been shown by i.a. Hsee (1999) and Bertini et al (2008), where providing test subjects with different reference points in terms of a different size ice cream cup or an extra product, changed their preferences. The point of reference that Ariely refers to seem to be the surplus information Gigerenzer refers to, which differentiates us from pure numeric logic systems.[Ariely, 2009] [Hsee, 1998] [Bertini, Ofek and Ariely, 2008]

Contrarily, in a study concerning use of metaphors by Thibodeau and Boroditsky (2011) results showed that test subjects were affected in their judgment of a given situation based on only one word. Two groups were given one article each. The articles' content was the same; however one word was changed between the two. In one article crime was described as a virus and in the other as a beast. After reading the articles the test subjects were asked what they would suggest in order to overcome crime. The virus primed test group primarily answered in terms of a "crime-reform programme" with social prevention reforms as a tool, whereas the beast primed group primarily answered in terms of a "law enforcement programme" with enforcement and punishment as a tool, despite having read the same surrounding article. Furthermore, when asked, test subjects did not point at the metaphors (virus/beast) as a basis for their answer, showing that they were oblivious to what made them choose as they did. [Thibodeau and Boroditsky, 2011]

This study shows that in some cases we do act and decide based on minor cues we pick up on subconsciously, which thereby becomes the determining factor for our opinions. From this view it seems more eligible to declare us as irrational beings since a word can alter our opinion radically. On the other hand had it been an experienced police officer who had been asked to assess the situation after reading one of the articles, his insights and experience on the area would most likely have influenced his answer, making knowledge and experience yet another factor in the decision making process.

As this section has shed light on, many factors seem to play a role in judgment and decision making, and even more important; what a rational decision is. Based on the previous it is concluded that the rational decision is relative. Setting up experiments within narrow frames makes it rather easy to judge a given decision as rational or not, but in real life settings there is much more to decisions. For example, what makes one person happy may not apply for another person, which was also found by Schwartz et al (2002). For some

people the best is never good enough, whereas for others the sufficient choice is the best choice, which makes the rational versus irrational decision very subjective.

Adding to the discussion, why should people be interested in making irrational decisions? As the maximiser/satisficer study showed what is rational is different for each individual. For the satisficer it may seem irrational to always strive for the best, when you can settle for "good enough", maybe even at a lower expense, why it is rational to them. We seem to strive for rational decisions, which in retrospect can be seen as irrational decisions, but based on the cues and circumstances surrounding the decision at the given time, why should people not strive to make the decision that seems more rational to them in the current moment. All of the above mentioned arguments enforces the notion that decisions is a subjective matter, which is hard to scale, since multiple contextual factors also play an obvious or latent role, which makes it hard to account for all influences in deciding upon the optimal, rational decision in real life.
For these reasons it is impossible to define the terms of a rational or good decision for a user in a use situation. Different use situations provide different contexts for decisions, such as the utility of a decision can potentially be ever changing. The stance taken in this project regarding the definition of decisions as either rational or irrational is therefore that a decision cannot conclusively be declared as rational or irrational. Instead of seeking an answer towards what decisions maximise utility in different use settings, the project seeks to examine the observations made regarding decision making in narrow framed experiments and see if they can be applied in more use relevant situations. If the decision patterns are transferable it is not a question of rational behaviour or not, it is a question of how to apply these prospective tools in order to support and precipitate certain use behaviour. Therefore, the mentioning of irrational or rational decisions in the consecutive sections will relate to decisions within the narrow frames of an experimental setting.

3.3 Demarcation

The subject of decision making spans a broad range of different subjects presented briefly this far (see section 3.1). In order to narrow down the subjects to fit the scope of the present project, three sub areas which are assessed to hold potential for being applied in use settings are chosen. Each of these three subjects will be examined and sought applied in use settings, with the object of observing if the choice patterns discovered in the literature also applies in use situations.
Some of the choice anomalie effects mentioned in section 3.1 are not as obvious to examine further as others, given the scope of the project. For example the subject of affective forecasting. Even though this effect makes us mis-predict future circumstances and thereby change our decision patterns, this effect is very closely linked to emotional states and prospective personal circumstances, why it is unsuitable for the focus of this project in relation to exploring the default effect for example. The latter is more obvious to implement in a use setting, as products always come with some sort of default setting, based on which it could be interesting to e.g. examine if use behaviour is affected by these.

Based on such considerations three sub areas of interest have been chosen for further examination, with the aim of applying these in a use based context. The three subjects are:

Sunk cost
Investing an amount of money and adhering in the present to the amount spent in the past constitutes an act of irrational behaviour in regards to SEU. However, this persistence might possibly be applied in a use setting as well in terms of investing time and effort into a product.

Framing
Framing a set of options is capable of changing our preferences of choice, despite of the options being the same. The interesting perspective with this subject is whether it can be transferred to a user interface, and if the effect can be observed.

Defaults
Defaults seem to impair us in terms of making a decision, if one of two options is a default we are more prone to choose that one, however, if it was the other option which was the default we would choose that one. This also appears irrational in terms of SEU, and if this applies in a use situation it can be utilised actively to guide user behaviour.

In the following all of these subjects will be examined more in depth in order to understand the subjects and the appertaining behavioural patterns. On the basis of that it is sought to examine if people act according to the same patterns in use settings, whereby it will be possible to design user interfaces proactively for behaviour.

CHAPTER 4

Problem Analysis

This chapter contains the problem analysis which entails a theoretical analysis of the sub subjects chosen in the previous chapter; sunk cost, framing and defaults. From this analysis the theoretic ground is laid in order to extract relevant information to implement in a subsequent empiric experimental process.

4.1 Sunk Cost, Time and Effort

It appears that previous investments can influence current choices, which is known as the sunk cost effect. This section provides an overview of the research done within the field of sunk cost, time and effort together with notions of application perspectives.

Sunk Cost

The sunk cost effect is a phenomenon which influences many decisions ranging from great leader decisions like the continuing of a war to small everyday decisions like finishing a dessert paid for despite being full. The phenomenon manifests itself in a feeling of waste when something is paid for and subsequently not used. An example is the case where continuous investments are made in failed projects to avoid having wasted previous investments, when in reality a lot more money is being wasted by proceeding with investing. The "irrationality" of sunk cost lies within making a decision one would not have made otherwise, due to the circumstances of having previously invested money, time and effort.

Sunk cost effect can be explained by the value function of Prospect Theory earlier described in section 3.1.

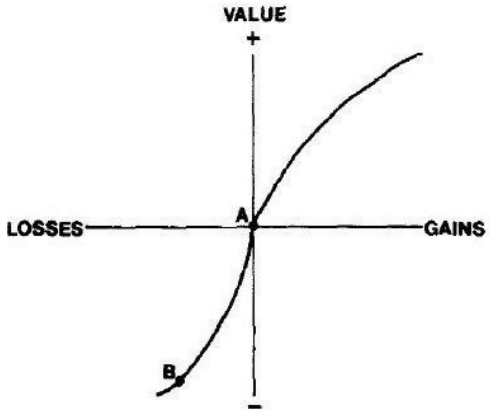


Figure 4.1: *The value function of Prospect Theory [Tversky and Kahneman, 1979].*

Having had a loss equivalent to point B in figure 4.1 a further loss will be valued less negative as if making the same loss at point A. Therefore, having invested an amount of money in a project which is not creating profit would situate the investor at point B, as he has obtained a loss. The willingness to invest further money will then be greater, since investing more

money will appear less mentally "painful" than admitting that the first investment was wasted.

Sunk cost has been studied and examined by Arkes and Blumer (1985) via several experiments. They argue that the effect is applicable for money, effort and time and that the behaviour behind the phenomenon is based on a human desire to avoid appearing wasteful. [Arkes and Blumer, 1985]
The majority of experiments made by Arkes and Blumer (1985) are based on test subjects imagining a certain situation in which they must make a choice. The results of these experiments illustrate the dilemma of sunk cost; that it causes people to make inexpedient decisions mostly by investing more money in failed projects. The reason why the choices are considered inexpedient or irrational is that other subjects provided with the same choices, but where no previous investments have been made, chooses the opposite decision. An example of a test scenario is illustrated below where the test subjects are being presented to one of the two scenarios.

A) *As the president of an airline company, you have invested 10 million dollars of the company´s money into a research project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed, another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company is building. The question is: should you invest the last 10% of the research funds to finish your radar-blank plane?*

B) *As president of an airline company, you have received a suggestion from one of your employees. The suggestion is to use the last 1 million dollars of your research funds to develop an airplane that would not be detected by conventional radar, in other words, a radar-blank plane. However, another firm has just begun marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company could build. The question is: should you invest the last million of your research funds to build the radar-blank plane proposed by your employee?*

The two scenarios are basically the same except 9 millions have already been spent in scenario A, whereas no money has been spent yet in scenario B however, the result demonstrate a significant difference in the answers (see in table 4.2). [Arkes and Blumer, 1985]

Condition	Continue	Discontinue
9 million spent	41	7
0 million spent	10	50

Figure 4.2: *The table presents the result from Arkes and Blumers airplane experiment demonstrating sunk cost.*

This result supports the theory of sunk cost, and thereby contradicts the axiom of independence described in section 3.1 about decision making.
A point of criticism of the experiments made by Arkes and Blumer is that they are primarily based on the test subjects´ ability to imagine different scenarios and evaluate which decision they would make in such a scenario. This raises the question if the experiments are measuring how humans actually act in these scenarios or only how test subjects believe they would act? Can test subjects' ability to imagine spending money provide the same basis for making a decision as test subjects actually spending the money? Theories concerning affective forecast have demonstrated how predictions of future behaviour is mis-judged in certain situations [Ayton, Pott and Elwakil, 2007]. Nonetheless, real examples of sunk time can be found through history; an example is the Concorde project. This project was continued even though it had gone far above budget as the Concorde flew for 27 years without creating a profit. The reason it is not in operation anymore is because of one crash in year 2000, which most likely have been a "good" excuse to retire the project instead of admitting

that it was an economic failure. This might be applicable for the prediction of imaginative decisions as well, consequently invaliding the results of these experiments.

Not all experiments however, have been imaginative. Arkes and Blumer (1985) executed a non-imaginative experiment with theatre season tickets sold with more or less discount and at normal price. In the following six months the number of attended plays were counted for each subject.
The normal price for the tickets was $15 for a season. The small discount given was at $2, making the price of the theatre season ticket $13 and the big discount was at $7 making the price for the ticket $8. The ticket seller explained the discount to the test subjects as part of a promotion by the theatre department. During the following six months it was monitored how often each test subject's season ticket was used. The results showed that test subjects who paid full price for their ticket used it significantly more times than test subjects who were given either a small or large discount.
This result demonstrates an effect of test subjects wanting to avoid feeling that the money they have paid for their ticket has been wasted, why test subjects who paid the most wanted to attend more plays to "get their money's worth".

The results of another experiment by Arkes and Blumer (1985) showed that test subjects are more affected by sunk cost when they are personally responsible for the choice they are facing than when they are answering on the behalf of a company or a third person. Also, sunk cost is still present in situations where the test subjects are given an item as a present and provided with the knowledge of the monetary value of the item. However, the results of the experiment showed less sunk cost than when the scenario stated that test subjects had paid for the same item. [Arkes and Blumer, 1985]

When viewing sunk cost from a user scenario aspect, the presented findings could be exploited e.g. by heightening the price of a product users would find hard to forsake later since the money spent buying it would then be wasted. Furthermore, if sunk cost is applicable for use scenarios, updates for a program could potentially be sold easier, since users who bought the original product are willing to invest or spent more money on the product to improve it in order for the original investment not to appear wasted.

Sunk cost is an established theory which is not particularly applicable in user interaction scenarios, as money is rarely part of the use situation, why it is more applicable in marketing settings. It can of course be discussed if it is a use situation when a World of Warcraft gamer has to pay in order to continue his use of the product however, it relates more to a marketing perspective and how to make most money of consumers.
Then why is the effect of sunk cost examined? The underlying psychological mechanism, which seem to make us persist in a situation or with a choice despite no one forcing us is interesting if it is transferable to the domain of time, which is more relevant to use situations. Time is always a factor when performing an action or interacting with a product, why the following sub section will examine sunk time and its implications, which are build upon the research of sunk cost.

Sunk time and effort

The effect of sunk cost, if applicable in use scenarios, has the potential of effecting users buying behaviour why it could prove to be a powerful marketing tool. However, marketing is not included in the scope of this project, why it would be far more interesting to see if an effect of sunken time and effort exhibit a similar effect on users. Spending time and effort is unavoidable in product- and user interaction, hence an effect on users' decision making of spent time and effort can have large consequences and be used both as guidance of the user

however, also to manipulate user decisions.

The effect of sunk time has been examined by Soman (2001), who demonstrated through several experiments that the effect of sunk cost is ineffective when the past investments made are temporal rather than monetary. Through analysis of literature concerning time, money and sunk cost, Soman argues that time is very different from money, why sunk cost does not apply for time. These differences are visible due to the arguments that:

- Time cannot be inventoried or replaced, as is possible with money
- Time is not as easily aggregated as money
- Accounting for money is a routine activity, but accounting for time is not

It can be argued if all these statements are valid, however, time and money are different on many levels. Time is not as easily replaced as money as Soman argues, however, replacing money often includes spending time earning the money, why it could be argued in that context that time is money. Furthermore, accounting for time can be said to be as much a routine activity as money, since everyday living is based on accounting time in order to e.g. abide appointments and retain a job. The difference in the accounting of time and money can be seen in the fact that money is a physical object that can be gathered and traded for other objects or values, whereas time is non-physical. This might make time harder to account for than money.

To support his argument, that time is not comparable with money and as a consequence sunk time is non-existing, Soman conducted a number of experiments with the purpose of exploring sunk time. One experiment included setting up a scenario where test subjects were to imagine having spent 5 and 15 hours on two different jobs and as payment been given a ticket to the theatre and a rock concert, respectively. The events take place the same night why only one can be attended. The test subjects are informed that they will enjoy the rock concert more (which they have worked 5 hours for to obtain a ticket) than the theatre performance (which was earned through 15 hours of work) and they have to choose which event to attend.

Result of sunk time experiment	
Theatre	4.8%
Rock concert	95.2%

Result of sunk cost experiment	
Theatre	61.7%
Rock concert	38.3%

Figure 4.3: *In the table to the left the results from the sunk time experiment by Soman are presented and in the table to the right the results of the same experiment only with sunk cost is presented.*

The results of the experiment can be seen in the left table in figure 4.3 and shows no effect of sunk time, since such an effect would have resulted in the theatre ticket being chosen more often than the rock concert ticket. When repeating the experiment with the variation that the subjects have bought each ticket - theatre ticket price $450 and the rock concert ticket price $150, the results are very different as seen in the right table in figure 4.3, thereby supporting the experimental setup layout.

It can be argued if the options of a ticket to the theatre or a rock concert are equal considering that the participants are undergraduate students. It is possible that a preference for the rock concert ticket is present, independent of the influence of sunk time; had the mean age of test subjects been 50 years a preference for the theatre ticket might have been present instead. Hence, it is questioned if the age of the test subjects is a masking factor and if repeating the experiment, with the variation that the scenario states that the test subjects would enjoy the theatre performance the most, would change the results. If these alterations would change the outcome, it will not necessarily prove sunk time, but disprove the experimental setup used in this experiment. The fact that sunk cost is present in the latter experiment where

the tickets are bought indicates that sunk cost might have a larger effect on decision making than sunk time, why the effect is measurable despite the possible masking in the experiment.

In a further experiment half of the test subjects got information on the wage rate for the hours they had worked. This indirectly provided them with the information of how much each of the two tickets was worth monetarily. The subjects were to rate on a 9 point scale how likely they were to stay at home from a sports event they had been given a ticket to due to an exam (1=definitely staying home, 9=definitely going to event).

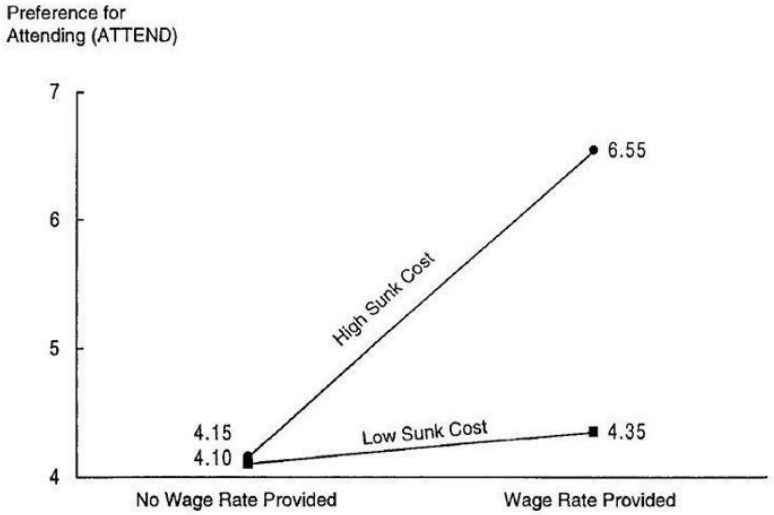


Figure 4.4: *Availability of wage rate demonstrates that sunk cost is present in the experiment, however sunk time is not [Soman, 2001].*

The graph in figure 4.4 illustrates the results of the experiment and the left side of the x-axis represents the assessed willingness to attend the sports event, where test subject have no monetary knowledge . No difference in assessment is seen between test subject working 10 hours (the uppermost curve) and test subjects working 5 hours (the bottom curve). The right side of the x-axis represents the assessed attendance willingness for test subjects provided with monetary knowledge and here a large difference between the two curves is seen, demonstrating that sunk cost is affecting the test subjects' assessment.

Another type of experiment was made where students were given two different gifts - a souvenir shot-glass and a paperweight - in exchange for completing two questionnaires - one took approximately 8 minutes to complete, other 25 minutes. The two gifts are valued approximately the same and the subjects are only to choose one. The question is, do they choose the present they got for the long questionnaire or do they choose the present they want the most? Only 44% chose the gift worked longest for indicating indifference in preference for the two gifts and thereby that the subjects chose the gift they want the most.
However, another group of test subjects were told that originally the reward for participating in the experiment was $10 per hour, but certain regulations had made them exchange the money with the two gifts. The test subjects in this group are therefore indirectly given a monetary value for each of the gifts and thereby the fact that the gift for the long questionnaire will be worth more. The results for this group showed that 80% of test subjects chose the gift from the long questionnaire.
The time difference between the two tasks in this experiment is very small (five hours) and according to Soman time is harder to account for than money, why a larger influence of sunk time might be necessary for the effect of sunk time to be measurable. Furthermore, if Arkes and Bulmer's argument is valid, stating that sunk cost effect results of attempting to avoid waste, the setup of the experiment might be invalid. Test subjects are to complete two questionnaires during the "sunk" time however, this time will not be wasted no matter

which gift they choose, since both questionnaires as far as the test subjects are aware, will still be useful for the result of the experiment. It seems unlikely that the test subjects are only willing to complete the questionnaires to be rewarded with a gift.
Soman argues that there is no sunk time effect, since we lack the ability to account for time in the same way as we account for money. Money spent can be substituted for other objects or values. However, time can never be inventoried or replaced. Furthermore Soman argues that accounting money is a routine for us, while accounting time is not and therefore it is not easy for us. This, he claims, is also why we have expressions like "getting your money's worth".

Whether sunk time exists or not is not clear from the results of Somans experiments, however Navarro and Fantino (2008) made an imaginative experiment based on the following scenario termed condition "Zero time". [Navarro and Fantino, 2008]

Imagine that you are the leader of a copper-mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek". **Today is the first day of the dig.** *Right away this morning, you and your group have found a vein of copper! However, there is a thick wall of quartz, which is easy to dig through, covering the most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small - about 10 pounds. Typical mines contain upwards of 500 pounds. You have a choice to make:*

A. *Dig the 10 days to collect the 10 pounds of copper*
B. *Abandon this "shady Creek" mine and go home*

The scenario was altered slightly in three different conditions. The second condition termed "Time-easy" differs from the first by having replaced the bold sentence with the following text: **"So far, your group has been digging at Shady Creek for the past 60 days. The ground so far has been soft, so the time spent digging has been easy."**

Condition three termed "Time-hard" has the bold text replaced with the following: **"So far, your group has been digging at Shady Creek for the past 60 days. The ground so far has been hard. The time spent digging has been very difficult and has involved a lot of effort from you and your group."**

Condition	Percentage choosing to persist the dig
Zero time	34% (11/32)
Easy time	53% (34/62)
Hard time	68% (40/59)

Figure 4.5: *The table contains the results from Navarro and Fantinos experiment concerning digging.*

The results in figure 4.5 shows the number of test subjects choosing to persist with the dig in each condition. A significant influence of time was detected when comparing the number of test subjects persisting with the dig in the two time conditions against the zero time condition. However, no significant effect of effort was found, but an increase of persistence was registered in the hard time condition.
In another variation of the experiment, the scenario was changed into being only the test subject themselves on a quest for a precious stone, where the outcome of the experiment resembled the previous. This demonstrates that sunk time effect is present as well, when the choice affects only the individual.

Furthermore, Navarro and Fantino (2008) conducted three non-imaginative experiments where test subjects were to spend either 10 or 50 minutes on solving a preliminary puzzle under certain circumstances, an experimental overview is seen in figure 4.6. Test subjects are divided into two overall test goups: voluntary and obligated. Test subjects in both conditions are informed that the experiment will take 75 minutes. After finishing the preliminary puzzle task lasting either 10 or 50 minutes in each condition, test subjects in the voluntary condition face choice 1 of either completing a number of questionnaires, which will last the remaining time of the experiment or get a new puzzle task and win the opportunity to leave the experiment 15-30 minutes early. Test subjects in the obligated group are obligated to do the puzzle task and are therefore never provided with choice 1.

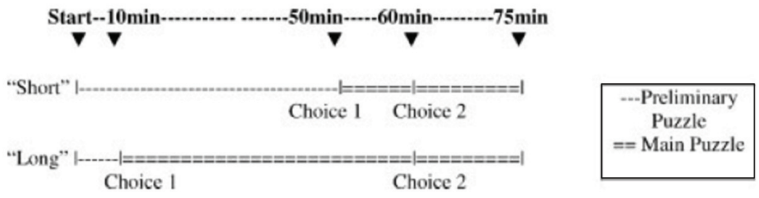


Figure 4.6: *The experimental setup of Navarro and Fantinos non-imaginative experiment concerning sunk time [Navarro and Fantino, 2008].*

Test subjects in the voluntary condition choosing to solve a new puzzle, are provided with a pair of dices when 30 minutes remain of the experiment. The subjects roll the dice and if the eyes total five or less they may leave the experiment. If the eyes on the other hand total six or more the test subjects must stay 15 minutes more and are faced with choice 2 consisting of the following two options:

A. *Continue the current puzzle task for another 15 minutes, then go home*

B. *Complete a series of questionnaires for 15 minutes, then go home*

Test subjects in the obligated condition was not provided with the dice, but were all to make the choice between A and B. The aim of the experiment is to measure how many test subjects choose option A dependent on the conditions of obligation and volunteering together with the amount of time spent on the puzzle before facing choice 2. Due to this, only the last decision between option A and B is of importance. The results can be seen in the graph in figure 4.7 where the percentage of test subjects choosing to persist with the puzzle in the voluntary and obligated groups is presented together with the information of time condition.

An effect of sunk time can be seen in the voluntary group, since far more test subjects in the long condition chooses to continue with the puzzle compared to test subjects continuing in the short condition (see the dotted curve in the graph in figure 4.7). In the group of obligated test subjects no effect of sunk time was measured.
Arkes and Blumer demonstrated how test subjects were of greater risk to fall for sunk cost when they were personally responsible for the previous investment of money and the result of this experiment by Navarro and Fantino indicates that the same might be evident with sunk time. When test subjects were obligated to do the main puzzle, fewer of them chose to persist with the puzzle in choice 2 than when having voluntarily chosen to do the main puzzle at choice 1. This makes the voluntary subjects responsible for performing the main puzzle task, whereas the obligated test subjects did not have a choice and therefore no responsibility.

The experiment was repeated with a different activity and the effect of sunk time was still observed however, the effect of obligation disappeared. Test subjects were to organise decks of playing cards after certain rules. A significant effect was found in terms of sunk time, but not of any other factors (see the graph in figure 4.8). According to Navarro and Fantino

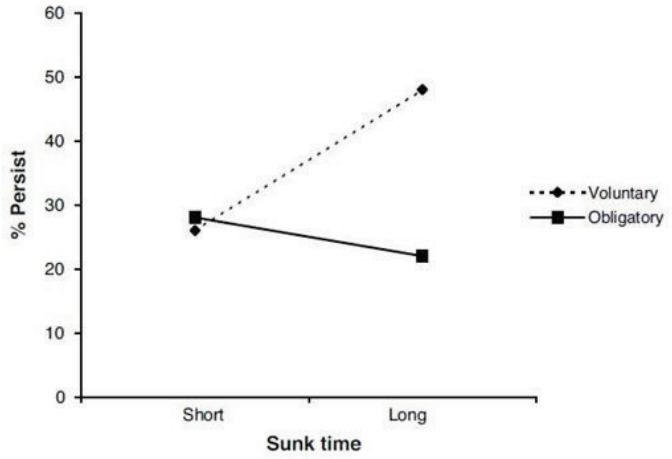


Figure 4.7: *The graph illustrates the small difference between the short and long conditions for the obligated test subjects, while sunk time has an effect on the test subjects in the voluntary group where a large difference in persistence is seen between the short and long condition [Navarro and Fantino, 2008].*

(2008) the difference in the results from the current experiment in relation to the previous, is due to sunk effort. They believe that test subjects might have put more effort into sorting the cards than solving the puzzles because they found the card sorting task harder.

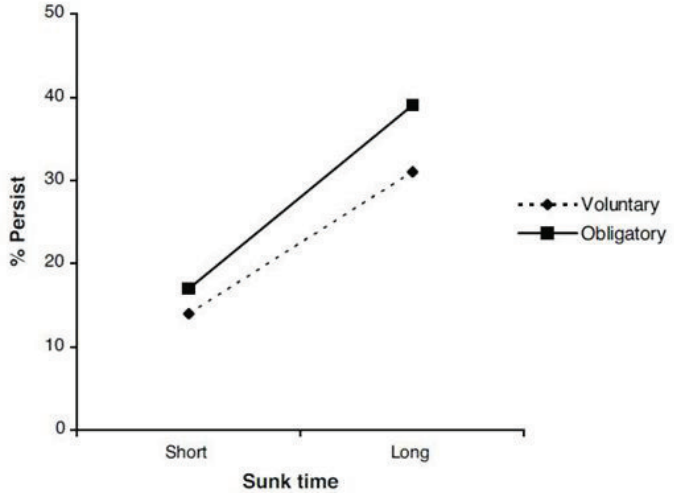


Figure 4.8: *The graph shows how an effect of sunk time is present in both the voluntary and obligated group when the activity of the experiment is changed to sorting decks of playing cards [Navarro and Fantino, 2008].*

To explore if this was in fact the reason, a follow up experiment was made where test subjects completed both the card sorting- and puzzle task. Following test subjects were to rate which task was more difficult and which task they had put most effort into solving. Significantly more test subjects rated the card sorting task as more difficult than the puzzle task and further stated to have put more effort into solving the card sorting task. This supports Navarro and Fantino's statement of the difference in results between the two experiments to be a difference in the amount of effort the subjects did put into each task.

Through several experiments Soman (2001) rejected the fact that sunk time effect exists, however, Navarro and Fantino (2008) has provided experimental grounds which suggests the

opposite. It needs to be noticed that different approaches have been used by the different researchers towards sunk time. In Navarro and Fantino's experiments the time spent will be a complete waste if the test subjects choose to quit opposite Soman's experiments, where the time test subjects spend will not have been wasted since the subjects used it on completing questionnaires. On this basis it could be questioned if Soman experiments only lacked to demonstrate sunk time due to a poor experimental setup.
It should however be pointed out that Soman found an effect in his experiments when he repeated them monetarily, which could be explained by the fact that monetary sunk cost might have a greater influence on us than time and effort, and the effect is therefore still detectable despite a poor setup. In that sense Soman might be right in his analysis and conclusions that we are better at accounting for money, since money has a clear value whereas time and effort can vary in value.

On the basis of the findings in the present studies the question concerning if sunk time can be applied in a user interface context becomes very interesting. If applicable, interfaces designed without regard for sunk time might unintentionally impact user's decision during interaction. As a consequence users might suffer under the influence of sunk time making decisions they would otherwise not have made. An example could be continuing the use of e.g. a text editing program due to the time put into installing and learning it, even though installing and learning another better program would save the user time in the end. The positive aspect of applying sunk time effect in user interfaces is that users might be more forgiving towards interfaces with long learning curves. Once time has been spent on the interface, users would potentially find it hard to give it up, since the time already spent would then be wasted. In fact it is hard to imagine a user interaction situation where sunk time would not have an effect, since time and effort always act as a premise in user interaction. However, are there certain requirements that must be valid for sunken time in order for a sunk time effect to apply?

The Influence of Occupied and Unoccupied Time on Sunk Time

Sunk time effect seems to be fully based on time and how this is spend, why studying the effect further includes the questions: What is time? Are there different types of time? And does other variables influence sunk time? Is the influence of sunk time fully dependent on the actual time that is sunk? Or does the perception of the sunken time have an influence on the effect?

Opposite money, time can have many different states; enjoyable or boring for instance. Another factor that distinguishes time and money is that money is a physical object that can be traded for other objects or events, while time is a non-physical constant almost only relying on individual perception. When waiting for an event to happen time is often perceived very slow while time can be perceived as having almost disappeared during a fun event.

The connection between time and money exists in the fact that time is traded with money in most types of work and time and money can in that context be said to be the same, or at least tradable with each other. Since time relies almost completely on perception, an interesting aspect is to examine if sunk time effect is influenced by how the sunken time is perceived or if the effect is only relying on the actual amount of time that is sunk.

In several experiments made by Oxoby and Bischak (2005) it was explored if the occupation of the time that is sunk has an effect on decision making. By experimentally comparing an amount of time in an occupied and unoccupied state it was found that the occupation has an effect on test subjects' decision making, and thereby sunk time. The experiments were conducted on the basis of a series of dictator and ultimatum games. In the dictator

games half of the test subjects (the proposers) were to offer an amount of money x out of 100 laboratory dollars to the other half of test subjects (the responders), individually paired. If the respondent accepted, the proposer would receive 100-x laboratory dollars and the respondent x laboratory dollars, however, if the respondent rejected the amount neither of the two would receive any money. Three different test conditions were set up: one third of the respondents and proposers were immediately provided with the choice of how much to propose and accept, one third had to answer questionnaires of a duration of five minutes before proposing or accepting the money, and the last third had to wait five minutes unoccupied before entering the game (the screen of the game was displaying "Wait a moment" for five minutes). [Oxoby and Bischak, 2005]

The ultimatum games had a twist compared to the dictator games. The proposer was still to offer an amount of money x out of 100 laboratory dollars, however the respondents had to estimate a minimum acceptable offer y of 100 laboratory dollars before receiving the offer from the proposer. Again, if the proposers' offer x was greater than the respondents minimum acceptable offer y, the outcome was the same as in the dictator games; that both participants received money. On the other hand, if offer x was less than offer y neither the respondents nor the proposers received any money. The aim of both experiments was to reveal if occupation during the sunken time is a factor that can affect the sunk time effect and thereby test subjects' subsequent decision making.

In the graph in figure 4.9 the percentages of monetary offers made by the proposers in the dictator games are visualised for all three time conditions: no time, five minutes occupied wait and five minutes unoccupied wait. The uppermost curve is the offers made by the proposer exposed to five minutes unoccupied wait before making the proposal. The two bottom curves illustrate offers from proposers with no waiting time and occupied waiting time. It is noticeable that only a very small difference in offers is detected between these latter conditions. The offers made by the proposers in the unoccupied condition are much lower, illustrating that the proposers wanted more money for themselves. In fact almost 40% of these proposers offer the respondents nothing, whereas only 13% of the proposers in the two other conditions offers nothing to the respondents.

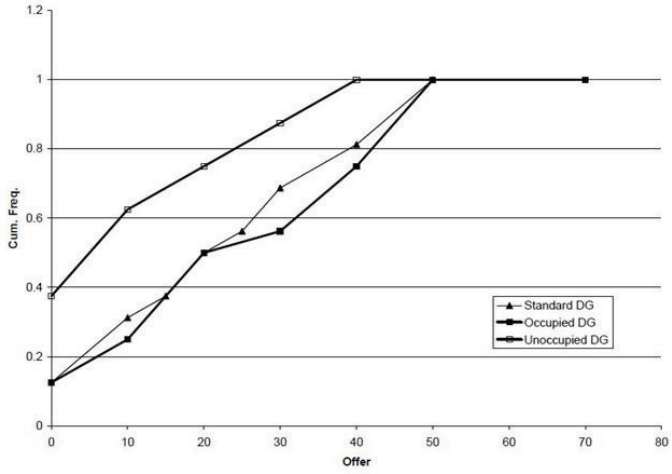


Figure 4.9: *This graph illustrates the value of the monetary offers the proposers give the respondents in the dictator games in each of the three time conditions [Oxoby and Bischak, 2005].*

In the ultimatum game both the proposers' offers and the respondents' minimum acceptable offers are registered. Again the results reveal that a significant difference is found between the offers made by proposers in the condition of unoccupied time and the offers made in the

two other conditions as seen in the graph in figure 4.10. The same tendency can be seen in figure 4.11 which illustrates the minimum acceptable offers requested by the respondents. Again no significant difference is measured between the offers in condition no time and occupied time for both the respondents and the proposers, despite the occupied condition lasting five minutes as well. Occupation hence seems to affect sunk time effect.

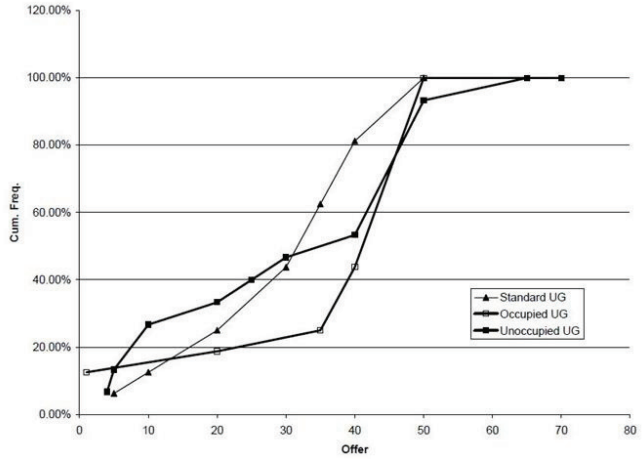


Figure 4.10: *This graph provides an overview of the offers made by the proposers in the ultimatum games in each of the three time conditions: no time, occupied time and unoccupied time [Oxoby and Bischak, 2005].*

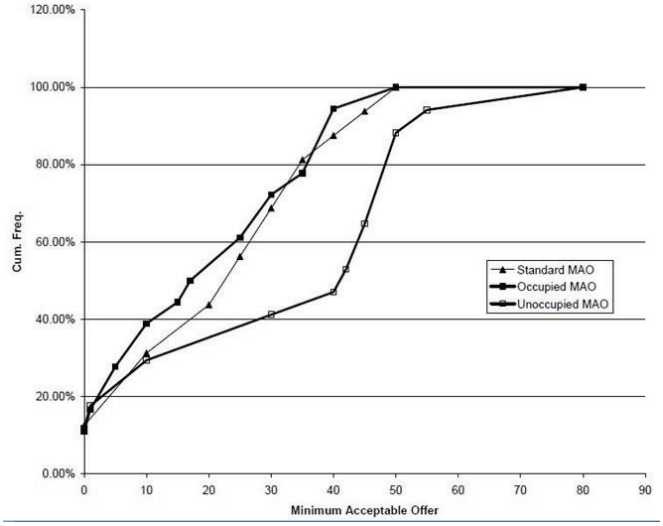


Figure 4.11: *A visual overview of the minimal acceptable offers requested by the respondents in the ultimatum games in each three time conditions [Oxoby and Bischak, 2005].*

These results comply with the statement by Arkes and Blumer (1985) that sunk time is caused by the fact that humans want to avoid wasting time, money and effort. In the condition of occupied time the time waste is not apparent for the test subjects since they are spending their time on completing questionnaires, hence not wasting the time. However, it could also be argued that sunk time has nothing to do with wasted time, but rather the perceived amount of time spent. The detected difference in this experiment could then be explained by the fact that test subjects in the different time condition groups perceived the five minutes differently. A number of sayings support this argument, claiming that time passes by much faster when having fun and that minutes feel like hours when bored.

On the basis of this, the difference in the results between the two time conditions could be explained by the fact that the test subjects in the bored condition might perceive the five minutes wait as ten minutes, why they are more affected by sunk time than the test subjects in the entertained condition, who might have perceived the five minutes as three minutes. It is not evident from the article if the test subjects have been asked to rate the amount of waiting time. Only by having the test subjects rate how long they perceive the 5 minutes wait in each condition, a conclusion about this can be made.

The consequences of these findings for a use context, could be that no sunk time would exist unless the users are wasting their time or that the effect of sunk time could be influenced by making the user perceive the sunk time as shorter or longer than the actual sunk time. During an installation of a program, which might take 2 minutes the effect of sunk time could thereby possibly be avoided by entertaining the user, so no time is wasted or the time at least is perceived as shorter.

Discussion

Sunk cost has been explored and demonstrated numerous times. Possible consequences of applying sunk cost in a user interaction context could be manipulation of users to invest more money on products that otherwise would have been discarded or the continuation of the use of product strictly due to the amount of money invested in the product when bought. However, consequences of sunk time are valued higher in user context, since cost is not necessarily included in user interaction whereas time and effort are always. The results however, of sunk effort and time experiments are more differentiated.

In general the experiments presented in this section showed a tendency of time and effort. However, Soman (2001) conducted a series of experiments with a very different approach to sunk time, and the result of those concluded that no sunk time exists. According to Arkes and Blumer (1985) however, sunk cost is present due to a desire to avoid wasting money, time or effort. The experiments conducted by Soman all lack the attribute of time being wasted. Furthermore the scenarios Navarro and Fantino use in their experiments were based on much more time e.g. 60 days compared to 25 minutes - 15 hours, which is the largest time period used in an experiment by Soman. This could be the reason why no sunk time is found in the research made by Soman (2001) that demonstrates a non-existence of sunk time.
Furthermore, the scenarios test subjects are exposed to in the imaginative experiments could be categorised as extreme compared to decision scenarios a user would experience in reality. When faced with the issue of whether or not to invest more money in a company, the choice would not be just between investing and not investing. Several other solutions would be considered such as, buying small parts of the plane and developing it further or only investing some of the money. Likewise in the copper mining scenario, the two options are quite extreme and in reality compromises like giving up the dig at “Shady creek”, but finding another place to dig would possibly exist. This questions, if the effect of sunk time measured in the scenarios is representative for how humans would react in scenarios where the situations are often more complex with more than two solutions.

Time is a phenomenon with many different states; fun, occupied, boring or hard while money only holds one state and one value. The same might be evident for effort. This might provide a large nuisance variable in the imaginative experiments, since test subjects are somehow impaired to imagine time, since they are not aware of the state of the time they are to imagine and thereby are much better at imagining cost.

Sunk cost is bound up on money or items representing a certain value, which does not apply

for sunk time. When it comes to user interaction with a product; money will rarely be a mean of interaction, whereas time can play a key role. Therefore, the sunk time effect might play a role in the interaction with a product. Having spent a certain amount of time on a product can as a consequence leave the user more willing to continue using the product, if sunk time is applicable in user settings. Also, waiting time especially in graphic user interfaces is a reality in terms of loading times, installation times etc., why the perception of sunk time might play a crucial role in relation to the use or perception of a program for example. If sunk time does exist given a use context, this may influence the users assessment of a program, why this will be examined further.

The question concerning if sunk time is affected by the perception of how much time is sunk or if sunk time is wasted or both, is unanswered in the present experiments. However, it remains a very relevant question in the aspect of applying the theory of sunk time in a user interaction scenario, as different ways of affecting users with sunk time may exist. Further examination of this subject together with replications of other sunk time experiments will therefore be conducted in an attempt to clarify the influence of sunk time in a user interaction context.

4.2 Framing

Whenever a decision is made, it is made within some sort of context. The context might determine the options available, but it also frames the problem and the available options. The focus of this section will be to explore if and how framing affects the decisions people make. The research made in this area suggests that humans are susceptible to be affected by the frames our decisions are made within. The scope will then be to determine if the findings can be applied actively in the design of a user interface.

Framing has been studied in multiple setups whereof research by [Tversky and Kahneman, 1981] forms the basis for understanding the way people react towards framing. In a series of experiments they show how decisions involving risks can be affected by the setup or the framing of the decision problem, which supports that framing can leave us to make contradictory decisions.

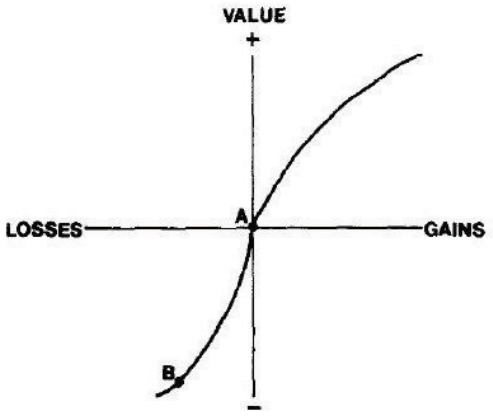


Figure 4.12: *Value function based on Prospect Theory [Tversky and Kahneman, 1979].*

The results of the experiments made by [Tversky and Kahneman, 1981] all relates to the Prospect Theory value function in figure 4.12. The experiments use different frames, which either assumes a gain or a loss. The results of the first experiment clearly exhibit how framing affects our decisions dependent on a positive or negative frame. The test subjects in experiment 1 are given the following problem:

Imagine that U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs is as follows:

One group of the test subjects get the following two solution options:

1 *If Program A is adopted, 200 people will be saved*

2 *If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved*

Another group of test subjects get the following two solution options:

3 *If Program C is adopted 400 people will die*

4 *If Program D is adopted, there is a 1/3 probability that nobody will die, and 2/3 probability that 600 people will die*

Chosen program	Share of test subjects
A	72%
B	28%
C	22%
D	78%

Figure 4.13: *Results from experiment 1.*

The results of the experiment can be seen in figure 4.13. It shows that the for two solution sets presented to the two groups of test subjects, which in fact presented the same condition, the framing of the options was critical for the option chosen.

This result is explained by the value function in figure 4.12. Test subjects presented with the first set of options are biased by the wording to think of gains, why they become risk averse. Therefore the majority choose to go with the option of a certain gain of 200 lives. In the second group of test subjects the outcome is reversed, since they are biased by the wording to think of losses. Therefore they are more risk seeking, and are willing to "gamble" the death of 600 instead of accepting that only 400 will die, because they have something to "win".

The experiments by Kahneman and Tversky (1981) are limited to include only numeric probabilities and quantitative outcomes. They set up experiments where test subjects are to choose which of two gambles they want to take. The majority chooses the same bets throughout the experiments, even though they choose the most non-profitable options. Furthermore, when confronted with all four options together, test subjects have no problem spotting which is the most profitable option and hence choose opposite from the previously obtained result. The effect occurs since two option sets are framed independently, and even with real money as incentive the same pattern is observed. The authors argue that some real life concurrent decisions are framed independently, which in turn can distort the preferences, why the framing problem extends beyond the lab experiments.
What Kahneman and Tversky implies is that if the information provided to us is only presented from one angle we are not able to extract excess latent information, such as if there is a 10% chance of dying from a disease, the latent information is that there is a 90% chance of surviving. If we are not able to extract this type of information, we will make decisions based on what is provided to us. If this statement holds, then framing is a powerful tool for manipulating humans, even more so if we are not able to detect what make us make a certain choice. If this also applies in use situations and the user is not able to grasp all information due to some of it being framed unintentionally, this could create issues. For example, a user using a program needs a tip concerning tab short cuts, the only button he sees is the "help" button, which he presses. He searches in the help function, but ends up not finding what he seeks. Half an hour later he discovers another button called "tips and tricks", which he then presses and discovers his answer. In this thought case, the user is framed by the word *help* and sees this as his only option, not looking specifically for other buttons as he is not able to extract the surplus information that help is when you are in trouble, whereas tips are for improvement. As mentioned this is a thought rather odd example, and multiple other factors such as experience would most likely also play a role, but if this could be observed in use situations, many "errors" in use might occur as a consequence.

In another reported experiment the authors report on another type of decision behaviour which contradicts normative rules. In the experiment three gambles are set up, where one of the gambles (A) is based on a sure win versus a risky gamble. In the second gamble test subjects are told that they are faced with a two staged game; there is a 25% chance to move to the second stage in which test subjects are faced with a gamble (B) based on a sure win

versus a risky gamble. The third gamble (C) has two risky options.
Test subjects are only assigned one of the three gambles, nonetheless they should choose the corresponding options across gambles, since the options across gambles are identical, if analysed. However they did not; in gamble A and B test subjects chose corresponding options, but not in gamble C.
The authors point at two effects to explain this incongruence; the "certainty effect" and the "pseudocertainty effect". The certainty effect rises when one outcome is certain whereas the other only has a probability attached to it. The certainty is more attractive to the test subjects than a high probability, why the majority chooses a certain gain in gamble A and B.
The pseudocertainty effect is displayed when the results from gamble B and C are compared. In gamble C test subjects actually demonstrate risk willingness in order to obtain a higher gain, since none of the options are certain. The "psuedo" comes from the fact that test subjects in gamble B ignores the fact that they have to make it through another gamble in order to get to the present gamble, which decreases the actual likelihood of the certain gain in gamble B, and thereby equals the same options as in gamble C. Therefore, the observed effect manifest itself by test subjects tending to forget or cancelling out the first decision in the process, and thereby they weigh the uncertain as if it was certain. The article also points to other experiments where the same observations regarding certainty and pseudocertainty effects have been observed.
The certainty effect holds interesting aspects towards user interfaces as it suggests that humans are more prone not to gamble a sure win with the chance of a large win and a loss. This might also apply in user interfaces, if users always work with a certain set of settings, then why should they risk the change from a setting they know and use, to another setting which only might be better, but is not guaranteed to be.

Another attribute of Prospect Theory is to evaluate ones' options based on a point of reference. If an outcome is evaluated as a gain or loss depends on the point of reference a person in a given situation sets, which in turn acts as a frame. This effect is illustrated by an example of horse race betting. A man has lost $140 and considers taking the last bet of the day of $10 with odds 15:1. He can consider this bet in two ways using two different reference points; first, he can use - $140 as a reference point, that is where he is now and by taking the last bet he can gain $140 or he can lose $10, or he can take the initial amount of money he had as the reference point, whereas he can even out his losses by winning ($0) or he can increase his loss to $150. This means that he can either loose $10 more, or win $150 and thereby even out his losses.
According to prospect theory continuously adjusting ones point of reference introduces a type of framing where one becomes more risk seeking than normally. In fact, for horse races it has been found that risky bets are more readily obtained at the end of the day [Tversky and Kahneman, 1981].

The loss-framing experiments made by Tversky and Kahneman (1981) have been transferred to the marketing domain in research made by [Ganzach and Karsahi, 1995]. Through a field study they examine how message framing can affect buying behaviour in a natural market setting. The hypothesis of the study is based on the prior findings suggesting that loss framing has a stronger effect than gain framing.
The study is made in cooperation with a credit card company and test subjects are credit card owners who are registered not to have used their credit cards for three months or more. All of the test subjects identified were contacted by the credit card company, first by a phone call and then by a follow up letter. Each test subject was either classified as a check or cash user, since benefits from using cash are different from checks and vice versa. The two groups of test subjects were then split in two more groups dependent on gain or loss framing.
When test subjects were contacted by the telemarketing communication centre and by mail they were confronted consistently with either a gain or loss framing. An example of a gain and loss telephone script could be:

*...It is worthwhile for you to know that there are many **disadvantages** in using cash instead of a ZionCard... Cash is not only **less** convenient, but also **less** secure.*

*...It is worthwhile for you to know that there are many **advantages** in using ZionCard instead of cash... ZionCard is not only **more** convenient, but also **more** secure.*

The same pattern is used in the follow up letters.

Having contacted the test subjects, the number of their credit card charges and amount were monitored for two months in order to observe effects of the framing manipulations. Six months later 20 of the test subjects from each group were randomly chosen and phoned for a short interview. They were asked if they remembered the message, if they could recall specific arguments used in it, how convincing the arguments had seemed and how important payment method was to them.
It was observed that test subjects exposed to the loss frame exhibited a higher tendency to use their credit card; 54.8% vs. 16.4% of the check users and 45.5% vs. 24.1% of the cash users in the second month. Also, the effect was seen on charging amounts, where loss framed test subjects in average charged up to twice as much as gain framed test subjects. Furthermore, when it came to the post interview it was found that 66% of the loss framed test subjects recalled the initial message, whereas this percentage for gain framed test subjects was only 43%.

The results obtained by Ganzach and Karsahi (1995) not only show the framing effect to apply in a real life setting, but they also add another dimension to the framing phenomenon in terms of a long term effect. The framing effect continued to linger since the sales information was still in memory of a large percentage of the test subjects six months later. The effect was most pronounced for check users, which is explained by the economic incentive of saving transaction costs. Nonetheless, the observed effect of payment method was very small compared to the effect of framing, why the "psychological" effect of framing seems to weigh more than the economic effect of saving. However, the authors hesitate to conclude that loss framing can be utilised in all areas as motivation, since it might also depend on other attributes such as consumer attachment to the product and context in general. Besides the general applicability of the findings, they also support the loss aversion phenomenon described by Prospect Theory.

As the previous reported experiments illustrate the framing effect seems to be evident and effective within both experimental frames and in real life settings, where people are directly affected by the frames. However, the previous studies only focused on loss and gain framing, which seems directly applicable in marketing situations especially, but a bit more difficult to integrate in user interfaces.
Marketing is a field where the semantic content is very important why framing appears as an obvious tool in the form reported this far. However, for user interfaces it is not as applicable as user interface interaction is more command driven in terms of pressing a "start" button, commanding a computer to open a certain program etc. Therefore, the type of loss and gain framing examined this far creates a foundation for understanding the framing effect, and what make us behave the way we do. However, framing based on other premises has also been examined; in a study made by Hartman et al (2008) the frames were not set up in terms of loss or gains, instead it was investigated if social opinion frames can affect the evaluation of a website.
Two experiments were made in order to investigate if the effect is observable both without actual use of a website and with use of a website. The first experiment is set up using a questionnaire which is distributed via e-mail. The test subjects are informed that the

experiment concerns users' expectations of websites. The test subjects are then presented with information of a hypothetical website. Each test subject is randomly assigned to a piece of information regarding usability, look and feel, or content and service quality. Each of the pieces of information is furthermore randomly assigned to be framed either positively or negatively (see figure 4.14). Having answered the brief questionnaire and obtained the information test subjects were i.a. asked to rate on 7-point Likert scales the quality of the website as well as the quality attribute that had been assigned to the test subject.[Hartmann, Angeli and Sutcliffe, 2008]

The findings showed that test subjects who were given the positive framed information in all cases rated the quality of that specific attribute significantly higher than test subjects exposed to the negative frame (see figure 4.14). However, the overall quality assessment was influenced by prior information of look and feel, and service quality, but not usability.

Usability
10% of users experience difficulty using the website
90% of users experience the website as easy to use
Look & feel
10% of users find the website visually unattractive
90% of users find the website visually attractive
Content and service quality
10% of the information on the website is unreliable
90% of the information on the website is reliable

Figure 4.14: *Framing sentences used in both experiments.*

The results from this experiment shows how framing does not need to be founded on gain and losses, and that "other" user's opinion does play a role. This shows how reputation can affect judgment of an interface, but not how actual use affects judgment. Therefore Hartman et al (2008) made a second experiment in order to examine if the observed framing effect in the latter experiment is evident after use of a website. The experiment is setup in a similar way; the same wording and information is used in the pre-questionnaire distributed to the test subjects and each test subject is still randomly assigned to one of the attribute lines in figure 4.14. However, this experiment differs from the first in that one group of test subjects is exposed to a website for one minute where they have to interact with it on an exploratory level. Another group is instead exposed to a screenshot of the same website, but only for ten seconds. Based on these exposures both groups of test subjects will have to answer questions concerning the quality of all three tested attributes despite not being exposed to all in the initial priming phase. In this way it is possible to observe if e.g. a negative priming of usability also affects the assessment of look and feel for instance. Furthermore, this experiment also includes a control group where test subjects are not primed by framing beforehand, but will give their immediate assessment of the three attributes upon viewing or trying the website as well, for comparison. The results of this experiment showed the same tendencies as in the first one. An overview of the results can be seen in figure 4.15. By the graphic layout it is easy to see that test subjects who were affected by a negative frame in general rated the different attributes lower, than test subjects exposed to the positive framing.

Concerning the impact of just viewing a screenshot of the website compared to interacting with it, it was found that test subjects who had interacted with the website generally rated aesthetics and usability a bit higher than test subjects exposed to the screenshot only (see figure 4.15).

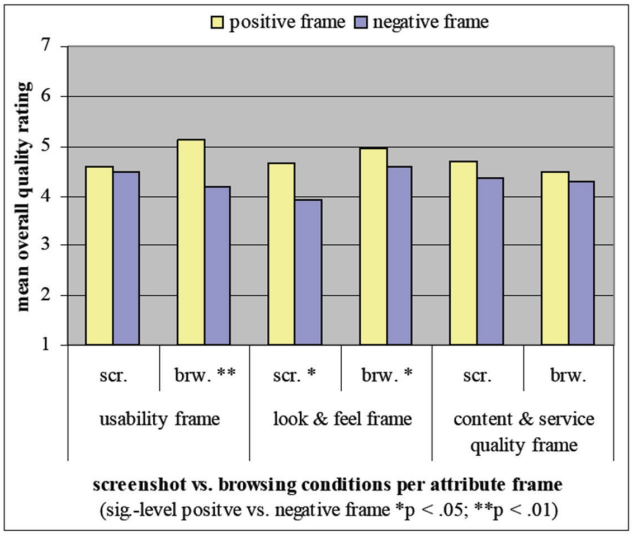


Figure 4.15: *Results from experiment 2 [Hartmann et al., 2008].*

In terms of a carry-over effect from framing one attribute positively for example, such was only found in one case. The usability framing had an effect on test subjects' evaluation of service quality, which is explained by the wording of the service quality scale, which resembles usability considerations. It was only the usability framing which had a direct effect on the overall quality assessment, for both screenshot and website use.

Overall, the results from both experiments showed that the framing effect was more subtle when test subjects were exposed to the website than when exposed to information only. However, the effect was still present which points to the effect reaching beyond decision tasks and into use situations. The results imply that if contextual information presented is not considered, it can lead to influenced decisions or evaluations beyond pure decision tasks. However, the use of the website was only limited to one minute, which might in some cases apply in real use situations as well, but nonetheless the setup did not include a specific task for the test subjects, why the motivation was not present for using the website. If there had been a task to solve, the test subjects would have been able to evaluate the website based on its actual utility and usability, which potentially could change the outcome.
Nonetheless, it is worthwhile noting that a statement can actually change the perception of a website if it is not used for too long, implicating that framing is of importance for the perception of usability, why it is necessary to account for. However, the experiment did not show if users can be affected by different means during use, instead they tested pre-framing effects.
As mentioned before, framing is a tool which appears to be able to manipulate peoples' judgments and choices, and it does not seem as if we are aware, as the Asian disease problem clearly illustrated. Thibodeau and Boroditsky (2011) therefore made a study containing five variations of one experiment, which supports the implications of the findings of Hartman et al (2008); that framing information is processed subconsciously and the effects can reach beyond decision tasks. The five experiments covered a variation of an article sample concerning crime in the city of Addison. For each of the five experiments the wording was slightly changed between two articles or a metaphor frame was provided at some point during the experiment. In one of the experiments crime was described as a *virus* in one edition of the article, whereas in the other edition it was described as a *beast*. Upon having read the article bits, test subjects were asked three questions: 1. "In your opinion what does Addison need to do to reduce crime?", 2. "Please underline the part of the report that was most influential in your decision" and 3. "What is the role of a police officer in Addison?". Overall test subjects suggested enforcement (62%) over prevention (38%), however, the majority (71%) who

had been exposed to the "beast condition" suggested enforcement compared to only 54% of the test subjects who had been exposed to the "virus condition". Concerning the second question, only 7% of the test subjects identified the metaphor as dominant, which shows how the actual frame is perceived and processed subconsciously. The third question posed was only used to clarify the suggestions made in the first question, and thereby strengthen the interpretation of answers from question number one. [Thibodeau and Boroditsky, 2011] These findings suggest that framing acts subconsciously, since test subjects are not able to identify the cues that influence their choice of how to fight crime. Also, the findings show that framing is indeed not restricted to only apply in loss and gain issues, but also in matters of vocabulary use.

As we do not seem to realise what affects our choices framing is a tool which can be used for direct manipulation, as it can be used unintentionally. However this far the mentioned research has mainly focused on word framing, which proves to be effective in multiple settings. Nonetheless, application of framing based on a physical premise also seems to induce the prior observed framing effect. Hsee (1998) examines valuing of "physical" options in a separate evaluation setup and in a joint evaluation setup. He reports of a less-is better effect, which is observed by providing test subjects with different points of reference. The effect is observed when normatively less valuable options are preferred to normatively more valuable options. The aim of the research is to examine a less-is better effect in contexts which differ from risk and gamble contexts. [Hsee, 1998]

Four experiments are set up, which all proves the "evaluability hypothesis" which states that; judging an isolated option a person will let the judgment be influenced more by easy-to-evaluate attributes, than by hard-to-evaluate attributes, despite the hard-to-evaluate attributes being more important. An attribute is hard to evaluate when the decision practitioner is not able to overview its neutral reference point, its effective range etc.
The experiments are constructed as between subject experiments so that one group of test subjects is exposed to one questionnaire or example only, and another group to another and a third to both. The following example is one of the studies, where two groups of test subjects are given the following question: (The text in the brackets is different for each of the groups, respectively.)

Imagine the following scenario: It is summer in Chicago. You are on the beach at Lake Michigan. You find yourself in the mood for some ice cream. There happens to be an ice cream vendor on the beach. She sells Haagen Dazs ice cream by the cup. For each serving, she uses a [10 oz cup / 5 oz cup] and puts [8 oz / 7 oz] of ice cream in it. What is the most you are willing to pay for a serving?

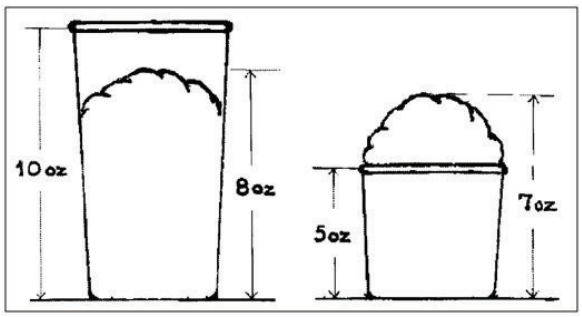


Figure 4.16: *Ice cream illustrations used in experiment 2 [Hsee, 1998].*

Together with each of the questions one of the two cups seen in figure 4.16 was shown, respectively. The third group was shown both of the cups simultaneously.

The results showed that test subjects who were exposed to the small cup with excessive ice cream were willing to pay significantly more than those who were exposed to the large under filled cup, even though the latter contained more ice cream. The third test group displayed "preference reversal", since they would pay significantly more for the under filled cup than the over filled.

All of the experiments reported in Hsee (1998) provide the same results where test subjects display a preference towards the lowest value object in the separate exhibition condition, whereas a preference reversal is found in the joint evaluation. The evaluations made in these experiments all rely on a reference point for judging how much the given object is worth. In the case of the ice cream cups, it is argued that the cup size becomes the reference and the filling is the value in relation to the reference. The test subjects all over valued the small cup with excessive filling, since they do not know if the amount of ice cream is much or little, but in regards to the cup it is much. The same judgment was made by test subjects who had the big cup, just in a reversed manner. For the test subjects who had both ice creams, the direct comparison between the ice creams led them to have a reversed preference as the others. [Hsee, 1998]

It can be portrayed as if test subjects valuing the low-value option as more valuable are wrong in doing so, however, it is only from an outside perspective that one of the options are more valuable than the other. The test subjects in the small cup group were not aware of the other option, why they are satisfied with their evaluation. In everyday life we therefore act according to the points of reference provided to us, which in turn might make us make non-optimal decisions, however, not being aware makes us unable of evaluating the decision in a bigger frame, why we not necessarily need to care.
It can be argued however, if the reported study exhibits a situation where a choice has to be made as there is no optional alternative. However, test subjects are asked to judge the value of the ice cream and upon that decide how much they would be willing to spend on the amount. This could also happen at a grocery store where a customer looks at the price of an item and decides whether she will buy the item or not. In this way it is a choice that is made, however without a clear alternative. Therefore, this reference point framing appears applicable in use situations where an adjustment needs to be made. An adjustment of a product setting for instance will provide the frame for the adjustment given the product setting options, whereby users potentially are affected by the frames of the product settings.

Framing do seem to affect us which makes us easy victims for marketing strategies, and maybe this behaviour extends into our use patterns as well. In the studies mentioned in this section, all of them were set up decision scenarios, which included judgment and valuing judgment, not direct use. The implications of framing and prospective further application are discussed in the following.

Discussion

This section presented applications of framing in different domains: a theoretical experimental setting, a real contextual setting and an objectified setting. For all contexts the framing effect was present, which indicates that we most likely are exposed and affected by this phenomenon every day. It could be argued, especially in the first case that the experiments are very much constructed to prove a point, and that these setups of decision-tasks are very unlikely to appear in everyday life. However, the results from the other sources fortify the theoretical implications that we are affected of the settings of a situation, as they were made in more "naturalistic" settings that test subjects can relate to.

The display of irrational behaviour throughout the set up experiments call to attention if it can be expected that people can make fully rational decisions, as laid out in the normative models of decision making. As found in the present experiments it seems that we are not cognitively able to command all information available to us, which in turn lead us to make incongruent decisions. However, instead of labelling incongruent decisions as irrational decisions, the adoption of a decision frame is the determining factor for the actual choice, and in general the findings seem to consolidate the s-curved value description of Prospect Theory, and the notion that losses loom larger than gains, which especially was apparent in the field research made by Ganzach and Karsahi (1995). Also, in the use situation with the website in the study made by Hartman et al (2008), it can be discussed if use over a longer period of time than just one minute, would provide the user with a more solid information foundation to evaluate the website on. If in fact a long term effect in that domain could be observed it would underpin the means of framing, which could potentially be used actively in design.

Besides supporting framing as a mean of altering peoples judgment of a situation the findings of Thibodeau and Boroditsky (2011) also showed that the vast majority of test subjects were not able to identify which word made them choose as they did towards a mean to fight crime. These findings show explicitly that the frames are perceived subconsciously, but nonetheless reacted to.

In terms of testing a direct use situation Hartman et al (2008) tested i.a. usability ratings of a website. However, this usability evaluation was based on word priming as framing, not interaction priming which is based on direct interaction with an altered user interface. The latter would be very interesting to examine further within the scope of the present project. In relation to this, the Hsee (1998) experiment setup is very intriguing, since he sets up a set of physical frames in terms of the ice cream cups, which then serves as references for judging the value of the amount of ice cream, instead of using loss and gain or positive and negative word frames.

The findings from this section suggest that the framing effect could be explored further in relation to use settings. If the effect is strong enough to withstand time would also be an interesting subject to examine, however, testing if the effect could withstand time would not be important in a website context for instance, since actions in such a domain are often transient. More immediate interaction within a framed setup would be interesting to examine further, to see if reference point shifting could prove to have an effect on judgment as found in the worded and illustrated studies mentioned.

4.3 Defaults and the Status Quo Bias

The word default relates to a set of options where one of these are preset. It can be that an option has been preset, but is changeable, and a default can also be the current situation that one wishes to stick with. Defaults are encountered multiple places e.g. on the internet in different registration forms, when one is installing a program or signing up for something. The default concept is also encountered elsewhere e.g. when ordering a menu in a fast food restaurant a medium sized menu is considered the default, and one would have to ask for a small or big one if that is desired (even though some fast food chains seem to consider the big menu as the default!).
Since these defaults are utilised in many different situations much research has been made on the area to deduce if we are affected by it, and if that is the case, how. The aim of this section is to examine the main findings within the field of defaults; with the prospect of investigating if possible effects can be used actively in the design of a user interface.

Johnson and Goldstein (2003) examined organ donation between countries where different defaults are set concerning whether to donate or not donate as well as setting up an experiment based on the organ donation theme. The experiment is set up with 161 test subjects divided into three different groups. Each group is presented with the scenario that "*you have just moved to a new state...*". However, the character of the following line is different for each group in terms of defaults. The following three lines are:

1. *...where the default is not to be an organ donor. Confirm or change that status.*
2. *...where the default is to be an organ donor. Confirm or change that status.*
3. *...choose if you want to be an organ donor or not.*

The results from the experiment showed that 42% wanted to become organ donors in the first scenario, whereas 82% and 79% wanted to become organ donors in the second and third scenario, respectively. The results from this experiment show that the default option has a strong impact on the option the test subjects choose. For comparison organ donor rates of 11 European countries are collected and presented in terms of presumed-consent- and explicit-consent countries. These rates verify the findings from the experiment, since countries with the opt-out condition has an organ donation rate of almost 100%, the opt-in countries has a maximum of 27.5 % (see figure 4.17). The attitude towards organ donation could of course be an influencing factor. However, an off-cited pole made by Gallup showed that 85% of Americans approve of Organ donation, even though only 28% had actually agreed to become an organ donor. [Johnson and Goldstein, 2003]

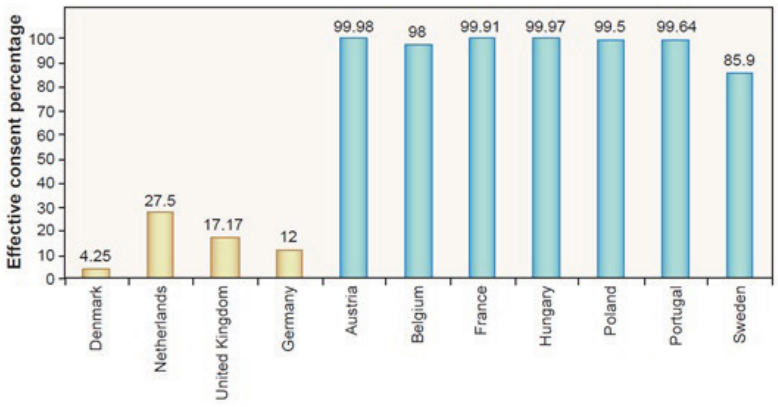


Figure 4.17: *Overview of countries with opt-in and opt-out organ donation programs, respectively [Johnson and Goldstein, 2003].*

The results show that despite the seriousness of the choice the default option plays a major part in relation to the final decision. It seems as if no choice is in fact made and that people just go with whatever has been chosen beforehand without doubting if they should just obey this preset choice or not. Multiple explanations as to this behaviour pattern has been put forward, i.a. that of cognitive laziness, whereby just following the preset options the decision maker is "spared" for the considerations involved in choosing. Other theories suggest that the defaults are perceived as the standard choice, which make people believe that it is the best option.
Despite the reasons for choosing (or not choosing) the default option, it is shown that people do choose the default whether it is to donate or not. Thereby the normative predictions do not hold as the choice is reversed according to the default option. If this holds for a decision as important as organ donation, it seems very likely that it will also apply in use situations, where users will use a product with the preset settings.

However, in order to underpin the observation and expectation more research is examined, i.a. research by Samuelson and Zeckhauser (1988) on status quo effects. The expression "status quo" refers to the notion of current state of affairs. It can be discussed whether it is exactly the same as a default, since a default also can be a change from status quo, which has been preset prior to a choice. However, the "status quo" phrase is used in the same manner as "default" by Johnson and Goldstein (2003) - which is deciding on the option which is already in use at the time of deciding. The object of Samuelson and Zeckhauser's (1988) research is to examine if a status quo bias can be observed using a series of decision making experiments. [Samuelson and Zeckhauser, 1988] [Johnson and Goldstein, 2003]

It is argued that real life decision problems often come with influential labels, where one of the alternatives unavoidably carries a status quo label. The way people react according to the status quo option is examined through an overall experiment containing sub parts focusing on different situations where the status quo effect can be involved. Also, the article includes reports from two field studies made on employees choosing health coverage and retirement investment type. Overall the findings show that a status quo bias exists in varying degrees according to type of decision problem and problem framing, but that it occurs for all of the set up decisions in the experiment.

The experiment is set up using a questionnaire containing 9 decision questions. The questions address different types of status quo influences, why 15 questionnaires in total are made to test across conditions. The number and combination of options are varied across questionnaires. The test setup is between subjects why one test subject is not exposed to the same question twice in spite of different framing of the questions.
Some of the questions are either made in a neutral way with a question and multiple choices, or with an implicit status quo and multiple options including the status quo. For an example see figure 4.18. In this way it is examined how test subjects choose if they are presented with a preset option in relation to how they would choose otherwise.
The number of options are as well varied for some of the same questions, and tested across groups with and without presets. By varying number of options it is tested whether number of options affects the status quo bias.

One of the questions is posed with different presets varying in magnitude, where the hypothesis is that greater status quo allocation will result in a greater actual allocation. See figure 4.19 for the question wording. In a final type of decision question posed in the questionnaire, the option chosen by the test subject in the first part of the question becomes a self-selected status quo option for the following part of the question. In this way it is tested if sequential decision tasks exhibit the status quo bias.

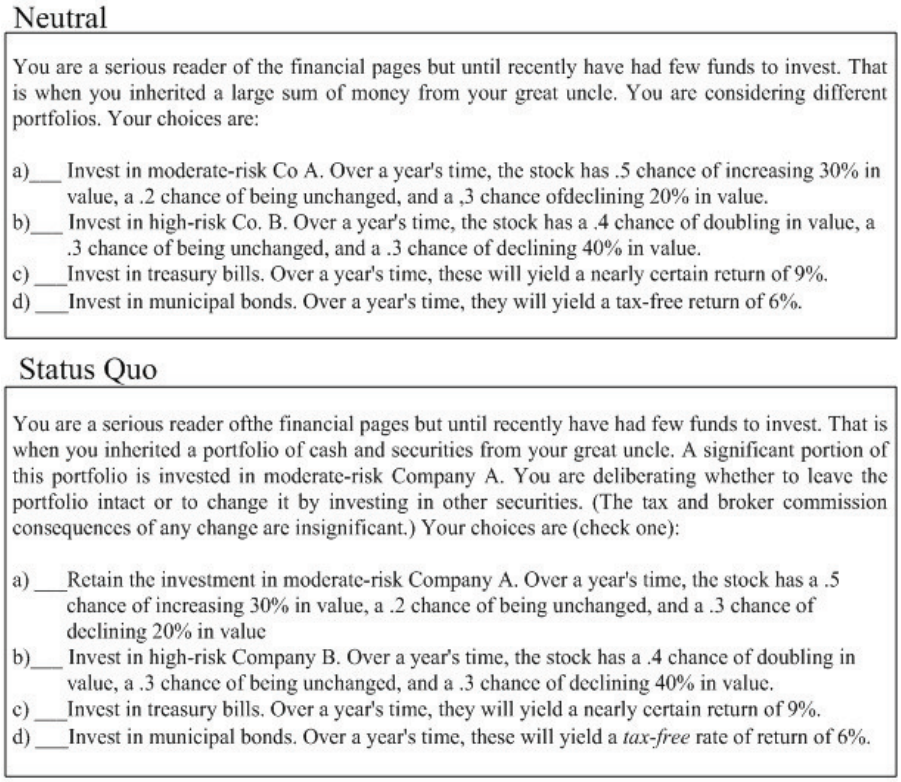
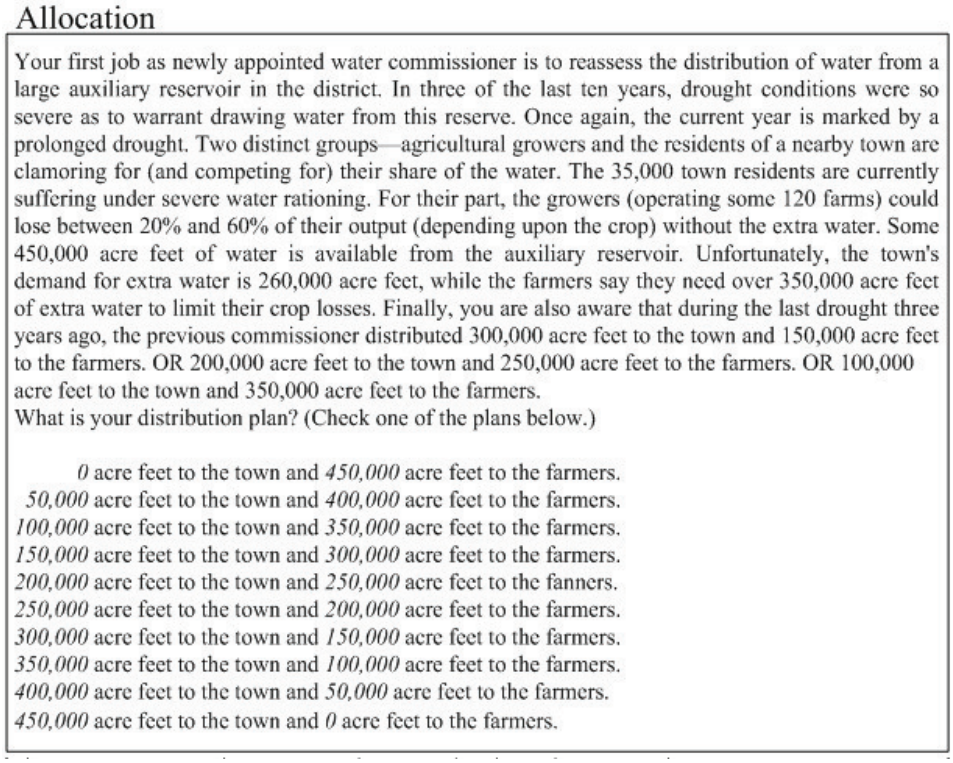


Figure 4.18: *Example on the neutral and status quo variation of one of the questions from the experiment [Samuelson and Zeckhauser, 1988].*

The results from the questionnaire support the notion of a status quo bias. For the questions where number of options and the option of status quo was varied, the status quo bias was recognised. For almost all of the question variations the majority of test subjects had been sticking to the status quo, and for more than 50 % of the questions the status quo option was chosen significantly more times than the alternatives.
This result supports the results from [Johnson and Goldstein, 2003] and implies that the default effect is applicable in more areas, as test subjects were asked to decide between more nuanced options, than in the organ donation setup where it was either opt-in or opt-out. Also, further results from [Samuelson and Zeckhauser, 1988] in relation to varying numbers of choice options show that the more options that are present, the larger the probability is for people to stick with the status quo option, and vice versa.
This is interesting as users presented with a few choices would maybe choose actively, but presented with more choices would be inclined to go with the default. Furthermore, this observation links defaults with another subject within decision making, the too much choice phenomenon where people seem to be paralysed by too many choices and might end up choosing not to choose, which in this case would be to choose the default as it is easier than consider it oneself (see Appendix I for more on the too much choice effect).

The results for the hypothesis concerning allocation of resources in relation to a status quo allocation (see figure 4.19) also underline the status quo bias. The results showed that test subjects did not allocate exactly the same amount as in the status quo, but they were affected as test subjects who were given the first status quo of “300,000 acre feet to the town and 150,000 acre feet to the farmers” did allocate more acre feet to the town than test subjects exposed to the “200,000 acre feet to the town and 250,000 acre feet to the farmers” did.

These results are very intriguing when it comes to user interfaces, certainly as users potentially can be manipulated to allocate more value if this type of default effect can be replicated in a use context. This could potentially be a disadvantage for the user, considering use of websites where the users are also customers. Being able to manipulate customers into paying more just because of a high default setting could be of great value to sellers, whereas not so much for the customers. However, if it does work in use settings as well, there might be other areas where such an effect can be applied in a positive way to guide user behaviour instead.



Allocation

Your first job as newly appointed water commissioner is to reassess the distribution of water from a large auxiliary reservoir in the district. In three of the last ten years, drought conditions were so severe as to warrant drawing water from this reserve. Once again, the current year is marked by a prolonged drought. Two distinct groups—agricultural growers and the residents of a nearby town are clamoring for (and competing for) their share of the water. The 35,000 town residents are currently suffering under severe water rationing. For their part, the growers (operating some 120 farms) could lose between 20% and 60% of their output (depending upon the crop) without the extra water. Some 450,000 acre feet of water is available from the auxiliary reservoir. Unfortunately, the town's demand for extra water is 260,000 acre feet, while the farmers say they need over 350,000 acre feet of extra water to limit their crop losses. Finally, you are also aware that during the last drought three years ago, the previous commissioner distributed 300,000 acre feet to the town and 150,000 acre feet to the farmers. OR 200,000 acre feet to the town and 250,000 acre feet to the farmers. OR 100,000 acre feet to the town and 350,000 acre feet to the farmers.
What is your distribution plan? (Check one of the plans below.)

0 acre feet to the town and *450,000* acre feet to the farmers.
50,000 acre feet to the town and *400,000* acre feet to the farmers.
100,000 acre feet to the town and *350,000* acre feet to the farmers.
150,000 acre feet to the town and *300,000* acre feet to the farmers.
200,000 acre feet to the town and *250,000* acre feet to the fanners.
250,000 acre feet to the town and *200,000* acre feet to the farmers.
300,000 acre feet to the town and *150,000* acre feet to the farmers.
350,000 acre feet to the town and *100,000* acre feet to the farmers.
400,000 acre feet to the town and *50,000* acre feet to the farmers.
450,000 acre feet to the town and *0* acre feet to the farmers.

Figure 4.19: *Example of allocation question based on a status quo allocation [Samuelson and Zeckhauser, 1988].*

For the sequential decision question, where test subjects set their own status quo in the first part of the question, the observation is the same. The question concerns a two-year decision where the test subject needs to make a decision each of the two years. The first year is forecasted to be good and the second to be bad and vice versa for the other half of the test subjects. The object is to decide upon leasing one of four different-sized airplane fleets for each of the two years. The bias is shown by test subjects who starts with a good year and leases a big fleet. Even though the next year is forecasted to be bad, they continue to lease the big fleet, whereas test subjects who started with a bad year leases a small fleet and continues to do so in the following good year. The evidence for this bias is primarily found when there is multiple options.
These findings imply that once a choice has been made by the decision maker, he is more prone to linger to that choice in the future as well, which further supports the notion of the default effect. In relation to other psychological aspects sticking with the same option as already chosen before makes good sense if the choice showed to be sufficient, then it is an easy solution to just use that same option later on. Relating to Prospect Theory, by sticking with an already made choice the decision maker is not risking anything by making another choice which potentially could be better, however it does not count as much as if it does not prove to be a better choice, as losses loom larger than gains.

In order to compare with real life situations two external field studies are as well examined in order to validate the observed bias in the experimental settings. The two field studies concern employees choosing between eight health plans and retirement fund allocation plans. For each of the field studies it is evident that employees persist in the first chosen alternative, in spite of the possibility to transfer at no cost.

From these results it can be concluded that people tend to stick with the status quo option despite the underlying logic of choosing another alternative. Furthermore, different status quo set ups do seem to affect us in varying degree, both in the experimental setup put forward in this examination, but also in real life decisions concerning health plans for example. [Samuelson and Zeckhauser, 1988]
All of the results support that we seem to make inconsistent choices when faced with a default or affected by a status quo state. However, if there is a difference in the way we perceive and react to default settings is examined by [Johnson, Bellman and Lohse, 2002]. They also examine defaults, however combined with framing of questions in terms of opting in or -out. It is examined if we are aware of the values attached to the choices we do or do not make for ourselves, hence if defaults are a tool to manipulate our behaviour given framing. Two experiments are set up to examine the curiosity.

In the first experiment test subjects answer a questionnaire with an option in the end of being notified with more surveys or not. The questionnaire itself is not the central interest, but how test subjects respond to the last option. Four groups of test subjects are each given one of the option formulations shown in figure 4.20.

Question			Percent Participating
(1)	☐	Notify me about more health surveys.	48.2
(2)	☐	Do NOT notify me about more health surveys.	96.3
(3)	☑	Notify me about more health surveys.	73.8
(4)	☑	Do NOT notify me about more health surveys.	69.2

Figure 4.20: *Survey questions and results from experiment 1.*

The way the questions are posed differs in two ways; the framing (positive or negative) and the defaults (opt in or opt out). From the results it can be seen that more test subjects chose not to tick off the negative option in question one and two, actually the difference was significant which shows the effect of framing. Comparing question three and four the results are almost equal, which the authors explain by the little preset check mark which signals to the test subjects that a choice is being made and leads them to pay attention. For all of the four questions posed a default option applies, which also showed to be significant.

However, the first experiment does not take an option without a default into consideration why a second experiment is set up. The second experiment is set up in the same way this time using six questions for six different groups of test subjects, and another checkbox system called radio buttons. The six questions and the results are shown in figure 4.21.

Comparing the results for question three and six, where question three by default is opting out and six by default is opting in, the default effect is very clear since 44.2% against 89.2% participate, respectively. Furthermore, there also seem to be a framing effect if questions framed negatively with the word “not” are compared to their equivalent positively framed question. In fact, the framing provides a difference of 15% - 20% difference in participation. Also, in the absence of a default the framing effect is observed by comparing participation percentage between question two and five.
Both the default and framing effects were found to be significant, however they did not seem to interact, why they are independent of each other. Furthermore, it was tested if age,

Question		Percent Participating
(1)	Do NOT notify me about more health surveys. ○Yes ◉No	76.9
(2)	Do NOT notify me about more health surveys. ○Yes ○No	70.8
(3)	Do NOT notify me about more health surveys. ◉Yes ○No	44.2
(4)	Notify me about more health surveys. ○Yes ◉No	59.9
(5)	Notify me about more health surveys. ○Yes ○No	88.5
(6)	Notify me about more health surveys. ◉Yes ○No	89.2

Figure 4.21: *Survey questions and results from experiment 2.*

gender, education etc. interacted with the results, but no significant interaction was found, why the format of the question overrules individual differences.

It is concluded that opting in does not equal opting out, since the first experiment showed that almost 50% more of the test subjects accepted to be notified when they had to opt-out than when they had to opt-in. The reasons for the effects are considered to be either cognitive laziness or perception of the default as being the recommended choice. Based on the results and observations it is concluded that the default effect can be avoided by making the consumer actively choose an option without any presets. [Johnson et al., 2002]

Discussion

The mentioned studies all report an observed default effect, which in many cases seem to affect us to choose whatever is preset or to stay in already established settings. However, each of the articles treated the subject slightly different, providing a broader basis for understanding the effect and the nuances within it. Therefore, the main findings will be outlined in short before commencing with a discussion. From the studies six main statements can be extracted:

1. Having a default setting evokes a cognitive laziness or misperception of the default option, which in turn can lead to people not actively consider the value or consequences of a choice and thereby sticking with the default.
2. Decision makers exhibit a significant status quo bias
3. The more options provided the stronger the status quo bias
4. Status quo allocation of resources affects decisions on allocation
5. Sequential decision task are also influenced by the status quo
6. Opting in is not the same as opting out

The first statement is observed in all three studies. Whenever test subjects were exposed to a default option or a status quo, the majority chose these options. The reasons for this i.a. include the notion of cognitive (and physical) laziness. One of the points made is that it often requires an effort to make a decision, whereas leaving it to the default is effortless, thus making it easier to go with that. With regards to the physical laziness Johnson and Goldstein (2003) mention that the physical act of filling out a form may serve as yet another incentive to accept the default.

Another reason mentioned across all three reports is the effect of the default itself; people perceive it as the recognised or even recommended option to choose. If this is in fact true, it can be critical for decisions such as in the donor case.
The second, third and fourth statement in particular relates to the study by Samuelson and Zeckhauser (1988). They discuss multiple reasons for people to stick with the status quo in spite of it seeming inconsistent. For example they argue that previous experience and learning play a central role in some decision cases; if a person has learned from previous decisions in a similar context that this was a good choice, why not make it again. Depending on learning and previous experience to make a choice might in turn also make the choice less cognitively straining. In fact, the cognitive cost of switching might not be topped by the outcome of the switch, which then causes people to stick with the status quo because the utility of that is sufficient. Also, the initial decision have cost some time and effort to make, which is spared to do again by making the same decision again.

Loss aversion (see section 3.1) is also considered to be of relevance since losses loom larger than gains. Thereby it might be evaluated that the potential loss of switching option will weigh heavier than the potential gain.
The effect of sunk cost (see section 4.1) is also considered to play a role in sticking with a previously chosen option. Even though economists claim that people make decisions on a pure cost-benefit basis, the sunk cost effect reports otherwise. If there has been a past resource investment in a choice it might prove harder to give it up, since it would feel as loosing the energy spent on that choice.
Regret aversion is as well a possible explanation closely linked with the sunk cost effect. Individuals tend to avoid the feeling of regret, which furthermore has shown a tendency for people to favour inaction over action [Zeelenberg, Dijk, Bos and Pieters, 2002]. In this case inaction would equal sticking with the status quo. In relation to this, choosing the default option may in some cases also relieve the decision maker if later consequences of a choice do not turn out as expected. In such a case the decision maker can "blame" the default option which he simply followed.

The sixth statement relates to the findings of Johnson et al (2001). Their results showed that opting in is not the same as opting out, however, the study made by Johnson and Goldstein (2003) did not show this effect. Instead it was indifferent whether the default was to organ donate or not, in fact the donation percentage was almost reversed. Therefore, it is assessed that this observation might not be valid in all cases. To make a choice towards organ donation requires more thought than deciding on whether to participate in a survey or not, why some of the previously mentioned reasons might affect the organ donation scenario and not the survey scenario. For example it is more convincing that a default on organ donation makes it a big decision, if it is default not to be an organ donor. Then it is up to the individual to analyse how one would feel about donating a heart or bladder after death, which can create a feeling of cognitive overload thus making the default the easy option. Also, it may seem as the recommended option. In contrast, if it is the default to donate, the choice has already been made and by doing nothing one would not have to analyse the situation. Furthermore, framing was used as an attribute in the study by Johnson et al. (2001), which was found to be additive, why this might have caused a difference as well.

Having established that we seem to choose according to- or affected by defaults, the resulting question is what consequences this behaviour entails? As illustrated by the organ donation study, adhering to a default setting can impact other people's lives that could be saved if the default here in Denmark for instance had been to organ donate. That we are biased by defaults to actually have reversed preferences in some cases implies that we are not always aware of the active deselection of other options we make. Of course it is not possible to account for all possible options we deselect when sticking with the status quo however, in user interfaces a fixed set of options are generally presented with one default setting for

example. Given a set of options and not considering the deselections made actively might impair the use of the product, as there is a reason for providing multiple options. On the other hand there is no reason not to consider the default setting as the standard, since it has been set as default.
However, in some situations it can make sense to spare cognitive resources for a choice which has been made before, and instead release resources for other matters. In such situations defaults in user interfaces can be used actively to relief the user of having to make a choice, when there is a need for the user's focus elsewhere. This could be applied in repeated user interaction situations, and maybe as a consequence of repeated choices the system interacted with could learn which default settings suit the user, and thereby support use even further.

men det er også kun når der er et limited set of options. det kan også give fin mening i nogen henseender at spare på tænkningen, så vi kan bruge vores ressourcer andre steder. Netop på sidstnævnte måde vil defaults kunne bruges aktivt til at lade os guide så vi har frigivet cognitiv kapacitet til at klare anden tænkning for os. hvis det vrker har vi så et behov for at det virker endnu bedre

The default and status quo bias seem to be affected by multiple psychological elements, which frame the way we act accordingly in most cases. In terms of applying defaults in user interfaces, it is already done. However, the basis for doing so is more questionable as all the experiments examined in this section relied on questions, not actions as such. In a user interface there will most often be a purpose of the use, why the motivation for using a product is different from deciding to become an organ donor or not. The latter question can appear out of reach to somebody, whereas using a product is done on the basis of obtaining a goal, hence the decisions might show to be of more importance to the user, why defaults might be undermined. Therefore, it could be interesting to examine whether the default effect can be observed in a user interaction situation, or if users are motivated to commit themselves in choosing the best option for themselves.

4.4 Recapitulation

In this section the main findings from the literature studies will be recapped in order to provide an overview of the aspects which have usability implications for interfaces and are going to be implemented in the following experiments.

Sunk time

Sunk cost is established as a psychological mechanism, as it has been verified in multiple studies. However, sunk time (and effort) is more debated as some results show that a sunk time effect exists, whereas others reject it. However, as discussed the research method may play a role in these contradicting results, as it appears that time needs to be wasted before it is perceived as an "investment" in the same way as costs.
When effects of sunk time were found, test subjects displayed the bias by continuing performing the task they had already invested time in. As to user interfaces this holds the implication that users, upon investing time in a program for instance, are not as susceptible to give it up again, why a "loyalty" towards the program can emerge if time to learn the program has been spent.
However, implications towards perceived time were also shown to have an effect on the sunk time effect. Entertained test subjects acted as test subjects where no time had passed, whereas test subjects who were bored showed clear signs of sunk time, as they demanded more in ultimatum- and dictator games. Whether time was actually perceived differently in terms of minutes between the groups was not found, why the different behaviour might as well have been caused by the fact that time was wasted in one of the scenarios, whereas it was exploited in the other.
In relation to use situations this aspect of sunk time is interesting as the sunk time effect in turn can lead to different decisions. For example it could be that a waiting time in a user interface would lead to the user demanding more of the product (as seen in dictator and ultimatum games). Also, the waiting times associated with a product might be diminished if the user is entertained during the wait, which in turn would lower the anticipation and improve the perceived utility of the program, in relation to the user having wasted time on the program.
If sunk time also applies in use scenarios, and it is affected by perceived time, could prove to be an important consideration to include when designing a user interface, which requires loading times for example. Therefore, the sunk time effect and perceived time impacts are examined in the following sunk time experiments.

Framing

In general framing is the concept of portraying the same fact using different means, however the true content of that fact is not changed. Mainly the literature has focused on framing by wording, where a set of equal options are described by different words, most commonly in terms of gains "save lives" and losses "lose lives". The findings made by Thibodeau and Boroditsky (2011) showed how one word can affect us to judge and decide differently, why word priming might also work in certain user interface environments such as a website. For example labelling a "buy" button "spend your money" versus "Invest your money" could potentially have the same effect as observed in the observational study made by Ganzach and Karsahi (1995). Despite this subject being very interesting it overlaps the area of marketing and would potentially be hard to apply in a user interface where buttons and other setting types are more instructional as "on", "off", "next" etc. Therefore, it would be interesting to examine if other more "physical" attributes of a user interface holds the ability to frame user behaviour, for example the visual domain as this is key in most user interfaces. Also, the wording provided to users is an aspect often taken into consideration when designing

a website or an interaction product, however, the designer might not be fully aware of the strong effect the wording can have.
Concerning the "physical" aspect, the study made by Hsee (1998) concerned framing in the sense of manipulating a physical reference point - an ice cream cup. This framing method is also found interesting in terms of manipulating use settings, which in turn may affect user choices. What is thought of in this case is directly transferring the ice cream experiment onto a "physical" setting tool commonly used in user interfaces, in order to examine if the same effect occurs given the circumstances of a use situation.
By actively applying the concept of framing in a user interface, user behaviour should change as a consequence, if the observed behaviour patterns found in framing research also apply in use settings. Based on this curiosity a set of experiments will be performed in order to examine if it is the case.

Defaults

Defaults and status quo have been shown to affect vital decisions such as organ donation, and multiple psychological processes seem to affect our decision patterns when it comes to defaults such as cognitive laziness, prospect theory as making a decision involves risk, framing etc. The default effect is interesting in terms of user interfaces, since most user interfaces come with some sort of default setup, or during the setup, some settings are made by default by e.g. a ticked checkbox. How do users react to these presets when presented to them in use situations? Do users "obey" the defaults and accept them as standard settings, or do they actively decide? And is there a recognition effect involved as discussed by [Samuelson and Zeckhauser, 1988]?
These questions are interesting as default settings may prove to be of great importance, if users (test subjects) do not make active choices and blindly believe that the default is the "true" or standard setting. If it is proven to apply in use cases then defaults can also be used actively by producers of software and physical consumer products to steer the user in the direction which is either best for the user or the company. It seems that this use of defaults is already applied as seen in program setups for example, where users often are presented with the options of standard- and advanced settings, where the preset is commonly the standard. Thereby users are supported in setting up the software as the standard settings are the most vital ones, presumably. However, if in fact users do follow default settings is interesting as to further application; if users follow defaults then they can be manipulated or steered in right directions, whereas not considering default settings might hold the consequence of steering users in the wrong direction unintentionally. Therefore, this concept will be examined further in the following experiments.

CHAPTER 5

Experimental paradigms

Experimental setups for studying decision making will be discussed in this section to form a basic understanding of the construction of decision making experiment setups and the implications involved. Examining previous experimental setups provides inspiration for how to test decision making in another domain. When the word "frame" is mentioned in this context it primarily refers to the experimental frame, not framing within the experiment as section 4.2 concerned.

Experimental paradigms used in classical psychological decision making research consists of forming a decision frame within which it is possible to evaluate a choice as either wrong or right, given a rigid statistical frame (as seen in i.a. section 3.1). Research made by Kahneman and Tversky often utilises this sort of decision making frame setup, which is illustrated in figure 5.1. A loss and a gain frame are constructed containing the same kind of gambles. Test subjects' answers are then compared between conditions in order to determine the differences in choice.

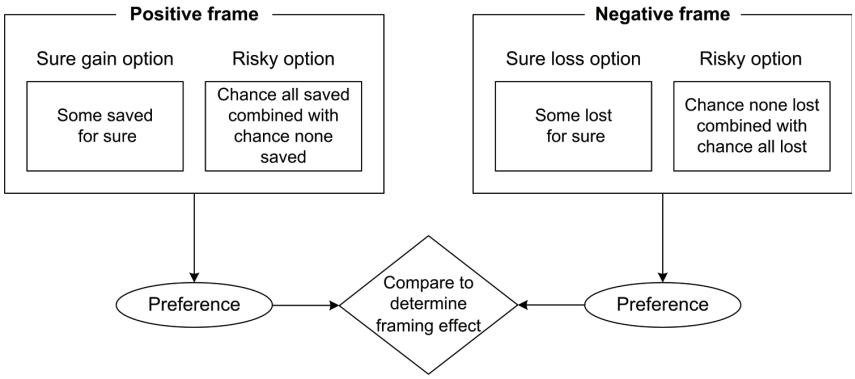


Figure 5.1: *Experimental paradigm often used for testing decision making in terms of framing.*

The following example illustrates such a setup [Tversky and Kahneman, 1981]:

Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

Decision (i). Choose between:
A. a sure gain of $240
B. 25% chance to gain $1000, and 75% chance to gain nothing

Decision (ii). Choose between:
C. a sure loss of $750
D. 75% chance to lose $1000, and 25% chance to lose nothing

This experimental setup is used to study choice under risk, and whether test subjects act according to the predictions of Prospect Theory or not. Given the frame, a very limited set

of external factors will affect the choices of the test subjects (such as mood, fatigue etc.), whereas more rich contextual stimuli such as visual or audio influences, as well as purpose and incentive effects are very limited and the hypothesis is somewhat easily tested.
The level of analysis has profound consequences for how the conceptualisation of decision making is made. The level of detail obtained in such an experimental setting comes with the cost of not being able to see the phenomenon within a more complex context, why it is difficult to predict if the observed phenomena in such settings have further applicability.

However, simulating a real life decision making context where multiple factors are in play in an experimental setting has proved to be cumbersome. Testing "just" one anomalie of choice at a time in a limited framework makes it relatively easy to objectively conclude upon however, not necessarily easy to transform to a use context experiment. Nonetheless, the principle of e.g. framing experiments in a simplified framework can function as a theoretical framework for setting up an experimentally bounded setup.

Setting up this kind of narrow framed experiments is in some cases making the content within the frames rather peculiar in order to test some of the decision making phenomena as "realistically" as possible. An illustrative example of this is a study made by Navarro and Fantino (2008) where test subjects are posed hypothetical questions as in Tversky and Kahneman (1981). The following question is posed in order to examine a sunk time effect:

Imagine that you are the leader of a copper−mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek" Today is the first day of the dig. Right away this morning, you and your group have found a vein of copper! However, there is a thick wall of quartz, which is easy to dig through, covering most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small − about 10 pounds. Typical mines contain upwards of 500 pounds. You have a choice to make:

A. dig the 10 more days to collect the 10 pounds of copper.
B. abandon this "Shady Creek" mine and go home.

Testing decision making, and especially a complex phenomenon as sunk time is hard, since the motivation in an experimental setting is different from that in a real world setting. Therefore, peculiar setups as the above emerge from attempting to test a phenomenon experimentally which is believed to exist in real life, but would however require a long term observation setup. The optimal test setup in which to test sunk time would be to observe a set of test subjects through their lives and observe in which instances they have wasted an amount of time and following made another choice than they would have otherwise. Moreover, establishing which choice the test subject would otherwise make with no influence of sunk time would not be apparent in such an observational setup. The problem is that no guarantees are given with regards to observation setups, why it is less cumbersome to try to set up an experimental "imagine paradigm" frame setup. However, as a consequence the content within the frames is at risk of becoming peculiar in order to provoke the effect.

Furthermore, most of the setup experiments rely on a set of binary choices, where the test subjects are forced to choose one of the given options. Such a setup does not leave much room for observing how test subjects could otherwise react in such a situation. Maybe they would come up with a third option outsmarting the problem stated, or maybe they would not choose at all. Given such rigid frames it can be doubted if it is in fact actual human decision behaviour which is observed, or if it is reduced to decision behaviour defined by the setting.

To ask test subjects to imagine a scenario and choose based on their imagination, given only a limited set of options, seems risky as the ability to imagine between test subjects may vary. Furthermore, it places all responsibility on the test subject as he needs to be able to imagine what he would choose of two choices in a situation he is imagining.
To rely on test subjects' analytical abilities can prove complicated as research has shown that when test subjects are asked to estimate a parameter, such as own arm reach, they overvalue their own capability, whereas asked to estimate the same parameter as a sub task of a primary task, test subjects are able to estimate the parameter almost precisely [Heft, 1993]. This suggests that placing test subjects in an analytical situation does not reflect their complex perception-action processes which seem to operate in a different more integrated manner than the analytical process, which is tested in these rigid experimental setup types.

Within the different subjects of decision making different experimental paradigms have emerged as a consequence of the decision phenomenon itself. An example is the sunk cost and sunk time research area, where three general ways of testing the properties of the effect have emerged. All are based on the test subjects concluding the experiment by making a choice however, this is mediated in different ways. The first relies on engaging the test subject in a task, which spans a certain amount of time. When the test subject has performed the task for a certain period of time, he is exposed to a choice where he can give up the task he has just spent time on, and continue on with another task and get off earlier from the experiment, or continue with the current task. The idea is then to observe if the test subject is inclined to continue performing the original task, as he has invested both time and effort in that task. The second test paradigm is basically constructed as the latter, however instead of choosing to continue or not, the test subject is offered a reward upon completing the task. The rewards are of different value and the idea is then to observe if test subjects choose the law value reward, which they have spent most time earning, as a consequence of sunk time.
The third test paradigm is build upon gambling in the form of ultimatum and dictator games, where it is examined if test subjects decide differently in these games upon having spent a certain amount of time.
This example shows how test paradigms have evolved as a consequence of trying to reveal choice mechanisms in artificial environments. It also exhibits how decision making matters are complex and often needs to be examined from different perspectives, as decision making is based on many differing internal factors such as knowledge, experience, mood, circumstances, motivation etc.

Tetlock (1985) discusses the decision making research paradigms based on a microscope metaphor, with which he argues that at intense levels of magnification the detail is observed, but as a consequence contextual premises of the phenomenon observed are diminished. However, at less intense magnification levels the observer gains the ability to observe the phenomenon in its context however, at the cost of not being able to observe the subsystem in detail. [Tetlock, 1985]
The less intense levels of magnification refer to field study observations for example, where the decision phenomenon is observed in a context, and not only in a diminished frame. The microscope metaphor highlights how it is assumed that a decision can be reduced to an analytic unit by offering test subjects one premise and a set of constructed options, on which a conclusion of decision making processes is based.

Another tendency within the decision making research however, is to zoom out the microscope to less intense magnification levels as ecological valid field studies are made, which functions especially well within the field of marketing and decision making. Iyengar and Lepper (2000) made an observational experiment in an American supermarket, where two tasting booths were set up. The behaviour of customers was observed with regards to the

too-much-choice phenomenon (see Appendix I) [Iyengar and Lepper, 2000]. Ganzach and Karsahi (1995) also made an observational field study, where they tested if framing a message with regards to credit card use, would affect consumers credit card use [Ganzach and Karsahi, 1995].
Using an observation paradigm reveals more about actual behaviour than the previously mentioned narrow framed experimental setups. The results from observational studies are more readily applicable since they are extracted from human behaviour in the environmental context itself. Therefore, to strive for such an experimental paradigm seems optimal, but as mentioned before, also contains implications if the phenomenon is not observable over shorter periods of time, for instance.

To be in favour of either of the experimental paradigms; the narrow framed or the observational ones, is ambiguous as both types induce different advantages and disadvantages. The narrow framed experiments reduce decision problems in the test setup to only encompassing two factors; the question posed and the answer options. Based on these factors it is relatively easy to conclude upon the result as no complex measures are counted in. On the other hand the premise for actually being able to claim that humans act according to the findings are somewhat questionable. As Gigerenzer (2000) discusses, the experimental framework of i.a. Tversky and Kahneman is mainly build on the conception that statistical solutions to problems are the only ones valid, why they continuously are able to observe that test subjects act and judge "irrational". However, Gigerenzer argues that these statistical solutions are not the only correct answers to the problems put forward, why it makes little sense to utilise the difference between test subjects' judgments and decisions and the statistical answers as a basis for describing human behaviour [Gigerenzer, 2000]. When acting and deciding in real life situations probabilities are not attached to the surroundings as explicitly as given in these types of experiments, and maybe we do not even assign probabilities to actions as is suggested in some of the classic decision making research. In that connection Gigerenzer adds a perspective to the notion made by Tetlock (1985) concerning detailed experiments missing the broader context; that narrow framed experiments forget to take into account what is natural behaviour and instead concerns itself with rigid statistical frames in which one answer is perceived as correct and the other incorrect.
With regards to observational experiments they encompass the opposite of the narrow framed experiments as they include a much richer context in which decisions are more likely made on an everyday basis. However, being able to control all parameters in observational experiments is more or less impossible, why disturbing elements might influence the results unexpectedly. Furthermore, observational experiments require more time and effort to set up, while at the same time there is a risk of not being able to observe anything if some circumstances not accounted for appears.
Therefore, concluding on one best practice experimental paradigm in the present project is ambiguous as it is desired to examine if the observations made in isolated setups, do also apply in more broad, complex settings. On the other hand time and resources are factors which also need to be considered, why narrow framed experiments are advantageous given such considerations. However, for the following experiments in the present project inspiration is found within both types of test paradigms and is going to encompass a compromise since ecologic, observational setups require a long time span, whereas the narrow framed experiments disregards complex stimuli. The experimental setups will resemble that of Hartmann et al (2008) where a compromise is made; the frames are narrow in the sense that the phenomenon of framing is tested using only two different worded frames, but visual and interactive stimuli is added as well.

Summary

Decision making is a natural ongoing process, which makes it challenging to test in a non-ecologic setup, since the motivation of the test subjects in an experimental frame most likely

will not resemble natural motivation, and might even change over shorter periods of time. Also the motivation for an individual choice can differ as the choice can be affected by prior knowledge, why some test subjects choose differently than others who have no experience on a certain subject. To motivate test subjects is therefore even more important in use setting experiments as the goal is not important for the test subject as it would be in real life if a decision towards using a specific object was made, in order to accomplish a goal. This will serve as a great challenge, and to some extent nuisance variable, which will be discussed for each of the experiments in the following.

The following experiments seek to uncover if specific decision patterns can be observed, however, it will be in the form of quasi experiments as many uncontrollable variables will undoubtedly affect the results. As mentioned, test subjects' different knowledge and experience will most likely affect some test subjects to make different decisions than others. Adding to this is the motivation of the test subjects, who really do not "risk" anything in an experiment, why the motivation will not resemble that of a real use situation where there is an individual purpose connected with a product interaction, and thereby which decisions are made. Factors such as these will be uncontrollable, but on the other hand if an effect of the experiments is observed nonetheless, it will support the hypotheses even more.

With these considerations in mind two preliminary and nine main experiments were set up which will be outlined one by one, subject by subject.

CHAPTER 6

Sunk Time Experiments

In this chapter two main sunk time experiments of four are presented. The two experiments not presented in this chapter are preliminary experiments, which are not directly corresponding with the thesis statement why they can be found in Appendix II instead.
The sunk time experiments presented in this chapter build upon each other. First it is examined if a sunk time effect can be observed in a user interaction scenario, where after the effect of perceived time is examined in relation to evaluation of usability parameters.

6.1 Experiment ST1

Indications of sunk time were found by Navarro and Fantino (2008) and will be examined further in this experiment in a use setting. If there is a sunk time effect in relation to a usability setting, it can affect the way users judge and decide. Therefore, an experiment is set up to examine if test subjects in a use setting are affected by sunk time. The experiment setup is based on an "ultimatum game" test paradigm to examine if test subjects do in fact make different choices based on time spent in a use process, given a use frame. An ultimatum game test setup is chosen where the test subjects are going to "gamble", contrary choosing between two options half ways. The latter seemed strange implemented into a use setting, where a user is seldom asked to choose between two alternatives during use. The choice of an ultimatum game test setup is sought transferred to a use setting as described in the following.

Experimental objective

as opposed to an imaginary task setup, the objective of the experiment is to set up a use based test environment in which test subjects will be exposed to different time conditions. Subsequently test subjects are to make a decision and the experimental objective is to measure if this decision is influenced by the previous time condition.

Hypothesis

The general hypothesis for this experiment is based on the findings from Oxoby and Bischack (2005), where it was found that test subjects who had spent (wasted) time before an ultimatum game were more inclined to demand more money. Therefore, in this setup the general hypothesis is that:

The amount of time wasted during the use process affects users' subsequent choice

As mentioned, the experiments made by Oxoby and Bischak (2005) were based on a variation in waiting times, why the difference in time spent in this experiment will manifest itself in a waiting time as well. If the hypothesis is accepted this will imply that waiting times can act as an influencing factor on subsequent choices made by a user, which for some user interfaces can prove critical.

Apparatus

The test setup is made in Visual Basic 2008 Express Edition, and is run on a MSI 340x laptop with a 13.4" screen. The test is made in a group room at Aalborg University with the curtains closed. All stimulating objects, pictures etc. are removed from the room, since test subjects should not be preoccupied during the waiting time. Camtasia, a screen capture program is used to record the screen during the test.

In order to be able to keep the time spent for each test subject constant, an experimental setup encompassing a clear course of action is made. The setup is made using the context of a factory where a production machine needs to be configured; in that way test subjects' course of action can be controlled as the machine settings are done according to different steps. More specifically, test subjects are asked to configure the settings of a machine for production of plastic containers for storing food. They are to imagine that the computer they face is the computer connected to the production machine.
In this way it is possible to "guide" test subjects to perform the tasks in the same order, since they are told that they have to adjust the machine settings in certain steps. When the test subjects are finished adjusting the settings they are exposed to one of two conditions; the first test group is to wait 10 seconds while the computer configures the settings just made, whereas the second test group is to wait 2.5 minutes while the computer configures the settings (see figure 6.1 for an experiment overview).
The waiting time of 2.5 minutes is based on a subjective assessment of how much time is needed for a test subject to feel that time has passed, while at the same time not pushing their limits as they have to look at the screen saying *"Maskinen skal nu tilpasse sig de indstillinger du har valgt. Dette tager et øjeblik da det er første gang den skal indstilles. Vent venligst..."*. The "Vent venligst" sign is flashing in order for the test subjects to be able to see that the computer is in fact still working and not freezing. Furthermore, two pilot studies were made in relation to the time to test if test subjects give up beforehand, which would otherwise ruin the point; both test subjects "survived" the wait why the time is kept at 2.5 minutes.

Following the waiting time, each test group is presented with the same task, where a scenario is put forward: A test container has just been produced by the machine according to the settings made by the test subject. The production executive will now expose the container to a temperature test to measure the quality of the plastic concerning the containers ability to handle heat from i.a. a microwave oven. The production executive will provide feedback about the temperature at which the container melted.

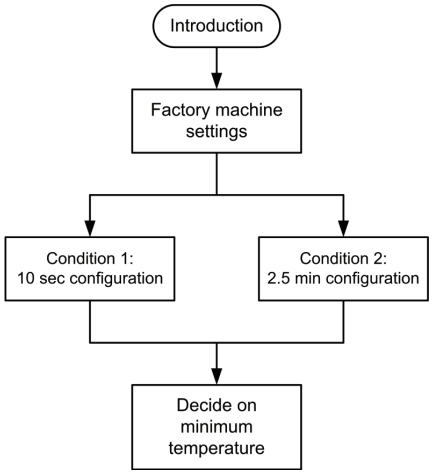


Figure 6.1: *Experimental course overview.*

Test subjects are provided with a slide bar (denoted TrackBar in VB) of length 500 pixels and 1000 steps. The endpoints of the bar are "50 degrees Celsius" and "190 degrees Celsius", and the default setting is 50 degrees Celsius in both conditions in order to avoid affecting test subjects decisions differently between the conditions (see figure 6.2). Test subject are to adjust the temperature they as a minimum are willing to accept as melting temperature for the plastic container. It is chosen to have test subjects answer using a slide bar in order to be able to register small differences.

Figure 6.2: *Slide bar used to judge minimum accepted temperature limit.*

Before adjusting the slide bar test subjects are informed that as a consequence, if the plastic container does not comply with the minimal acceptable temperature they have put forward, the setup of the machine must be adjusted again to improve the quality of the plastic containers produced by the machine.
It is this adjustment which resembles that of the ultimatum game, where test subjects have to assess how much money they are willing to gamble. In the same way test subjects in this setup need to set a minimum temperature limit, which they are willing to accept in order not to restart setting adjustments.
Since test subjects who have waited 2.5 minutes might not want to reset the settings as they expect that the configuration of their changed settings once again will take 2.5 minutes, they could be inclined to set their acceptable minimum much lower than they otherwise would have. Therefore, they are additionally told that the first configuration time is longer than the following, why a second configuration will only take 10 seconds. In this way the prospective time should not be a factor.

The experimental setup is a between subjects experiment, as one group of test subjects are exposed to 2.5 minutes of configuration time, whereas the other test group is only exposed to a 10 second configuration time. The independent variable is therefore the waiting time (short and long), whereas the dependent variable is the minimum temperature accepted by test subjects, adjusted on the slide bar.

For the developed program , see Attachment A and the attached DVD.

Procedure

The test subjects were placed in the group room in front of the computer. They were asked to remove their mobile phone and watch in order not to disturb the test, as well as they were told how the test would progress (See Attachment B). An introduction text was placed besides the computer (see Attachment C), which the test subjects were asked to read. The experimenter then pressed a computer button which initialised Camtasia (screen recording), without the test subject being aware. Following the test subjects were left alone in the group room, and were asked to come out when having pressed the "Afslut" button.

Afterwards, the test subjects were rewarded with some cake and asked to sign a declaration of consent (see Attachment D) since their mouse clicks had been captured. All test subjects signed the declaration.

Test hypothesis

The test hypothesis for this experiment is based on test subjects' minimum accepted temperature estimate. It is expected that test subjects who only wait for 10 seconds are not as

demanding as test subjects who wait for 2.5 minutes, as a consequence of the time put into the use process. The hypothesis therefore states that:

H.ST1.1 *Test subjects exposed to the long wait condition request significantly higher minimum acceptable temperatures than test subjects exposed to the short wait condition*

N.ST1.1 *Test subjects exposed to the long wait condition do not request significantly higher minimum acceptable temperatures than test subjects exposed to the short wait condition*

Participants

40 students participated in the experiment (mean age = 23.975±3.13 years). The gender distribution was 17 women and 23 men. None of the participants in the long wait condition gave up during the experiment, and no one reported on any irregularities.

Results

The result from this experiment is in the form of the test subject's minimum temperature demand. In order to translate the slide values chosen by the test subjects to a degree Celsius quantity, the following equation is used:

$$50° + \left(v \times \tfrac{140}{1000}\right) \text{ , where v is the slide value.}$$

The data is parametric and normally distributed as seen in figure 6.3. In order to test the null hypothesis N.ST1.1 a two sample t.test is therefore used, which examines if the means of distributions between conditions are significantly different. With a mean of 118.24° for the 10 sec. condition and a mean of 140.78° for the 2.5 min. condition, the p-value of the t.test shows that hypothesis N.ST1.1 is rejected (p=0.007766), whereby H.ST1.1 is accepted. The difference between the groups corresponds to a 19% difference in accepted minimum temperature on average. A graphic overview of the data is presented in figure 6.4.

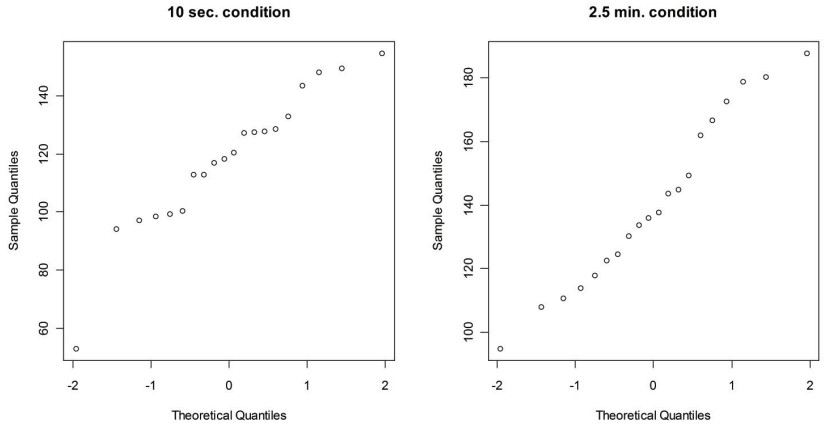


Figure 6.3: *QQ-plots of the stated temperature minimum between conditions.*

As can be seen in the boxplots, an outlier appears in the 10 sec. condition. By removing this outlier the p-value in a t.test is diminished, however it is still significant (p=0.01439), why no further examination is made.

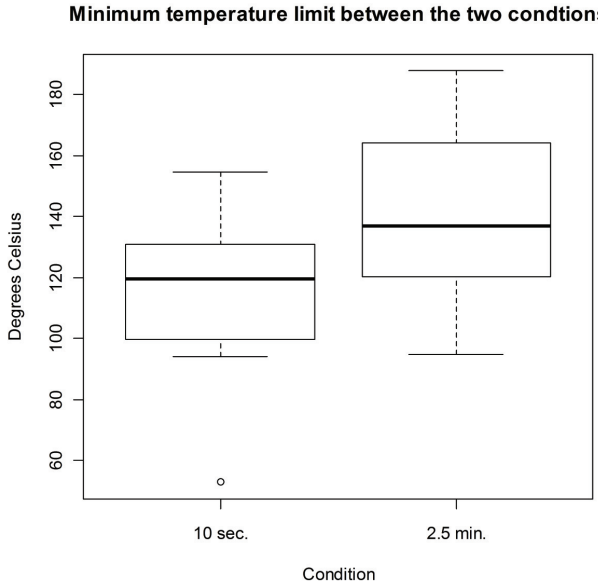


Figure 6.4: *Boxplot showing the distribution of temperature limits between the two conditions.*

In order to test for gender inferences a two sample t.test to test for effects of gender on temperature minima is applied. According to the results gender does not affect the results (p=0.5014).

Discussion

The results show that sunk time does seem to have an effect in a use based setting, as the hypothesis was accepted. However, many uncontrollable variables may influence such an observation given the context for example. In this particular setup test subjects were asked to perform a task; they did not do it for themselves. Even though they had to make a decision given a use based context their answer would not have a large impact for them personally, why the motivation most likely is distorted. Also, the specific setup might have been hard to fully grasp as it is not ordinary to be in control of a plastic container factory. Furthermore, test subjects were all placed in a situation where they had no alternatives but to look at the screen during the configuration for 2.5 minutes. It is highly unlikely that users in the comfort of their own homes or at work for example, would not find something else to do while waiting for the configuration to complete. In this way the experimental setting is far from ecological.
However, the purpose of the experiment was to examine if user choices can be affected by waiting time, which the experiment demonstrated for the current setting. This finding is important in the context where a user is "forced" to wait, where after a decision needs to made. In fact, it seems that the time spent waiting constitutes a carryover effect for following decisions.

If waiting time leads to a different choice it is the waiting time which is of importance, and it might be that it is the feeling of "waste of time" which is the influential factor, as mentioned by Arkes and Blumer (1985). This raises an interesting question to follow through on; is sunk time affected by perceived time? And if it is the case, does this have any further implications towards a usability context? This question is examined in Experiment ST2.

6.2 Experiment ST2

In the latter experiment it was found that test subjects did alter their judgment and decision affected by the time spent waiting for a set of settings to be configured. A question concerning waiting time was raised; is sunk time affected by perceived time? By perceived time is meant the way people judge the "length" of an amount of time. For example, when queuing in a supermarket it feels as if time passes more slowly than when shopping. Therefore, it is interesting to examine if a certain amount of time spent during use can be perceived differently, in relation to "wasting" time and using time.
A great source of inspiration to this issue can be found in the computer game world, where long loading times often are inevitable. In The Sims 3 and Age of Empires to name a few, tips and description of civilisations occur while loading, why gamers can actually learn while waiting. In this way by "entertaining" it might be that the sunk time effect found in the previous experiment can be diminished. The present experiment therefore seeks to examine this aspect of sunk time.

Experimental objective

The objective of the experiment is to examine an attribute of sunk time which might have a direct impact on sunk time itself; occupation during waiting time. This will be tested in a setup similar to that in the previous experiment, which included waiting time.
However, in this experiment the design will not be based on either the ultimatum game or the "switch option" setup, as used in former sunk time experiments by i.a. Oxoby and Bischak (2005) and Navarro and Fantino (2008). Instead it is decided to examine if the perception of sunk time affects test subjects' judgment of a given user interface, as a measure of the effect sunk time has on the evaluation of a user interface, if any. These two main goals are examined in the present experiment.

Hypothesis

Despite having two main goals they still appear interconnected, since the judgment of a user interface will serve as the "measure factor" of the perceived sunk time. Therefore, one overall hypothesis is stated, which accounts for both objectives.

Judgment of a user interface is affected by the perception of sunk time

This hypothesis assumes that sunk time is directly connected to perceived time, which later test hypotheses will elaborate on.

Apparatus

The test setup is made in Visual Basic 2008 Express Edition, and is run on two laptops; a MSI 340x with a 13.4" screen and a Fujitsu Siemens Amilo si1520 with a 12.1" screen. Furthermore, an external photo editor program is used (Photo! Editor). This program is the subject of evaluation by test subjects, since it is assessed not to be a commonly known photo editing program. The test is made in a seminar room at Aalborg University with the curtains closed. All stimulating objects, pictures etc. are removed from the room, since test subjects should not be preoccupied during the waiting time.

For this experiment a more realistic setting is chosen where test subjects are told that they are about to evaluate a program (Photo! Editor). They are told that they need to evaluate

the program as a whole, why they also need to install it. The installation setup is made by copying the commonly known Windows setup guides into the program, which accommodates most program setups in Microsoft Windows (see Attachment E and the attached DVD for the program). Having "installed" Photo! Editor, test subjects are asked to perform a task in the program and afterwards evaluate it (see figure 6.5 for an experiment overview).

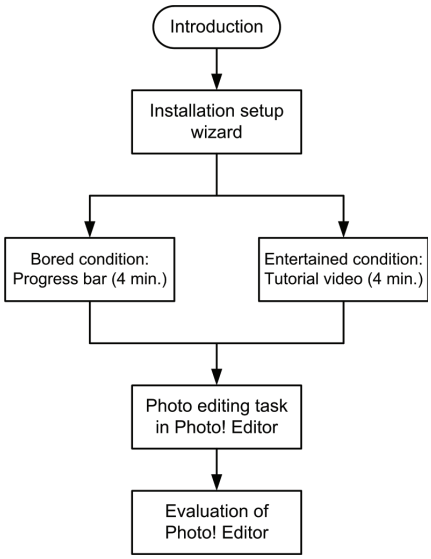


Figure 6.5: *Experimental course overview.*

The program is made in two editions where the only thing differing is the installation; in the first program an installation progress bar is shown on the screen, whereas in the second program a tutorial video for Photo! Editor appears above the progress bar and runs during the full installation time minus 10 seconds in the beginning and the end. The installation time (4 minutes) in both scenarios is the same, however the entertainment video is only supplied in one of the programs. It is chosen to show a progress bar in both conditions especially in the first one (denoted "bored" condition), as it is important that test subjects do not give up or misunderstand the installation time as if the computer has crashed. In the second condition (denoted "entertained" condition) the progress bar is also present as it will help test subjects to perceive the process as an installation process instead of a tutorial lesson.

Multiple circumstances caused 4 minutes to be chosen in this setup, as opposed to 2.5 minutes. First, both test groups will be exposed to the same amount of time why there is a risk of the results not being pronounced to the same degree as in the previous experiment where there was a time difference between groups. Second, during this test an experimenter will be present in the test room why the test subjects are easier controlled during a 4 minute wait. Third, the Photo! Editor tutorial video fitted the time span.
The chosen tutorial film does not offer any tips towards the task the test subjects are to perform in Photo! Editor subsequently. A simple editing task is to be performed by the test subjects in order for them to get a feel of the program. First they need to crop a picture followed by a light setting. In a pilot study a more complex set of tasks were posed, but pilot test subjects reported on the tasks being too complex considering the circumstances of an experiment, why the task was simplified to contain the two mentioned steps. For the task description as well as the picture for manipulation, see Attachment F.
Following the installation of the program, test subjects will be presented to the Photo! Editor program, where they have to perform the tasks just mentioned. Upon completing the task, test subjects are exposed to the screen seen in figure 6.6.

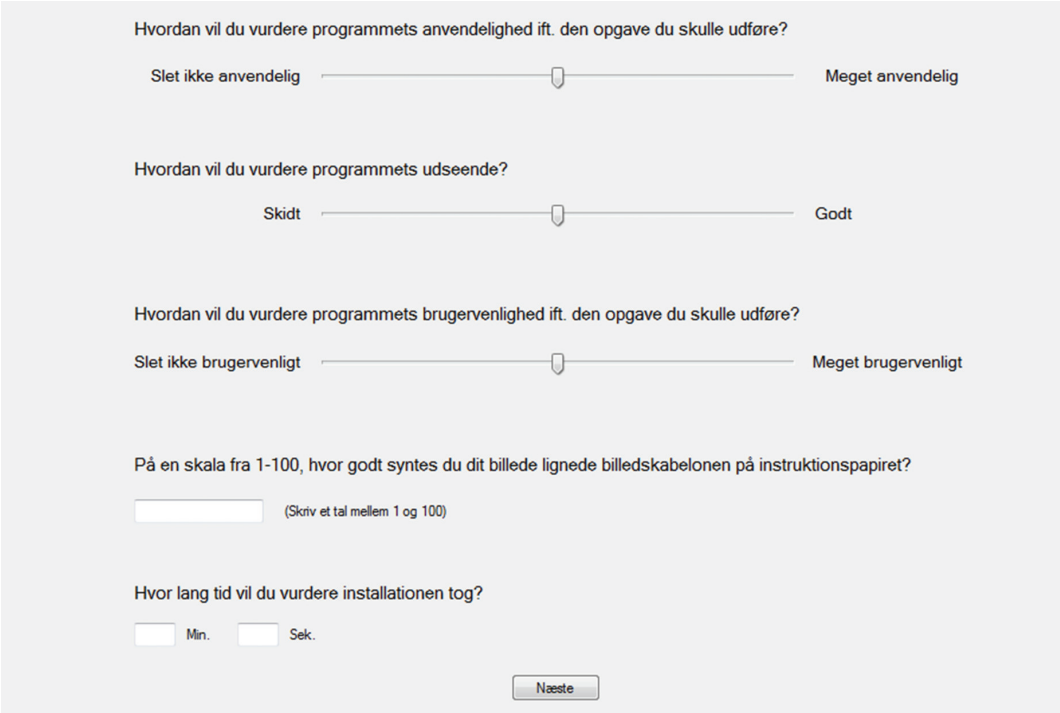


Figure 6.6: *Screenshot from the Visual Basic program.*

As can be seen, test subjects are asked to assess the program they have just explored based on three different parameters. The first is the utility of the program in relation to the given task. This question is posed as it will reveal if each test subject judge the program based upon a bad or a good experience, since some might be better at photo editing programs than others, or just in general luckier. If this utility assessment is poor due to e.g. an unfortunate experience it may mask the possible effect of perceived sunk time.

The second parameter concerns the look of the program. With this question it is examined if the sunk time effect expands into this aspect of a user interface. If this applies, the "design" of waiting times needs as well to be considered when designing a given user interface.

The third parameter concerns the user-friendliness, denoted in the rest of this experiment as the usability, for ease of reference. The test subjects are asked to rate this parameter in relation to the task just performed, since it would otherwise be ambiguous for them to judge. With this question it is examined if a difference in program judgment between the two time conditions occurs. If a difference is found, this will weigh heavily in terms of design considerations, since the usability is of great importance when it comes to e.g. reputation and loyalty of users.

All of these three questions are asked using a slide bar (denoted TrackBar in VB); only with end point anchors, and all with length 500 pixels and 1000 steps. The risk of using such a setup is that it is individual for test subjects how they understand the bar and its "steps", however, the end points are opposites why it is easy to understand that the middle position (where the adjuster is placed as default) is a "neutral" setting. By letting test subjects adjust on a slide bar smaller differences are detectable, whereas providing five boxes with a statement for instance, forces test subjects to choose one even though they might not be able to choose between option three and four. If there is a difference it should be possible to detect using slide bars.

The next question posed concerns the picture the test subject has just made in Photo! Editor. Data derived from this question will only be used to analyse and possibly discard test subjects who have had a bad experience with the task (possibly supported by their utility rating). In the final question test subjects are asked to estimate how long time they think the installation lasted in minutes and seconds. It was considered to ask how long the installation "felt" however, if the effect is present the test subjects' true estimation should still be affected by perceived time.

The experimental setup is a between subjects design, as one group of test subjects are entertained by a tutorial video during the installation time of 4 minutes, whereas the other test group is only exposed to a progress bar during the installation time of 4 minutes. The independent variable is therefore the occupation during the installation (bored and entertained), whereas the dependent variables are: rating of utility, look, usability, time and own performance.

For the developed program, see Attachment E and the attached DVD.

Procedure

Two test subjects were tested at the same time in a large seminar room where each test subject was placed in front of a computer at opposite ends of the room, facing opposing wall. An introduction text was placed besides the computers, which contained an explanation of the course of the experiment, together with the task and the picture they were to remake in the Photo! Editor program (see Attachment F).
The two test subjects were given a verbal introduction together (see Attachment B), and were asked to turn off their mobile phones and remove their watches before starting. During the experiment an experimenter was present in the room, in order to assure that the test subjects were not speaking to each other during the installation. Test subjects tested simultaneously were both exposed to the same test condition. The experimenter did also time the test subjects' performance on the task. From the moment the program started on the screen until the test subject finished the task, the time was measured.
Upon finishing the tasks and having answered the questions, the test subjects were escorted out of the room. Outside an informal interview took place, to assure that none of the test subjects were familiar with Photo! Editor beforehand, which was not the case for any of the test subjects.

Test hypotheses

The following test hypotheses are based on the evaluation questions posed, seen in figure 6.6. It is the only way to "measure" sunk time since the test subjects' actions cannot reveal their thoughts in this scenario. The overall hypothesis is therefore broken down into test hypotheses based on the measurements, in order to reveal if sunk time is related to perceived time, and if it in fact affects test subjects' attitude towards the program.

The first hypothesis concerns the perceived time, based on time estimates made by the test subjects.

H.ST2.1 *Test subjects in the bored condition assess the waiting time to be significantly longer than test subjects in the entertained condition*

N.ST2.1 *Test subjects in the bored condition do not assess the waiting time to be significantly longer than test subjects in the entertained condition*

If this hypothesis applies it shows that time perception is not constant, but changes according to the given situation. This hypothesis solely tests perceived time.
The following hypothesis concerns the rating of the look of the program. This is examined of mere curiosity if the perceived sunk time effect can extend into another domain, in this case the evaluation of appearance of the program.

H.ST2.2 *The look ratings are significantly different between conditions*

N.ST2.2 *The look ratings are not significantly different between conditions*

If this hypothesis is accepted it will also hold great implications for future design of products.
The last hypothesis concerns the evaluation of usability (user friendliness) in relation to the given task. Therefore, the task itself affects this rating directly, and thereby may diminish the measure of usability as a consequence of the time priming. Therefore, all test subjects will be screened in relation to their performance on the task, and by that it is hoped to diminish a potential effect of task and task success.

H.ST2.3 *The usability ratings are significantly different between conditions*

N.ST2.3 *The usability ratings are not significantly different between conditions*

Once again, if this hypothesis is accepted waiting times cannot be disregarded in design considerations of a product. However, implications will be discussed more thoroughly later.

Participants

30 students from Aalborg university participated in the experiment (mean age = 22.43±2.54 years). The gender distribution was 12 women and 18 men. None of the participants claimed to be familiar with Photo! Editor beforehand and none of them gave up during the 4 minute installation waiting time.

Results

During the experiment some test subjects were better at performing the picture editing task than others, why this might affect the premises on which the subsequent judgments towards the program are made. Therefore, a screening of the test subjects is made initially, based on the gathered data. In particular the test subjects' assessment of utility, their assessment of own performance, and how much time they used solving the task are of interest in assessing whether they evaluate according to the time spent or other influential factors such as task success. In figure 6.7 and 6.8 all of the measures for the bored condition are shown in boxplots, which makes it easy to identify the outliers.

As can be seen in figure 6.7 only one test subject act as an outlier (no. 5), however this test subject has not answered diverging in any of the other instances, such as concerning the utility question, why it is not considered as problematic. In figure 6.8 it is once again test subject no. 5 who act as an outlier as the time to finish the picture editing task was longer for this test subject. Again, these deviations cannot be linked to a bad experience with the task as this particular test subject's ratings of the utility and usability does not differ from the rest.

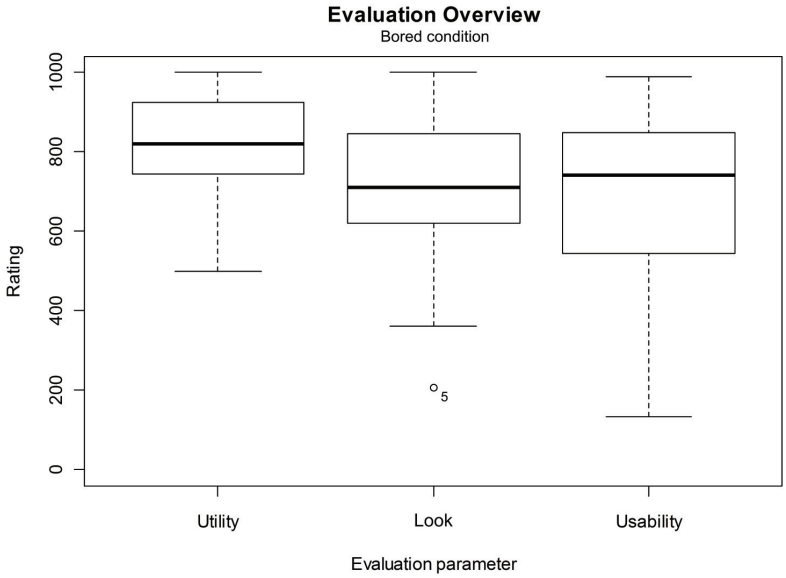


Figure 6.7: *The three evaluation parameters plotted for the bored condition.*

The procedure is repeated for the "entertained" condition as seen in figure 6.9 and 6.10. In this condition test subject no. 22 stands out by evaluating the utility rather low, furthermore this subject estimates the waiting time significantly different than the rest. Adding to this test subject 22 is the only one who evaluates own performance deviating low. This particular subject used more than 4 minutes to accomplish the task where the mean time was 2 minutes and 4 seconds. This test subject is therefore removed from the data set as it seems that the task has caused frustration, which seems to have had an impact on the evaluation.

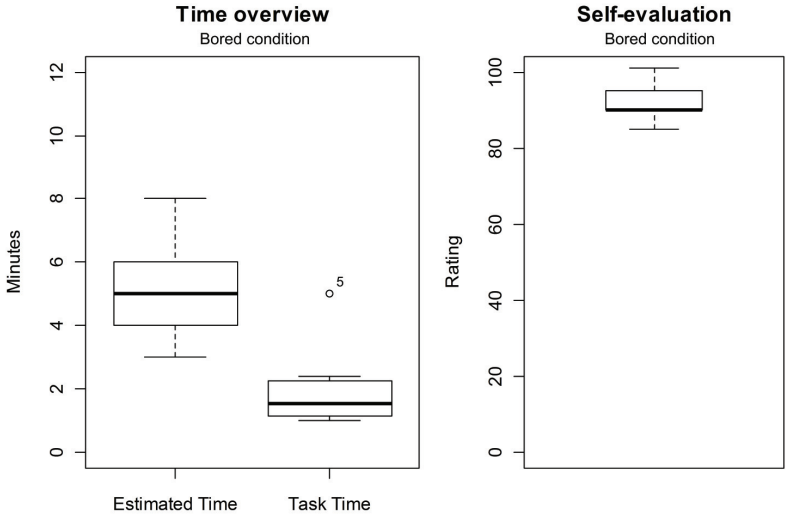


Figure 6.8: *The first time (Estimated Time) represents the time the test subjects in the bored condition estimate they have waited during the installation. The second time represents how long it took the test subjects to finish the task. The third measure is the evaluation of own performance on the picture editing task.*

In order to obtain an overview of the difference in ratings of the three attributes between conditions, a boxplot overview is presented in figure 6.11 in which test subject no. 22 has been removed. It is interesting to notice that the two groups of test subjects do not entirely agree on how to rate the program, however, if there is any significant differences will be examined in the following.

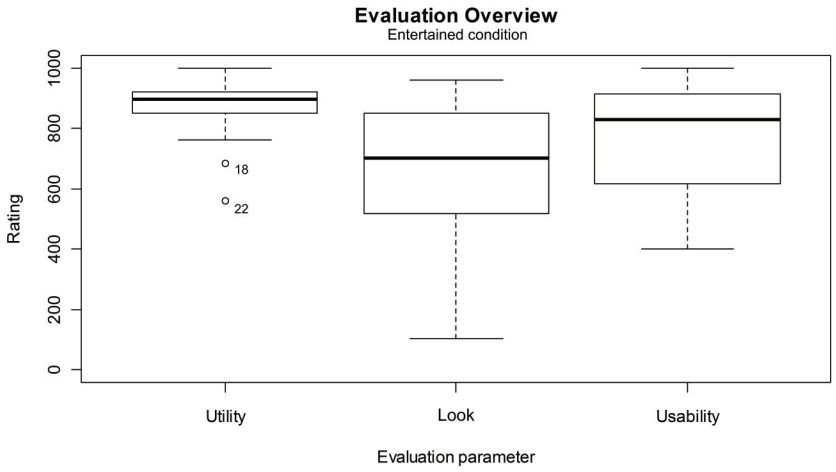


Figure 6.9: *The three evaluation parameters plotted for the entertained condition.*

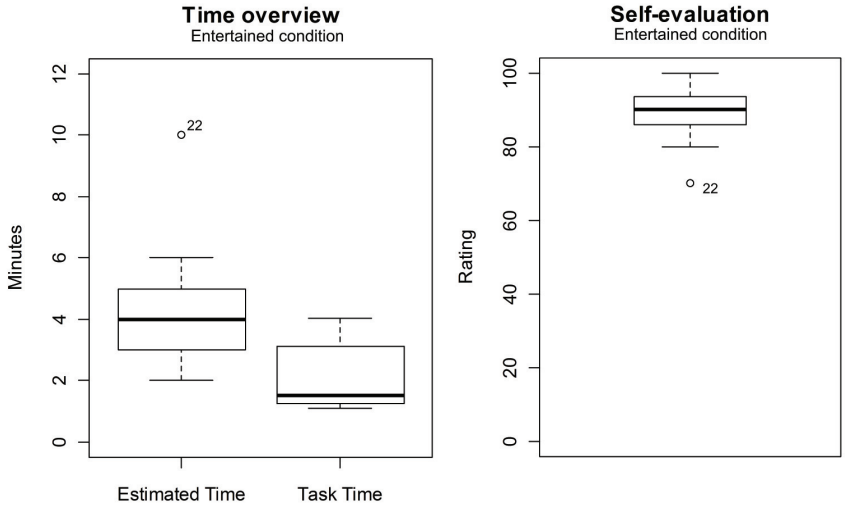


Figure 6.10: *The first time "Estimated Time" represents the time the test subjects estimate they have waited during the installation. The second time represents how long it took the test subjects to finish the task. The third measure is the evaluation of own performance on the picture editing task.*

Upon removing one test subject the test hypotheses are tested. As the obtained data of the evaluation parameters is parametric null hypothesis N.ST2.1, concerning the perceived time in minutes and seconds, is tested using a two sample t.test, in order to identify if there is a significant difference between the means of the two conditions. The data is checked to see if it is in fact normally distributed, see figure 6.12.

The data is fairly normally distributed why a t.test is performed. The result of the t.test shows that the alternative hypothesis N.ST2.1 is accepted and the H.ST2.1 is rejected with a p-value of 0.0632. Despite rejecting the hypothesis it was very close to being accepted, why an effect is still suggested as the mean of estimated times for the bored condition is 5 min. 11 sec. and for the entertained condition 4 min and 9 sec. In general test subjects in the bored condition assessed the time to be 24.9% longer.

The second hypothesis H.ST2.2 concerns the look ratings of the program. As before it is checked to see if the data is normally distributed (see figure 6.13) before continuing with performing a t.test.

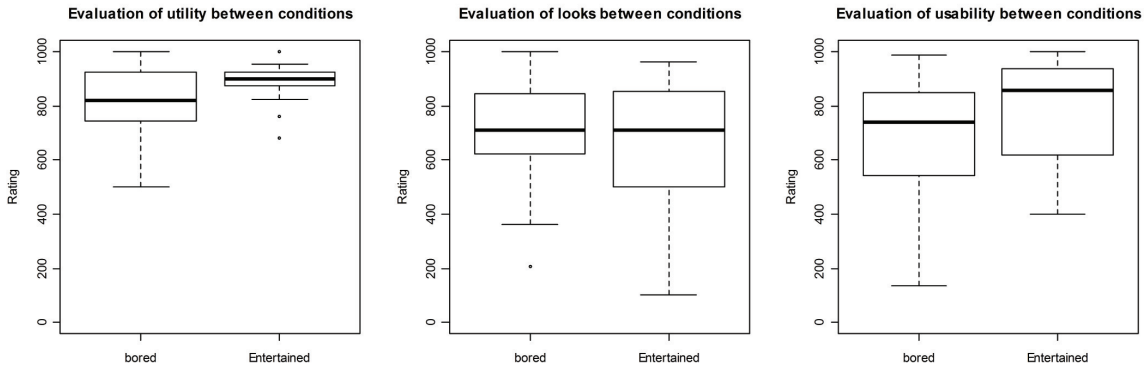


Figure 6.11: *Overview of test subjects' ratings of the three parameters.*

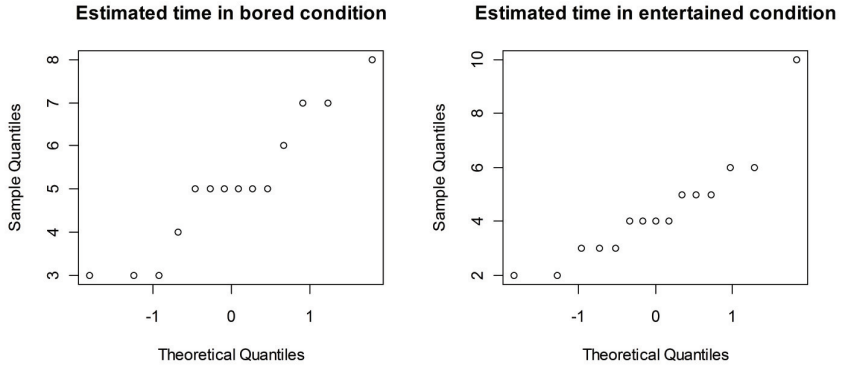


Figure 6.12: *QQ-plots for the estimated times for both conditions.*

Running a t.test shows that hypothesis H.ST2.2 also needs to be rejected. With a p-value of 0.6402 the ratings do not differ nearly significantly. This shows that the perceived time does not influence test subjects' perception of the graphic layout of the program. Furthermore, on average the look was evaluated 4.5% better in the bored condition.

The third hypothesis H.ST2.3 concerning usability ratings is tested in the same way as the two previous ones. For the normal distribution see figure 6.14. The p-value from this t.test also rejects the hypothesis, and instead leads to acceptance of the alternative hypothesis N.ST2.3 with a p-value of 0.2399. This result also suggests that evaluation of the programs' usability is not affected by a preceding time factor, however on average the usability was evaluated 9.6% better by test subjects exposed to the entertained condition.

All of the tested attributes were also tested for effects of gender, but none were found, since the lowest p-value was 0.1677.

Discussion

None of the hypotheses were accepted, why the overall hypothesis that judgment of a user interface is affected by the perceived sunk time can also be rejected for the cases tested in this experiment. However, the time was perceived almost significantly different between the two conditions, where test subjects exposed to the bored condition judged the installation time to have taken 24.9% longer than test subejct in the entertained condition.

This implies that the perception of waiting time can be diminished by adding some sort

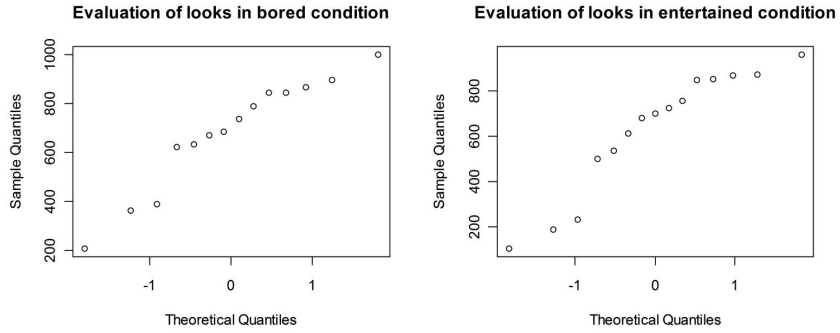


Figure 6.13: *QQ-plots for the evaluation of the program's look for both conditions.*

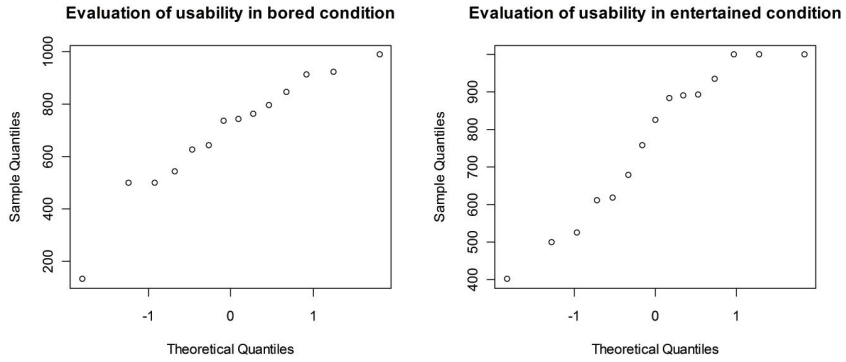


Figure 6.14: *QQ-plots for the evaluation of the usability for both conditions.*

of entertainment. Despite the different perception of the installation time test subjects did not evaluate the program significantly different, which in turn can prove positive for bigger programs with extensive setup times.

For the present setup it can be discussed whether the evaluation of a program has anything to do with decision making. In relation to a regular use decision, the setup does not supply a decision making pattern given the circumstance, instead it provides knowledge concerning the users' perception of the program. What right does this observation then have? The general idea was to see if test subjects change their perception and choices (in this case an evaluation which reflects a choice) based on a differently perceived waiting time. This was not the case why it can be concluded for the given scenario that sunk time does not cross into the domain of attribute evaluation of a computer program.

Many reasons for this outcome can be identified i.a. the program itself; the program Photo! Editor is fairly standard in its expression and interface, why the baseline for this program in relation to the three measures might be relatively high, why another program with another, lower baseline could potentially create a larger difference in evaluation. Also, test subjects might perceive the two processes (installation vs. program use) as two separates why the sunk time effect identified in the previous experiment (see section 6.2) is not carried over into the program.
If the sunk time aspect had been incorporated into a more coherent setup, not divided into steps, the sunk time effect could potentially persist, which in turn would influence choices made further on in the interaction process.

It can also be discussed if it is in fact sunk time which has been tested in the given setup, since none of the established sunk time experimental paradigms have been used (see chapter 5). The test subjects are not forced to choose to continue or replace a use process or asked

to gamble based on the time spent. Nonetheless, this experiment encompasses spent time under different circumstances and use, and places these two attributes against perception in order to identify if the perception of time influences perception of a use experience. In this way sunk time is tested for its property to extent into another domain, as it was established in the previous experiment.

The implications of the present findings together with that of the previous sunk time experiments will be further discussed in the following recapitulation section, which also discusses further test setups.

6.3 Sunk Time Experiment Recapitulation

Four different sunk time experiments have been made in the present project. However, two of them are considered main experiments, whereas the other two found in Appendix II are considered preliminary. An overview of the results from the two main experiments are shown in figure 6.15.

Hypothesis	Content	State
H.ST1.1	*Test subjects exposed to the long wait condition request significantly higher minimum acceptable temperatures than test subjects exposed to the short wait condition*	Accepted
H.ST2.1	*Test subjects in the bored condition assess the waiting time to be significantly longer than test subjects in the entertained condition*	Rejected (Nearly accepted p= 0.0632)
H.ST2.2	*The look ratings are significantly different between conditions*	Rejected
H.ST2.3	*The usability ratings are significantly different between conditions*	Rejected

Figure 6.15: *Overview of hypothesis status for the sunk time experiments.*

The two preliminary experiments on sunk cost did not directly translate to the scope of the project, nonetheless they act as a foundation for understanding limitations of the calssical experimental paradigms within decision making research. The points of the discussion in chapter 5 were supported as adding comment boxes to each of the decision problems for test subjects to explain themselves in classic imagine experiments, revealed that their choices are found upon more than a "yes" and "no" answer. These results therefore acted as a further motivation for making more integrated test setups, where the test situation resembles more of a use situation than just an imaginary one. Therefore, the following two main sunk time experiments were based on more "authentic" settings.

In the first of the main experiments (Experiment ST1) sunk time was identified, as test subjects made different decisions dependent on how much time they had waited previously. The results are very interesting as they imply that upon having invested a certain amount of time on a program, higher demands are made towards the program. For further applicability, this holds the implication that a wait time is not irrelevant in a use setting, as the following decisions seem to be affected by it. The choice in this case was in the form of a minimum accepted temperature which varied according to time spent. This might seem farfetched to try to apply as a general rule in user interfaces, that time spent will influence choices in general. However, it might not be as obvious as displayed using this experiment, as it was designed to identify such a behaviour, but it may nonetheless be important to consider as test subjects did get higher demands towards the product. On the other hand, test subjects in the long wait condition also displayed that they would be more willing to readjust the settings if not satisfied with the outcome, which also was a consequence of the waiting time.
This appears complex as test subjects exhibited a higher expectation, but on the other hand were more willing to spend further time on the settings. Therefore, the effects of sunk time do not appear to be all negative, as a higher expectation and a higher involvement in the program seems to be the result. However, concerning an evaluation of the program as consequence of sunk time was not examined why the second main experiment was put forward.

It could be that the only reason that test subjects were more involved in Experiment ST1 upon having spent an amount of time waiting, was that they have invested time. Therefore, the second experiment examined if there would be any different effects on the evaluation

of a program given the time invested. If time invested also influences the evaluation of the program, then it could prove to be either positive or negative in relation to applying waiting time to an interface, in principle. If the evaluation is influenced negatively as a consequence of time spent, then waiting times should be reduced in user interfaces. However, if the evaluation becomes more positive as a consequence of waiting time, then it seems to be positive to add, as users will have higher expectations, be more committed and evaluate the program better. Finally, if waiting time does not have any effects on the evaluation of a program, then actively applying these or not reducing them in a user interface will commit the user in the use of the program as time has been invested, as found in experiment ST1. These conclusions are in principle as it will require more empiric data to conclude for such a broad perspective.
The results from experiment ST2 did not show a difference in test subjects' evaluation of the program tested, why it supports that having a waiting time in relation to a user interface will only commit the user more, without affecting the evaluation of the program. This is good news, as waiting times consequently are not purely negative. However, in a test setting test subjects were not allowed to do anything meanwhile waiting, which they probably would do otherwise, why this distorts the results.

The effect of perceived time was also tested in experiment ST2, as one half of the test subjects were exposed to the waiting time while entertained, whereas the other half just waited. The difference in perceived time was almost significant between the conditions, with the entertained condition estimated to be shorter in waiting time. Adding this to the found results suggest that entertaining test subjects during a wait will alter their perception of time, and as a consequence the foundation for the sunk time effect is diminished. However, test subjects still did not estimate the entertained time significantly lower than the actual time of four minutes, why adding entertainment during the wait will most likely not affect the overall effect of sunk time found in experiment ST1.

Even though the experimental setups used in experiment ST1 and ST2 are not as rigid and narrowly framed as classic imaginary experimental setups (as used in experiment STA and STB, see Appendix II), they still limit the test subjects in relation to an authentic decision making situation.
Test subjects are not allowed to do anything else during the wait; they are asked to assess certain parameters and thereby explicitly consider the inputs provided, the frames in which they can act are laid out for them, as they cannot go online to check for alternatives to the given parameters for example, or whatever else a test subject would have done if it had not been an experimental setup.
All of these factors do affect the results, and for the latter experiment the program of Photo! Editor itself also establishes a different baseline for the evaluation measures than another more or less fancy program would do. Therefore, the results cannot be directly applied before examined further. Nonetheless, it was interesting to observe that sunk time was in fact observed in experiment ST1, which also indicates that the frames of a test setup is of great importance, and even though the results are only based on one experimental setup in this case, it seems as if there is one such effect as sunk time and that it can affect users.

CHAPTER 7

Framing Experiments

In this chapter five experiments will be presented. The focus of the experiments is to explore different aspects of framing tested in use settings. Several of the experiments presented are expansions of each others. In the first two experiments framing in the form of slide bar lengths is explored, whereas the three last experiments test if colours can be used as different frames to affect user decisions.

7.1 Experiment FM1

The framing paradigm serves as the inspiration for the experiments. In section 4.2 concerning framing, it was seen how people respond to a question focusing on loss compared to the same question focusing on gains, which showed that framing affects our decisions. Many of the studies reported were based on the wording of the question and the wording of the incentive to act or judge, which showed to be effective. However, one of the studies reported made by Hsee (1998) took a more direct approach in terms of valuing a physical substance based on different reference points; the valuing of an amount of ice cream was judged in relation to different cup sizes. This test setup triggered a curiosity towards testing if direct reference point framing can be utilised in a use situation.

Different test objects were discussed, i.a. a boiler kettle. By moving the maximum point for water (a reference point), it would be interesting to see if it would affect how much water a test subject would pour into the kettle. However, many complications in relation to a test situation were identified; how to incorporate the task making it seem natural, how to manipulate the maximum point on the kettle, should the task be implicit or explicit etc. Therefore it was chosen to make an illustrative test setup on a laptop presented in the following experiment.

Experimental objective

The objective of the experiment is to test if change of "physical" frames will lead test subjects to perceive and use the different frames as different reference points for judging, and as a consequence change the interaction with the test object.

Hypothesis

The idea is to test if varying a frame, and thereby providing another point of reference, will affect judgment of a given attribute. The general hypothesis based on previous findings, especially those of Hsee (1998), then states that:

The judgment of a given attribute in a use setting will vary according to the frames provided for the judgment

If this hypothesis is accepted, it implies that users are affected by outer frames for making a judgment and thereby a decision. For example, if a user judges and decides how much he

Figure 7.1: *Screenshot of the circle which test subjects are asked to judge the diameter of.*

needs to adjust a setting in a user interface his choice will change according to the frames of that setting; an example could be a volume adjuster. If the adjustment bar is long he will turn up the volume more than he would if it was short, or vice versa depending on the frame and how the effect expresses itself.

Apparatus

The experimental setup is made in Visual Basic 2008 Express Edition and run on two laptops: a MSI 340x with a 13.4" screen and a Fujitsu Siemens si1520 with a 12.1" screen. The setup consists of a slide bar (denoted TrackBar in Visual Basic) and a grey, filled, drawn circle with a diameter of 300 pixels (see figure 7.1).

Test subjects are exposed to the drawn circle for 2.5 seconds, which enables them to register the circle, without being overly exposed. Afterwards they are asked to assess the diameter of the circle, using the slide bar to represent the perception of the diameter of the circle just seen. A circle shape is chosen since a vertical or horisontal line seemed to be easily transferred to the slide bar in pilot tests. The circle is made in a grey nuance denoted "LightGray" in Visual Basic, since pilot tests revealed that a black circle left a "print" on test subjects' retinas. The grey provides less of a contrast, and thereby minimises the print effect.

The experiment is designed as a between subject experiment since one group of test subjects will be provided with a slide bar of length 800 pixels, whereas the other group will be presented with a slide bar of length 500 pixels (see figure 7.2). As can be seen in the figure the marker on both slide bars is by default placed at the far left of the slide bar. This is done deliberately in order to avoid affecting test subjects in the two conditions differently. Despite the different lengths of the two slide bars both are provided with 1000 steps and the circle diameter is held constant for both scenarios.

Figure 7.2: *Slide bars used in the experiment in lengths 500 and 800 pixels, respectively.*

An overview of the experiment is presented in figure 7.3. The dependent variable which is measured is the estimated circle diameter, and the independent variable is the slide bar lengths. If framing in terms of varying slide bar lengths is affecting the test subjects'

assessment of circle diameter, this will be measured as a difference in assessed circle diameter between the two different slide bar conditions.

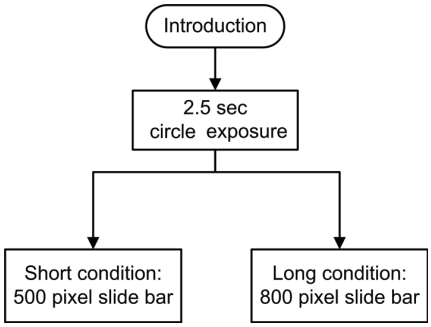


Figure 7.3: *Experimental course overview.*

For the developed program see Attachment G and the attached DVD.

Procedure

The task of adjusting the slide bar is considered simple and easily executable, and together with the fact that the experimental setup only consisted of a program run on two laptops lead to the experiment being executed in two different canteens on Aalborg University. This environment was not optimal, since many uncontrollable variables were present such as noise, various visual inputs, etc. but the advantage of the location was the large number of test subjects willing to spend a few minutes on the experiment. Before participating each test subject was given an introduction and explained that they were to solve a simple task (see Attchment B).

Test hypothesis

In this section the test hypothesis for the experiment is put forward. Since the independent variable only holds two levels, the test hypothesis will concern a comparison between these two levels, why it states that:

H.FM1.1 *A significant difference in slide bar adjustment will be observed between the two different slide bar conditions*

N.FM1.1 *No significant difference in slide bar adjustment will be observed between the two different slide bar conditions*

If framing is present and applicable in this experimental setup, test subjects who get the longer slide bar should assess the circle diameter to be different than those who get the shorter slide bar.

Participants

78 students at Aalborg University participated in the experiment (mean age = 23.22±2.89 years). The gender distribution was 35 women and 43 men. None of the experiments were discarded.

Results

To enable a comparison between the two conditions the data registered is converted from steps into pixels via the following two equations:

Long condition:

$$v \times \left(\frac{800}{1000}\right) \text{ ,where v is the slide value.}$$

Short condition:

$$v \times \left(\frac{500}{1000}\right) \text{ ,where v is the slide value.}$$

An overview of the results obtained between groups is presented in figure 7.4. As the boxplots show the median values are not equal; the long condition median is slightly higher than that of the short condition (311.52 vs. 294.95). However, more test subjects seem to estimate the diameter both shorter and longer in the short condition than in the long condition, and in average the diameter is estimated to be 311.62 in the short condition and 303.20 in the long condition (removing the outlier in the long condition provides a mean of 307.09 and a median of 312.04).

The result clarifies that test subjects do not seem to be affected by the length of the slide bar in terms of estimating the circle diameter. However, in order to effectively test the hypothesis the data is checked to identify if it is normally distributed in order to perform a t.test, in which it is calculated if there is a significant difference between the mean of the assessed diameter in the two groups.

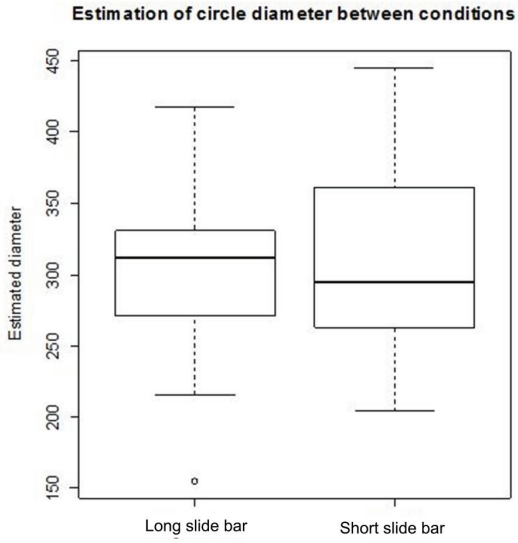


Figure 7.4: *Boxplot based on the estimated diameter lengths between conditions.*

The data was found to be fairly normally distributed (see figure 7.5), and a t.test showed no significant difference between the diameter estimations of the 2 groups (P=0.5331) why null hypothesis N.FM1.1 is accepted, thereby rejecting H.FM1.1.

Since it is a widely spread perception, and women in certain experimental setups have scored significantly lower in spatial tests than men [Linn and Petersen, 1985], the results are checked to see if an effect of gender exists. In the boxplot in figure 7.6 the diameter estimation in regards to gender for each condition is plotted. As can be seen women in the long condition are disagreeing more than the men, however, the men are in general aiming too high. In

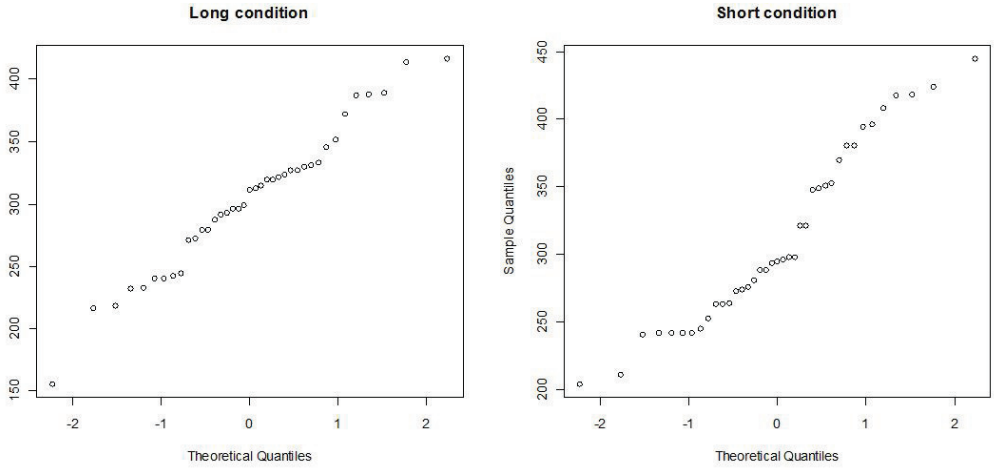


Figure 7.5: *QQ-plots for the distribution of diameter estimates*

the short condition it seems more equal. Performing a t.test to examine if the diameter estimate is affected by gender shows that gender is not a factor important to the outcome of the diameter estimation (p=0.623).

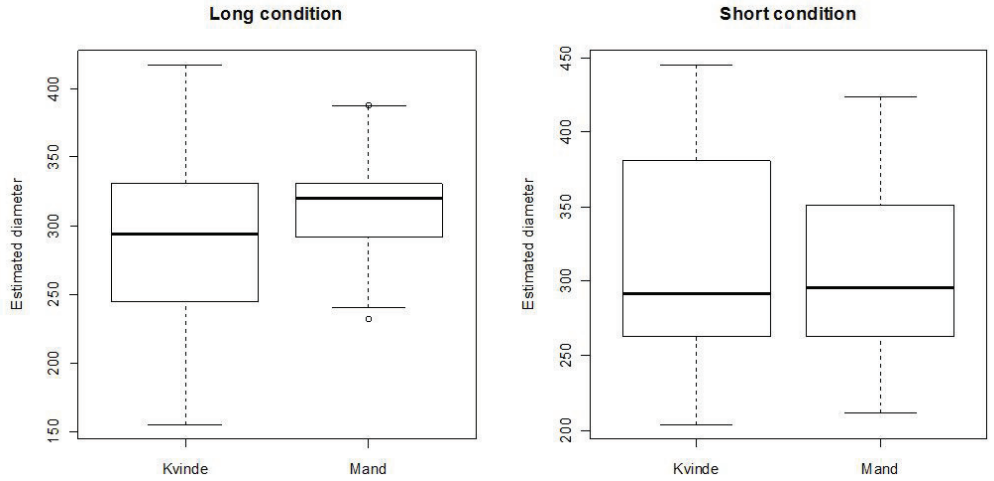


Figure 7.6: *Boxplot illustrating the diameter estimation in terms of gender for both conditions.*

Discussion

The present experiment was made from the curiosity concerning the impact framing seems to have on decision making ([Hsee, 1998], [Tversky and Kahneman, 1986], [Hartmann et al., 2008], etc.). Most of the referenced framing experiments however, are all executed as imaginative experiments, why the present experiment sought to take it a level further towards an applicable setting. The results showed that test subjects were not affected by the surrounding frame setup as a premise for their judgment. The experimental setup is rather simple, why a more complex setup could potentially reveal an effect. Also, some circumstances surrounding the setup in terms of the spatial setup could function as a simplifying circumstance in relation to a website where one may have to judge a given service using a slide bar.

Several test subjects reported having been able to "see" the circle before their eyes when adjusting the diameter on the slide bar, despite the circle being grey. Has this been the case for most test subjects then the assessment aspect of the task can be questioned together with the existence of framing in the experiment. In this case only an adjustment takes place to fit the slide bar length with the diameter of the circle "printed" on their retina. This questions whether visual geometric objects can be used as assessment objects in this kind of experiment.

Another discussion point however, is that the experimental setup does not resemble that of a use situation; a boiler kettle was considered since it is an everyday use object. The current test setup was inspired from the principle of having a frame (the potential volume of the kettle indicated by a maximum) and having an exterior measure (e.g. a coffee cup) to manage within that frame. When translating this setup the frames in this experiment have become the length of the slide bars and the exterior measure is the circle diameter. That is, in translation the setup is the same, however, there is no practical need for test subject to estimate the circle diameter as there would be in estimating a cup of water. Contrarily, it can be argued that in a real use situation people do not over think the situation and act more naturally, than they do in an experimental setting. In fact, if two sets of kettles with manipulated maximums could be distributed to a series of private homes, and the general level of water could be measured, another pattern might appear as a consequence of the everyday setting, where the process of filling a kettle is more habitual. Consequently a potential effect could be observed.

Also, the experimental setup is based on the setup from Hsee (1998), where an ice cream cup acts as a reference point in order to estimate the value of an amount of ice cream. It can be discussed if it is fair to directly translate such a setup, where valuing an amount given a frame is the goal, to a setup where judging a spatial dimension is the goal. The present experimental setup then assumes that valuing an object can be translated to assessing a geometric attribute, which might seem incongruent. In the light of this argument an experiment, where the object to be assessed is of a more evaluating character than a given geometric object, will be made.

Based on the present obtained results it can be concluded that the length of slide bars do not affect the judgment of a geometric exterior attribute, in this case a circle. However, as mentioned the experimental setup of judging a circle diameter can be hard to directly translate into a use situation, why a follow up experiment will be made in order to test if the observed results for this experiment also applies when tested in a more use based setting, where judging an external geometric figure is not the object.

7.2 Experiment FM2

In Experiment FM1 no significant difference was found when subjects were to assess the diameter of a grey circle on two slide bars of different length. As a result of this no framing effect was found of the length of the slide bars. However, the slide bar framing was only tested in the scenario with a geometric object to be assessed; the diameter of the grey circle. Several test subjects in Experiment FM1 stated that they could almost see the circle before their eyes when adjusting the slide bars despite precautions was taken to minimise the contrast between the background and circle colour. The assessed object being of visual and geometric character raised concerns and as a result led to a reconstruction of the experiment in a different scenario.

Experimental objective

The objective of the experiment is as in Experiment FM1, to test if change of "physical" frames will lead test subjects to perceive and use the frames as reference points for judging. In this experiment however, test subjects will not be asked to assess a given geometric feature, instead the setup will encompass test subjects making a decision, which is framed in terms of slide bar lengths.

Hypothesis

The scope of the experiment if varying a frame, and thereby providing another point of reference, will affect judgment of a given attribute. The general hypothesis then states that:

The judgment of a given attribute in a use setting will vary according to the frames provided for the judgment

Apparatus

The test setup is made in Visual Basic 2008 Express Edition, and is run on a MSI 340x laptop with a 13.4" screen. The test is made in a group room at Aalborg University with the curtains closed. All stimulating objects, pictures etc. are removed from the room. Camtasia, a screen capture program, is used to record the screen during the test. The screen capture is found necessary in order to subsequently observe if each test subject has executed the tasks in the correct order or if any irregularities have found place.
The framing effect is tested as part of the experimental setup used in Experiment ST3 (see section 6.1). A slide bar is set up on an initial machine screen, where test subjects are to adjust multiple settings for a plastic container factory machine. One of the parameters is the texture of the plastic used, which is adjusted using a slide bar.

Figure 7.7: *Slide bars used in the experiment in lengths 500 and 800 pixels, respectively.*

As in Experiment FM1 this experiment is designed as a between subject experiment (see figure 7.8 for an experimental overview). This design means that one group of test subjects

will be provided with a slide bar of length 800 pixels, whereas the other group will be exposed to a slide bar of length 500 pixels (see figure 7.7). The labels at each end is the same for both slide bars and an explanation of each label is provided in an attempt to uniform the test subjects perception of the slide bar values. The default location of the marker on both bars is as seen in figure 7.7 identical to assure that test subjects judge based on the same starting point.

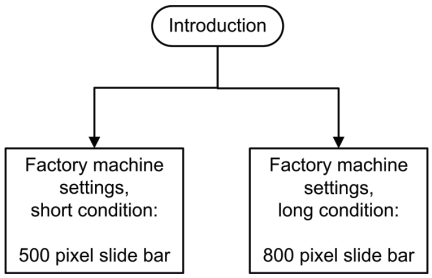


Figure 7.8: *Experimental course overview.*

The lengths of the two slide bars are deliberately kept the same as in Experiment FM1, to avoid this parameter influencing the result. Furthermore, contemplations were made concerning the direction of the slide bars. In FM1 the bars are horizontal, however changing the direction to vertical might change the test subjects´ adjustments on the bars. To keep the result of this experiment clear in the sense that if the hypothesis is accepted this time it will be due to the change in assessment object and not the change in direction of the bars, thus they are directed horisontally in this experiment as well.

Test subjects are to decide the hardness of the plastic of the plastic containers in both slide bar conditions. This object has been chosen since there is no right or wrong decision opposite the circle diameter. However, subjects will possibly have a somewhat common conception of how soft or hard the plastic should be of a food storing container. Furthermore, the object is not visual as the grey circle in Experiment FM1, why no "print" of the object can be left on the test subject's retinas.

The purpose and goal of this experiment is to measure small however, significant differences in the adjustment of the plastic hardness dependent on the slide bar length. The experimental setup is a between subjects experiment, as one group of test subjects is exposed to a short slide bar, whereas the other test group is exposed to a long slide bar. The independent variable is therefore the slide bar lengths (short/long), whereas the dependent variable is the decision value registered for each of the decisions. If framing in terms of the two lengths of the slide bars is affecting the test subjects decision making, this effect can be measured as a difference in the decision value of plastic hardness.
For the developed program see Attachment H and the attached DVD.

Test hypothesis

The test hypothesis is a repetition of that in Experiment FM1 and concerns a comparison between the decision values in the two conditions, and states that:

H.FM2.1 *A significant difference in slide bar adjustment will be observed between the two different slide bar conditions*

N.FM2.1 *No significant difference in slide bar adjustment will be observed between the two different slide bar conditions*

Procedure

Test subjects were placed in the group room in front of the computer. They were asked to remove their mobile phone and watch in order not to disturb the test, as well as they were told how the test would progress (See Attachment B). An introduction text was placed besides the computer (see Attachment C), which the test subjects were asked to read. The experimenter then pressed a computer button which initialised Camtasia (screen recording), without the test subject being aware. Following the test subjects were left alone in the group room, and were asked to come out when having pressed the "Afslut" button.

Afterwards, the test subjects were rewarded with some cake and asked to sign a declaration of consent (see Attachment D) since their mouse clicks had been captured. All test subjects signed the declaration.

Participants

40 students participated in the experiment (mean age = 23,98±3,17 years) with a gender distribution of 17 women and 23 men. Originally 41 participants conducted the experiment, but one was discarded, since the participant did not complete all the tasks. Two pilot tests were executed before conducting the experiment, however these did not raise any complications with the experiment, why no changes were made to the experimental setup. No irregularities were detected in the screen captures, why no further results were discarded.

Results

Boxplots of the assessments made in the two slide bar conditions are plotted to provide an overview of the data (see figure 7.9).

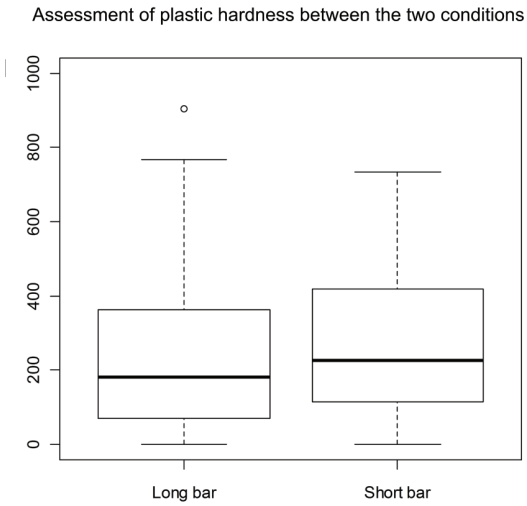


Figure 7.9: *Overview of the decision value distribution between conditions.*

As can be seen in the two boxplots the median of adjustments are nearly the same in the two conditions. However, test subject adjusting the short bar have adjusted the plastic hardness slightly softer than the subjects adjusting the long bar. A t.test will be used to reveal the statistical difference, hence the data sets are checked and identified to be normally distributed (see figure 7.10). From the t.test it is apparent that the adjustments made on the

two bars are not significantly different (p-value = 0.9451). On the contrary it is noticeable how similar the two sets of data are.

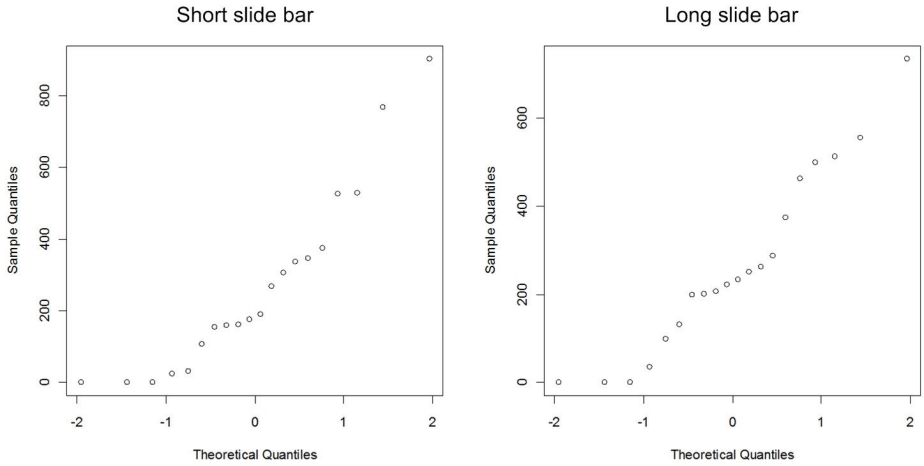


Figure 7.10: *QQ-plots of the decision values.*

On the basis of these data null hypothesis N.FM2.1 is accepted and H.FM2.1 rejected. This complies with the previous findings in Experiment FM1. To examine if gender has influenced the results boxplots of the decision values of the degree of plastic hardness between genders are presented for each bar condition (see figure 7.11). The medians of all four boxplots are similar, however, the standard deviation differs, especially between the decision values of men in the two conditions. This indicates that men disagree much more on the values on the long slide bar than they do on the short. Furthermore it is remarkable that women in the short bar condition have adjusted the plastic softer than the men, but in the long bar condition this is reversed. This result can be coincidental and the boxplot in figure 7.12 of the gender distribution across conditions shows only a slight difference. This is supported by a t.test showing no significant difference between gender (p-value = 0.5379).

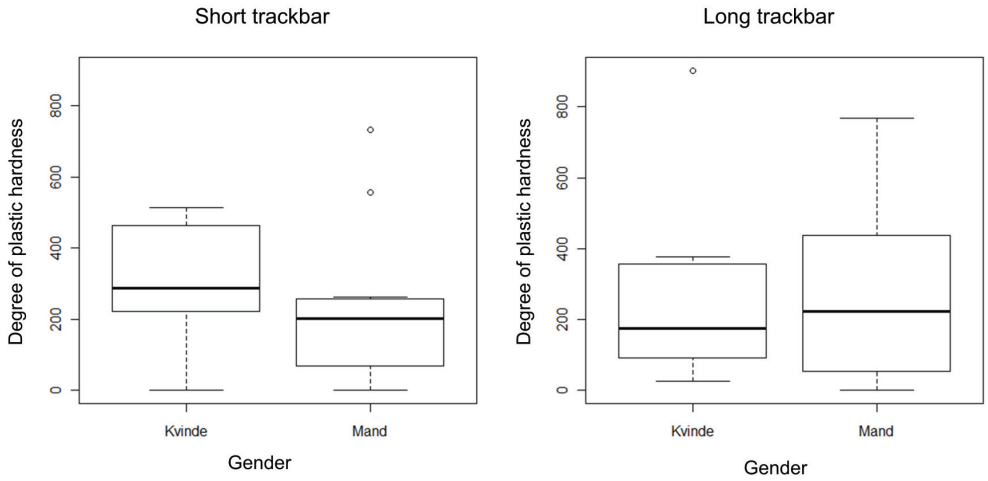


Figure 7.11: *Boxplots of the gender differences in the two different slide bar conditions.*

Discussion

In this second experiment of testing framing with different reference points in the form of different slide bar lengths hypothesis H.FM2.1 is also rejected. This indicates that different

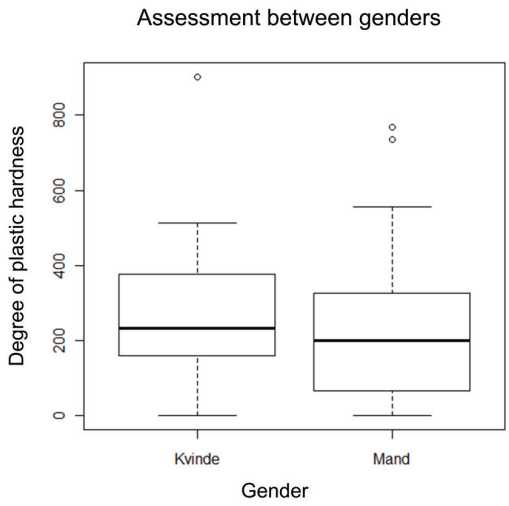


Figure 7.12: *Boxplots of the gender differences across the two different conditions.*

slide bar lengths cannot be used as framing since this hypothesis has been rejected in two experiments with different parameters. One could argue that test subjects are not framed by the length of the slide bars due to the fact that the end labels of the slide bars are the same, hence the length of the slide bar is perceived as quantiles instead of an actual length. If this perception of the slide bars is valid, it seems that a strategy for using slide bars applied by the test subjects is to find the 50-percent quantile and then judge if the marker should be above or below half the scale and how much. If such a method is used, the length of the bar becomes an insignificant factor for the adjustment and as a consequence no framing effect is present.

It is noticeable how much in agreement the test subjects are about the hardness of the plastic of the containers. The p-value illustrating the difference between the means of the answers of the two groups is very close to 1 (p-value = 0.9451). Furthermore the standard deviation of the data sets are 204.74 for the short slide bar and 253.52 for that of the long of a possible 1000, indicating a uniform conception of the degree of hardness of plastic containers among the participants.

It can be questioned if the two assessment and decision value objects in Experiment FM1 and FM2 were too narrow and thereby did not provide enough decision room for the test subjects to actually make individual decisions. The test subjects seem to have a common and beforehand established idea of the objects resulting in the very similar results between the test groups.

The outcome of Experiment FM1 and FM2 strongly indicates that framing cannot be executed using different slide bar lengths. However, this does not exclude that framing can be done with other objects in use situation. Users might be framed by the shape of buttons , the colour of different choices the size of objects, etc.

7.3 Experiment FM3

In the two previous experiments framing has been tested in terms of different lengths of slide bars, but the results revealed no effect of framing. However, this does not exclude framing in use scenarios. Framing is an effective tool to manipulate decisions in wording and it is actively used by insurance companies for instance. In light of this a continuous curiosity concerning framing in user interfaces has led to the set up of this experiment. In most user interfaces there is no need of-, possibility for- or it might seem inappropriate to describe and use words to guide actions. Therefore, it would be interesting to examine if an established visual tool such as colour can act as framing. Is it possible that some colours are associated more than others with positive, neutral or negative circumstances than others? And can they be used to influence users´ decisions (see Appendix III for a brief basic overview of colours)?

Not only colour is suspected to have a potential framing effect, but also sound, size, shape etc. However, for this experiment it is chosen to focus on colour since colours are included no matter if the object of interest is a physical- or a virtual product, whereas sound for example is not always an integrated part of a design, and might be annoying in the use of a website, why it is disregarded. Shape on the other hand also always play some sort of role, however it is harder in some instances to manipulate shape since some objects for practical reasons need to have a specific shape. Contrary to shape, colours can always be modified why they always need attention in a design process, making it an interesting test subject.

In the following the framework for a colour framing experiment is put forward.

Experimental objective

The experiment is conducted in order to examine if colour can act as a framing attribute. More specifically, the experiment seeks to examine if we seem to combine certain colours with certain circumstances; positive, negative and neutral, and if we are affected by the presence of a colour to choose differently. If such a pattern can be observed, it will provide implications that colours can affect our choices.

Hypothesis

An overall hypothesis is put forward to set the scope of the experiment. The overall objective is to test whether colours can be used as a framing attribute, why the following general hypothesis is put forward:

Colours can be used to frame users' choices

The general hypothesis is sought measured through the following experiment and "predicts" that users' choice will be affected by the colours attached to a set of options. The overall hypothesis will be answered through a subset of test hypotheses specified later for the experimental framework.

Apparatus

The experimental setup is made in Visual Basic 2008 Express Edition and is run on two laptops: a MSI 340x with a 13.4" screen and a Fujitsu Siemens Amilo si1520 with a 12.1" screen. The setup consists of 16 questions; five positive, five negative and six neutral,

which can be found in Attachment I. Test subjects are mainly students, why all questions concern subjects they can relate to, such as group work, parties and everyday events. The experiment is made as a between subject setup where half of the test subjects get eight questions (condition A: two negatives, three positives and three neutral) and the other half get the remaining questions (condition B). This was chosen based on the consideration that 16 questions was too many to pose to each test subject, since it might lead them to become unfocused as well as it would take more of the test subjects' time.
Within the two groups a further division is made, where one half gets colours attached to the answers (F conditions) whereas the other group gets answers with no colour. Furthermore in the two sub conditions the questions and answers are varied in two orders (1 and 2) to prevent a negative question from affecting the following or vice versa. These considerations resulted in a total number of eight conditions, which are presented in figure 7.13 and 7.14.

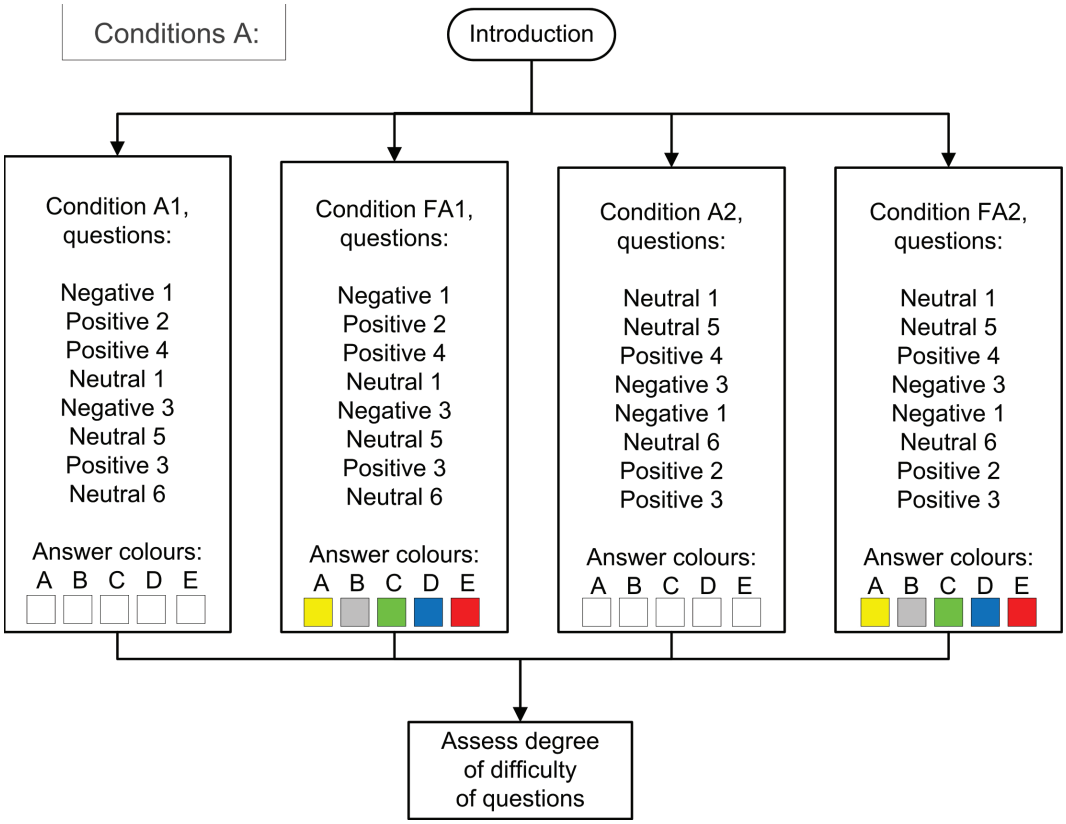


Figure 7.13: *Overview of the experimental setup of the four A conditions consisting of eight questions. The order of the questions is varied between conditions 1 and 2 and within each order condition a variation of colour and no colour is applied.*

Therefore two questions are not following each other between the two groups within the one group (sequence A1 vs. A2 and B1 vs. B2). Also, a positive question is not followed by a negative if that is the case in the other group for example.

For the actual setup one question is presented to test subjects at a time, with five options in the form of buttons (see figure 7.15). Whenever test subjects answer by pressing a button, their choice is gathered in a separate CSV file. When creating buttons in Visual Basic the uppermost button on each page is automatically highlighted when entering the page, consequently functioning as the default choice. This function is switched off to avoid affecting test subjects to choose the uppermost answer (see section 4.3 about defaults).

The questions are posed in a way where there is no obvious answer; in fact the answer options for each question are synonyms for each other, with which it is hoped to observe that

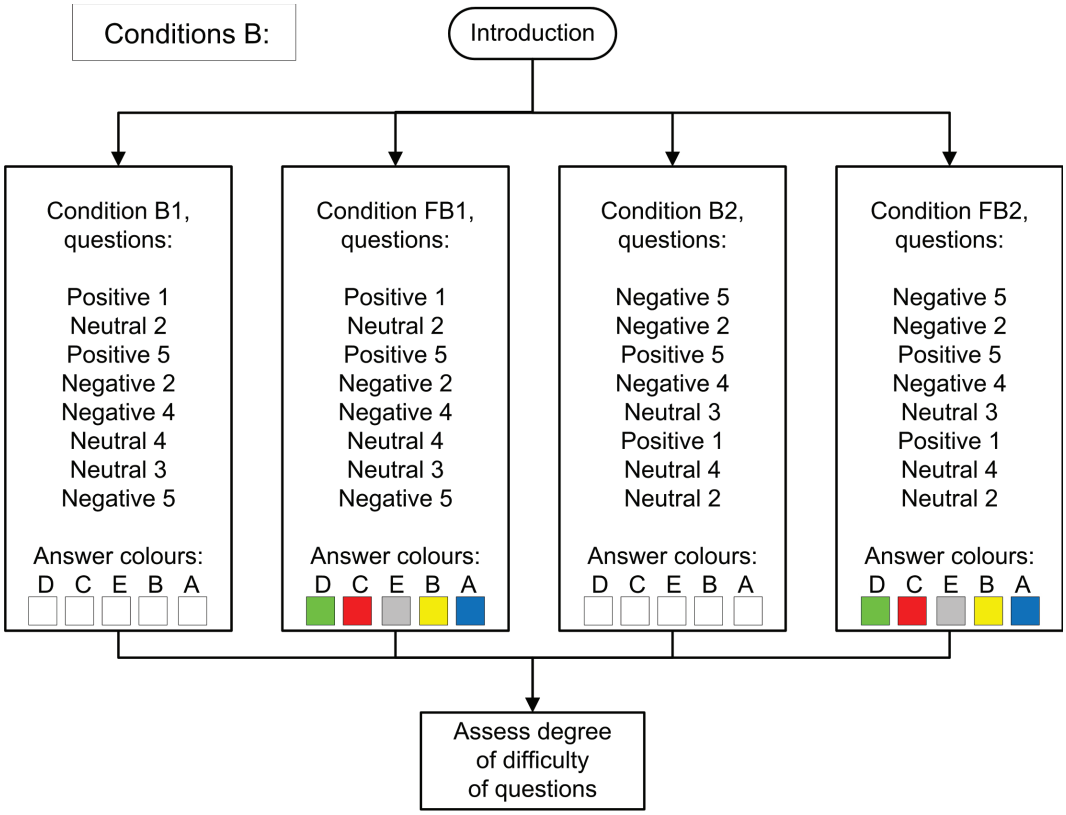


Figure 7.14: *Overview of the experimental setup of the four B conditions consisting of eight other questions. Again the order of the questions is varied between condition 1 and 2 together with the order of the answers within each question and colour.*

colours will control the answers instead. As an example a positive question may address a positive situation such as someone returning ones lost wallet. Upon receiving such a scenario test subjects are asked how they would react in the given situation; "happy", "content", "pleased", "delighted" or "glad". Since all of the choice options are synonyms for the same feeling it should not matter which option the test subject chooses. This also applies for the neutral and negative answers. However, within each of the three categories one of the questions will have an answer option which is more obvious than the others (see Attachment I concerning the questions). If colour seems to be superior to the wording, adding an obvious answer option will reveal if colour is also able to exceed these and thereby what might seem a more "obvious" choice.

For the control condition, no colour is added to the buttons. This is done in order to establish a foundation for comparison with the colour coding conditions. The answer options are based on synonyms, which should make them more or less equal, thus the distribution of answer choices should be equally distributed for all five options, however, this cannot be guaranteed. Therefore, by adding a control condition it is possible to see the distribution of answers without colours, which enables a comparison with the coloured condition, to reveal if all test subjects choose one option because of the colour, or because of the option itself.

Concerning the number of answer options, five options were chosen. From too-much choice theory (see Appendix I) too many choices have been found to have an effect on our decisions, and theoretical implications suggest a limit of six options to avoid an effect of too much choice affecting the test subjects´ answers. However, since the answers are to be very similar a selection of six might still be overwhelming to the test subjects and not that many

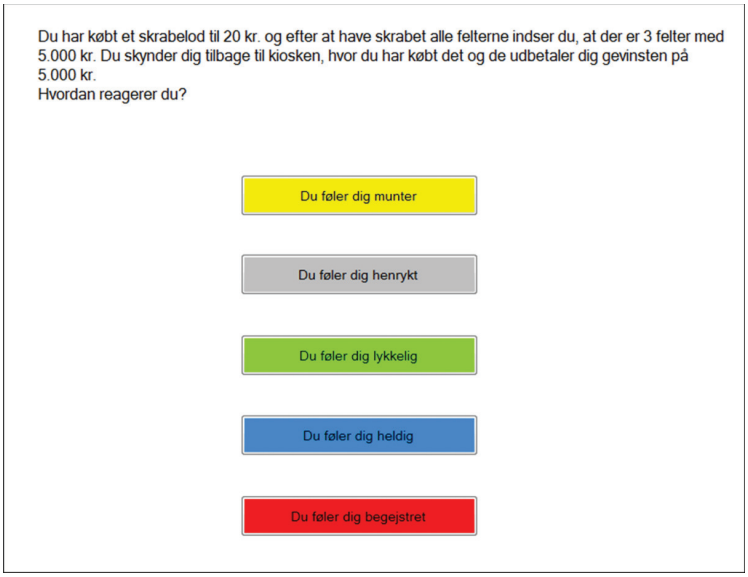


Figure 7.15: *Screenshot from the program from colour condition FA1.*

colours need to be tested. Furthermore, finding that many synonyms is rather difficult. As a result of this it was chosen to have five answers to each question and therefore five colours were chosen as test objectives. The chosen colour scheme is based on a simple colour count on 20 different websites, see Appendix IV concerning the colour count.

The five colours for the experiment was selected to be the five colours used most frequently on the 20 websites, since it can be argued that there is a bigger need to examine colours more frequently used as opposed to colours less frequently used. However, the colour white was most often used (19 out of the 20 cases), but mostly as a background colour. The experimental setup necessitates a background colour, thus white is chosen and consequently disregarded as an answer colour.

The colours blue, grey, yellow and green followed in the count why they are used in the experiment. However, the fifth most used colour would have been orange, but it was assessed to be very close to yellow, why this could affect the results in a way where a clear effect as a consequence could be omitted. It was therefore decided to take the next most frequent colour, which in this case was black. Again, as with the white, this colour is often used as the background colour as well as it is not exactly a colour in category with the other four colours chosen this far. Therefore it was decided to take the next most frequent colour, which was the colour red. Thereby the five colours chosen ended up being: blue, grey, yellow, green and red. The colours chosen in visual basic to represent the five colours are: “DodgerBlue”, “Silver”, “Yellow”, “Lawngreen” and “Red”. These were chosen in order for the test subjects to be able to read the button text, since the colour “Green” was too dark for example. Furthermore, the colours will change according to which screen they are portrayed at, why the colours were chosen on the basis of which looked more representable than others.

Concerning the colours, it was discussed whether to keep one colour with the same answer options at all times or to counterbalance the colours between answer options. The latter is chosen based on the concern of not being able to identify an effect if the colours are always attached to the same answer option. One scenario when keeping the same colour attached to the same answer all the way through the experiment, would be that the test subjects have chosen the answer “happy” in the control condition. In all colour conditions the happy button is red and the subjects have again chosen this answer. This would make it impossible

to observe if the answer happy in the colour condition was chosen because of the colour red or because of the wording. By switching the colours which changes the situations so that red is now in one condition attached to the answer "happy", but in another colour condition to the answer "glad", this provides a clear answer to whether the subjects choose by colour or by wording. Do the test subjects choose "happy" in the colour condition where happy is red, and then "glad" in the colour condition where this is the answer connected to red, this will suggest that the subjects make their decisions based on the colours and especially the colour red in this case.

Test subjects are screened to discard subjects who are colour blind since this experiment is based on the ability of the subjects to detect different colours. The subjects are following presented with the eight questions (depending on which group they are in) on a laptop. They are also asked of age, gender and field of study in the case that these could act as influential factors. After answering the eight questions the test subjects are to assess on a slide bar the degree of difficulty they experienced in answering the questions (see figure 7.16). This assessment is asked for since it is "predicted" that test subjects will find the question with colours attached to the answers easier to answer than the control questions. This is predicted on the basis of the assumption that colours influence test subjects' choice and the choice should on that basis be easier to make.

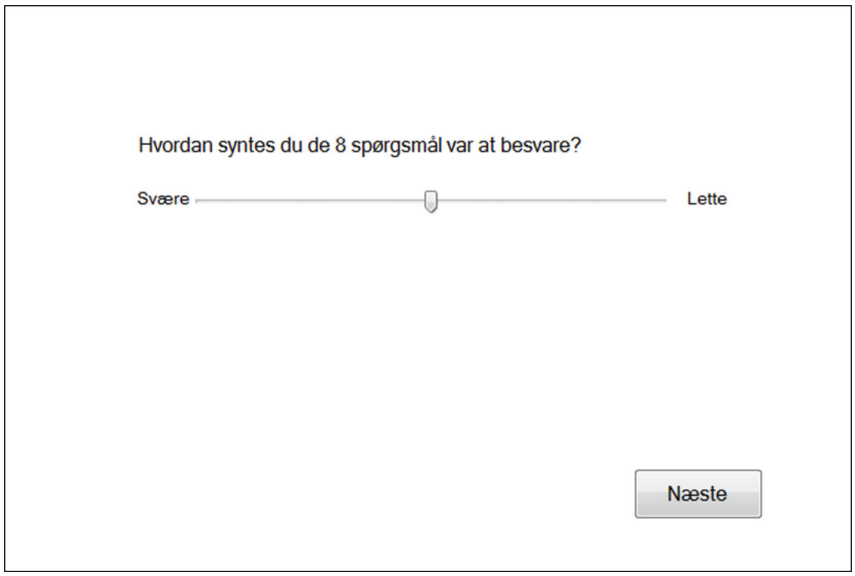


Figure 7.16: *Screenshot from the program of the last question where test subjects are to assess the degree of difficulty of the questions.*

If colours have a framing effect on test subjects will be measured via the dependent variable: the answers chosen in each condition, dependent on the independent variable: the colour of the answers.

For the developed program see Attachment J to view control condition A1 and Attachment K to view colour condition FB1. One edition of the eight programs can be found on the attached DVD.

Procedure

The experiment was executed on two laptops; a MSI 340x with a 13.4" and a Fujitsu Siemens Amilo si1520 with a 12.1" screen. Two subjects at a time participated, one at each com-

puter. They were beforehand explained that the questionnaire consisted of 8 questions that might appear strange in relation to each other. They were also told that there might not be an answer that complied with the test subject's opinion completely in which case they were to choose the most compelling answer.

After answering the eight questions the test subjects were asked to assess the degree of difficulty they experienced answering the eight questions.

Test hypotheses

In this section the test hypotheses for the experiment are put forward. Each of the hypotheses is accompanied by a null hypothesis, which is going to be tested. The objective of the experiment is to establish if colours have a framing effect and thereby can influence user's choices, why the test hypothesis states:

H.FM3.1 *Test subjects in the colour conditions assess the degree of difficulty of the questions significantly easier than test subjects in the control condition*

N.FM3.1 *Test subjects in the colour conditions do not assess the degree of difficulty of the questions significantly easier than test subjects in the control condition*

H.FM3.2 *Answer options chosen in the colour conditions are significantly different from answer options chosen in the control condition*

N.FM3.2 *Answer options chosen in the colour conditions are not significantly different from answer options chosen in the control condition*

The test hypotheses and appertaining null hypotheses will test if an effect of colour has influenced the test subjects' answers, since this should result in a significant difference between the answers in control conditions and colour conditions.

Participants

160 students participated (mean age = 23.01±3.62 years) with a gender distribution of 65 women and 95 men. None of the participants reported any colour blindness.

Results

In figure 7.17 two boxplots of the parametric data of test subjects' assessment of the degree of difficulty of answering the eight questions, is presented. The boxplot shows that test subjects in the colour condition assessed the questions slightly easier to answer (5.31% easier on average) than the test subjects in the control condition, since medians in the two boxplots are different.

To test hypothesis N.FM3.1, which concerns the difference in the degree of difficulty of the questions, a t.test is made. The data is identified to be normally distributed (see figure 7.18) and the following t.test reveals that null hypothesis N.FM3.1 is accepted (p-value = 0.5998), whereby hypothesis H.FM3.1 is rejected.

In figure 7.19 the non-parametric data from question Neutral 1, Neutral 5, Positive 1 and Negative 3 is presented. In each graph it is possible to see how many test subjects have

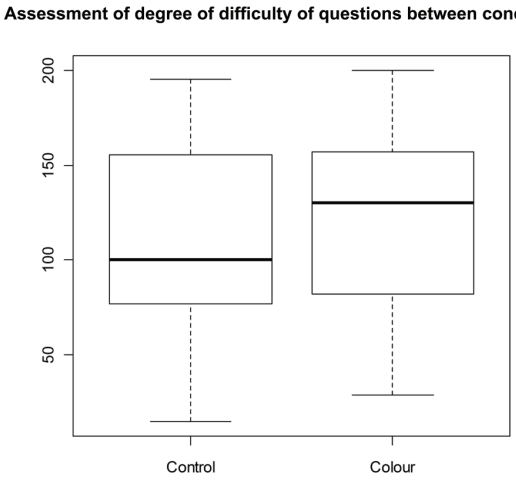


Figure 7.17: *Boxplot showing the distribution of evaluation of question difficulty level.*

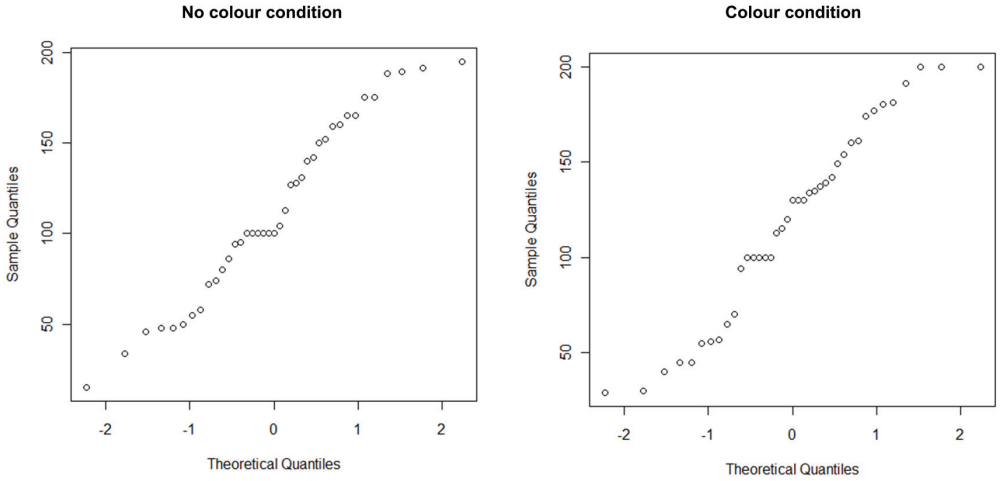


Figure 7.18: *QQ-plot of evaluation of question difficulty level.*

chosen each answer depending on which of the four conditions test subjects were exposed to. The x-axis of the four graphs represents the five answers by A, B, C, D ,E in each question. For each answer there are four pillars, each representing one of the four conditions (A1, FA1, A2, FA2), and the colour of each pillar represents the colour the answer had in the specific condition. The graphs are meant to provide an overview of the raw data, since the number of answers in each condition can be seen together with the specific colour the answer had in each condition.

In the graph of Neutral 1 it can be seen that answer A is chosen much more in condition FA2 where it is attached to the blue colour, than when attached to the other colours. In an attempt to control the majority of nuisance variables in this experiment and to obtain results on different levels, the experimental setup has become somewhat complex. As can be seen in the four graphs in figure 7.19 no clear tendencies can be registered and due to this, the data from the 12 other questions will not be represented graphically.

The data obtained is non-parametric, composite count data. As a consequence of the effort put into counterbalancing the different aspects within the experiment such as question

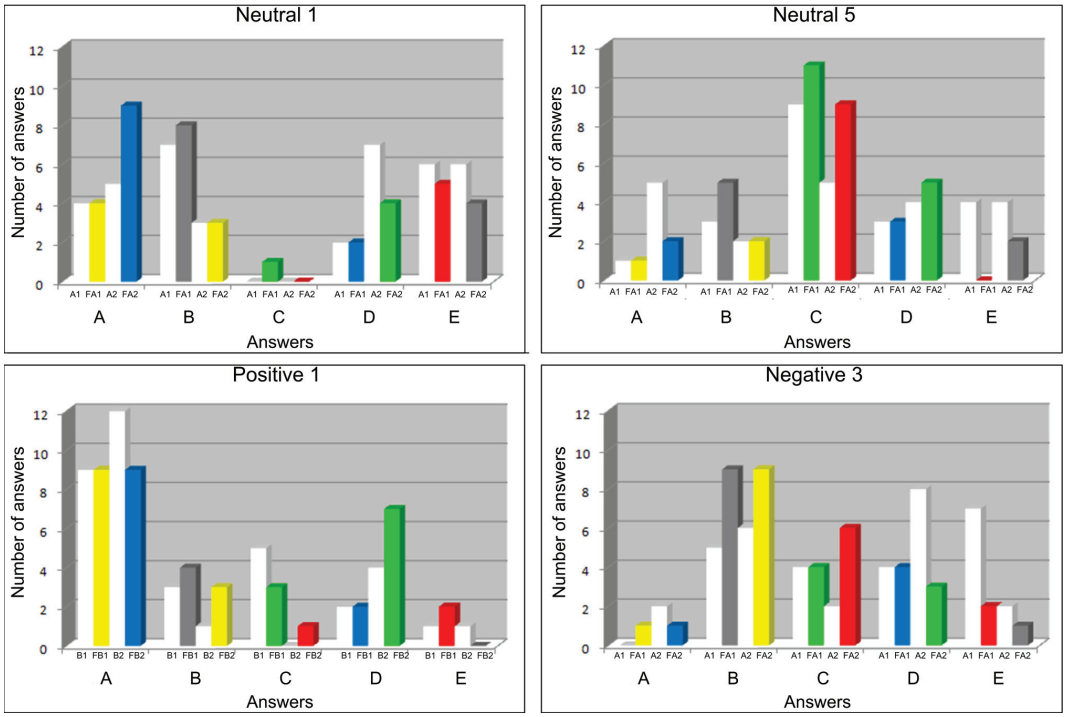


Figure 7.19: *Four graphs of the answer distribution within each question: Neutral 1, Neutral 5, Positive 1 and Negative 3.*

order, colour order, etc. the data has become somewhat complex to handle statistically. Furthermore, the number of test subjects is limited in relation to the number of factors and as a consequence, not all colours have been attached to all answer options, which creates a disparity in the data between groups. Therefore; data from the coloured and non-coloured answers are counted as if one test subject had answered all 16 questions. In this way the data is simplified to contain only two conditions (colour versus non-colour) instead of four, where the counterbalancing was different. By handling the data in this way a more homogeneous data frame is obtained, based on which statistical analyses can be performed.
A chi-square test is employed in order to observe if the difference between the no colour condition and the colour condition is significant i.e. if there is an effect of colour. However, the p-value obtained from the chi-square test showed that hypothesis H.FM3.2 is rejected and N.FM3.2 is accepted, as no significant difference was identified (p=0.2283, X^2=5.6333).

Positive condition		
Colours	No colour	Colour
Blue	40	25
Green	43	52
Red	33	35
Yellow	38	50
Grey	46	38

Negative condition		
Colours	No colour	Colour
Blue	32	20
Green	57	48
Red	56	62
Yellow	26	30
Grey	29	40

Neutral condition		
Colours	No colour	Colour
Blue	66	61
Green	65	75
Red	36	38
Yellow	25	21
Grey	48	45

Figure 7.20: *Overview of counts for answers chosen based on the direction of questions. Numbers in the "no colour" column indicate how many test subjects have chosen the same answer in the control condition as in the colour condition, when the answer chosen in the colour condition was blue, green, red, yellow or grey.*

Despite having rejected hypothesis H.FM3.2 the data is examined further to reveal if there should be any sub effects not identified by the overall test of the data set. It is therefore examined if colour has had an effect on answers in the different direction questions

(pos/neg/neu). Therefore a chi-square test is applied on a table containing the frequencies of the combinations of answer and colour condition for each condition (see figure 7.20). However, no effect of colour was found in terms of direction of question, see figure 7.21.

Direction	X^2	P-value
Positive	6.7713	0.1485
Negative	5.8851	0.2079
Neutral	1.4098	0.8425

Figure 7.21: *Results of the chi-square tests according to direction.*

Nonetheless, it is interesting to observe that the p-value related to the neutral questions is very close to 1 compared to the others, which indicates that colour has very little influence, if any in the neutral questions, but approximates an effect in relation to negative and positive questions.

	All conditions		
Colours	No colour	Colour	P-value
Blue	138	106	0.0037
Green	165	175	0.4432
Red	125	135	0.3338
Yellow	89	101	0.2182
Grey	123	123	0.9287

Figure 7.22: *Overall overview of counts for answers chosen in the "no colour" and the colour condition. The numbers in the no colour column indicate how many test subjects have chosen the same answer in the control condition as in the colour condition, when the answer chosen in the colour condition was blue, green, red, yellow or grey. P-values are obtained based on a chi-square test between conditions.*

The overall counts for answers in the no-colour versus colour condition seen in figure 7.22 show that blue has a significant effect as adding blue to the options otherwise chosen in the no-colour condition, makes test subjects deselect that option. This indicates that some kind of effect occurs when blue is added. The differences in numbers of test subjects choosing differently as a consequence of colour might nonetheless be affected by multiple external influences, which will be discussed later.

To assure that the result has not been affected by gender a chi-square test is run on a table containing the frequencies of the combinations of answer and gender, which did not show any effects of gender, with a p-value equalling 0.2289.

Discussion

Both hypotheses put forward were rejected, which implies that colour cannot function in terms of framing peoples' choices. The colour blue did however affect test subjects to deselect answers which they chose when no colour was added, but several nuisance variables may be accountable for this result. The experiment contains a large number of factors and as the number of factors grows the number of test subjects should grow proportionally. 160 test subjects participated in the experiment; however this is barely sufficient since dividing the number of test subjects out on all the conditions, only 20 test subjects were tested in each condition. In terms of consolidating an effect which is not clearly demonstrated, a larger number of test subjects is needed.

As the experimental setup was rather complex, it could be simplified by reducing the number of colours to two or three and reducing the number of answers in each question. Also the amount of questions could be reduced meaning that the individual subject could answer all questions without becoming bored or too conscious of her/his own answers.
Furthermore, only a control condition with no colour and a colour condition where the questions and answers are in the same order could be tested. The experiment could also have been simplified if only negative, neutral or positive questions were to be tested instead of all three conditions.

In terms of nuisance variables in this experiment; the experimental scenario can appear rather odd to test subjects. The questions are all out of context and they do not make much sense, which was a consequence of assuring that the questions were neutral, negative or positive and that the answers within each question were as equal as possible.
Furthermore, test subjects verbally commented that they noticed following a specific decision making pattern such as having chosen three red answers in a row, and therefore in the next question chose a yellow answer, just to break the pattern. This was mentioned by several test subjects and therefore acts as a nuisance variable in the experiment. Had each test subject only been given e.g. three questions this nuisance might have been avoided and the result of the experiment been another. Furthermore, having presented the questions in a meaningful context could make the test subjects less aware of how to answer the questions, and more immediate answers might have been given.

The experimental setup included attaching different colours to each of the five answers in the 16 questions. This as well might serve as a nuisance variable since each colour is intended to function as a frame. By framing test subjects with all five frames (colours) simultaneously the framing effect might be camouflaged. Furthermore, it is possible that the colours affect the answers placed next to them, resulting in several colour frames affecting each answer and diminishing the framing effect.
A further questionable aspect of the fact that the colours are attached to five answer buttons in each question is that the buttons consist of a rather small limited area. Consequently each colour is only represented on a small area and might thereby not be able to fully influence test subjects decision making. Had each colour been represented on a larger area an effect of the colours might have been detectable.

The experiment included too many parameters compared to the number of test subjects, why it was not possible to deduct any clear tendencies, despite the suggestion of a blue colour effect, where test subjects displayed behaviour towards not choosing the answers they would have in the no colour condition, when they were blue. As mentioned, if this effect is genuinely applicable is questionable. Furthermore it was found that neutral questions were not at all influenced by the colour attached, whereas the negative and positive questions were more affected by colour, though not significantly, why conclusions on this part cannot be made clearly.
The fact that test subjects were aware of their decision pattern according to the mixed colours attached to the decision buttons, can have caused the unclear results. As some tendencies were shown in this experiment, a new version of Experiment FM3 is put forward, where test subjects are not exposed to all colours, ruling out the colour pattern nuisance.

7.4 Experiment FM4

In Experiment FM3 it was sought tested if colours have framing potential, however no significant effects were found, why no hypotheses were accepted. Several nuisances in the experiment could have diminished the effect of colour framing; the colours having affected each other, test subjects not being framed enough by each colour since the buttons were small, each colour was only represented on one button, etc. Due to these considerations a new experiment seeking to test colour's potential ability to frame user choices is set up.

Experimental objective

This experiment is conducted in order to retrieve clearer results concerning colours' ability to influence the choices of people as a framing attribute. It is believed that colours to some extent can influence user choices, which is sought demonstrated in this experiment.

Hypothesis

The overall hypothesis is the same as in Experiment FM3 and concerns whether colours can be used as a framing attribute, and therefore states:

Colours can be used to frame users' choices

As in Experiment FM3 the hypothesis is sought measured through the following experiment and "predicts" that choices of test subjects will be affected by the colour attached to a set of options. An actual test hypothesis will be put forward later in this section.

Apparatus

Nine questions (three neutral, three positive and three negative) are chosen from the 16 questions in Experiment FM3. Six of the questions are chosen as they are the ones with least significant difference in answer distribution, thus the most neutral questions from the control condition. The last three questions (Positive 4, Neutral 4 and Negative 5) are chosen as they are the ones with most significant difference in answer distribution and they are chosen in order to be able to recognise if the effect of colour framing is strong enough to over rule "obvious" answers. In figure 7.23 the nine questions are presented together with their appertaining p-value of each question's answer distribution in the control condition in Experiment FM3 (for the wording of the questions see Attachment I).

Questions	P-value
Positive 1	0.1124
Positive 2	0.6452
Positive 4	0.0014*
Negative 1	0.2403
Negative 4	0.2352
Negative 5	0.0010*
Neutral 4	0.0012*
Neutral 5	0.3432
Neutral 6	0.4067

Figure 7.23: *Table of the nine questions chosen from Experiment FM3 with the appertaining p-values of each question's answer distribution in the control condition with no colour.*

To prevent the colours from influencing each other as they possibly did in Experiment FM3, each test subject will only be exposed to one of the five colours: grey, yellow, red, green or blue (the same colours as used in Experiment FM3). Each of the five answer options to each of the nine questions will as a consequence have the same colour (see figure 7.24).

The experiment is a between subject design as each test subject will only be exposed to questions with either blue, green, yellow, grey or red answer options (see figure 7.25). The dependent variable is the answers chosen in each colour condition, and the independent variable is the variation of colour between conditions.

Contrary to Experiment FM3, which was run on two laptops it is chosen to execute this experiment using paper questionnaires. This is decided since distribution of paper questionnaires is easier and quicker, since test subjects can fill out the questionnaires simultaneously contrary to two at a time on the laptops. Furthermore, it is assessed that printing the questionnaires on paper will not cause any nuisance for answering the questions. The reason why Experiment FM3 was run on laptops, was due to it being part of a larger experimental setup, where Experiment FM1 was executed as well, which necessitated it to be executed on laptops.
The disadvantage of printing the questionnaires on paper is that test subjects are not bound by programmed rules, why they can e.g. choose more than one answer in each question and avoid answering questions, however, the advantage of saving time distributing the questionnaires was valued higher.
For the questionnaires see Attachment L to view condition blue.

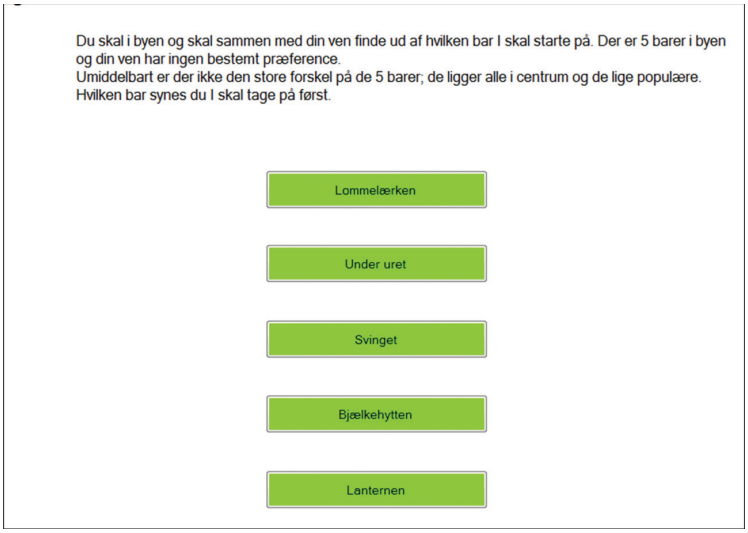


Figure 7.24: *Question Neutral 6 used in Experiment FM4.*

Procedure

The participants were all approached in their group rooms and asked if they wanted to participate in the experiment for the reward of a cookie. Before participating test subjects were informed that the questionnaire consisted of nine questions of somewhat odd subjects. Furthermore they were explained that the answer options to each question might appear very similar, but that they were to answer each question with only one answer and that they should not discuss answers with each other before having filled out the questionnaires (see the introduction in Attachment B). Test subjects were beforehand screened for any kind of colour blindness.

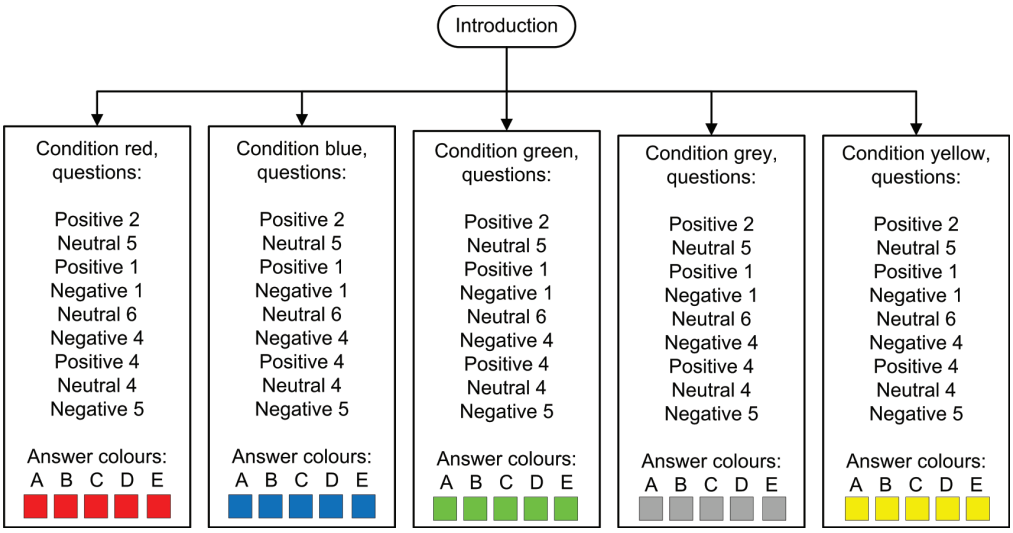


Figure 7.25: *Visual overview of the course of the experiment.*

Test Hypothesis

To test the overall hypothesis a test hypothesis is put forward with an appertaining null hypothesis. The answers in each colour condition are to be tested against each other, why the test hypothesis and null hypothesis states:

H.FM4.1 *Answer options chosen between the five different colour conditions are significantly different within each question*

N.FM4.1 *Answer options chosen between the five different colour conditions are not significantly different within each question*

If no influence of colour is present, no significant difference should be found between the number of test subjects having chosen e.g. answer "Harm" when it is red than when it is yellow, blue, grey or green.

Participants

100 test subjects participated in the experiment, all students from Aalborg University. The mean age of the test subjects is 21±1.41. The gender distribution is 36 women and 64 men. None of the participants reported on being colour blind.

Results

To provide a visual overview of the results of this experiment, graphs of three of the nine questions are presented in figure 7.26.

In question Neutral 4 most test subjects have chosen "Hammer" and "Skruetrækker", however, it is noticeable that "Skruetrækker" is chosen much more when it is yellow or red and "Hammer" is chosen when it is blue, green or grey. Furthermore the answer "Rund" in question Neutral 5 seems to be a more attractive choice to test subjects when it is red than when it is one of the other four colours.

In question Neutral 6 a tendency can be seen of the colour red and the answer "Bjælkehytten". The tendencies will be examined further with generalised linear models (glm) accounting for a poisson distribution.

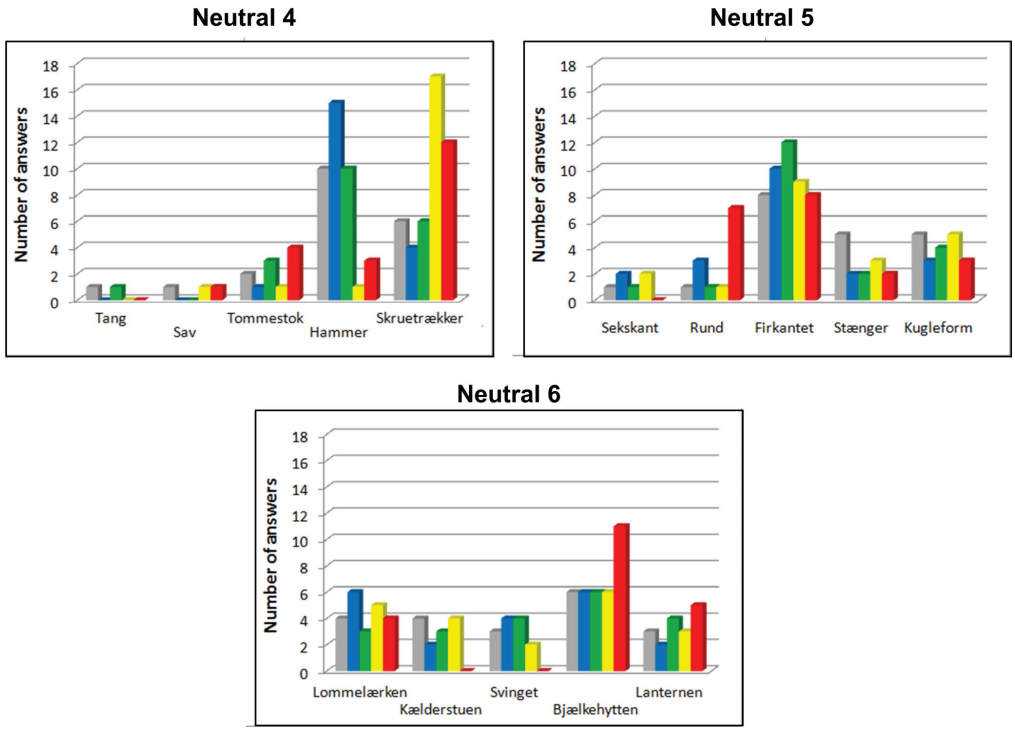


Figure 7.26: *Three graphs presenting the results of the three neutral questions.*

A glm is made for each of the nine questions; comparing the answers within each question across the five colour conditions. Thereby, calculations are made to examine if the answers given in the blue condition are significantly different than those in the green, red, grey and yellow conditions for each of the nine questions. Of the nine questions and appertaining glm calculations, only three showed an effect of colour: Negative 4, Positive 4 and Neutral 4. It is noticeable that both question Neutral 4 and Positive 4 are questions with one "obvious" answer of the five (see figure 7.23).
In figure 7.27 graphs of the three questions are presented. The glm for Negative 4 indicates that an effect of colour has influenced test subjects to select and deselect answer A: "Du bliver ked af det" (see figure 7.28). From the graph it can be seen that test subjects choose answer "Du bliver ked af det" when it is grey and blue, but deselect it when yellow, green and especially red. The number of test subjects who have chosen this answer in total is not substantial why it is difficult to conclude upon, however, when the answer is attached to the colours grey and blue twice as many test subjects have chosen it.

The second question which displayed an effect of colour is Positive 4, in which answer D: "Du føler dig heldig" is very close to being chosen significantly different between colour conditions (p-value=0.0727), see figure 7.28. In the graph of Positive 4 in figure 7.27 it can be seen that the number of times answer option D is chosen varies substantially according to which colour it is attached to. When attached to yellow 75% of test subjects choose answers D compared to 30% when attached to green and grey, which shows implications of colours framing test subjects to select and deselect the answer.
The last question displaying a significant effect of colour is Neutral 4, where mainly the answers D and E have been chosen (see figure 7.28). However, if D "Hammer" or E "Skruetrækker" is chosen seems to depend on the colours attached to the answers: when grey and green 50% of the test subjects choose "Hammer" and when blue "Hammer" is chosen

by 75%. "Skruetrækker" is on the other hand chosen by 60% of test subjects when attached to red and 85% when yellow. In the control condition with no colour in Experiment FM3 "Hammer" has been chosen by significantly more test subjects as it is the obvious answer in this question. By adding yellow and red to the answers test subjects are persuaded into choosing "Skruetrækker" instead of the obvious answer, which indicates that test subjects' decisions in this situation have been framed by colour.

In these three questions a suggestion of colour framing on test subjects choices is seen, however this is only the case in three out of nine questions why hypothesis H.FM4.1 cannot be accepted.

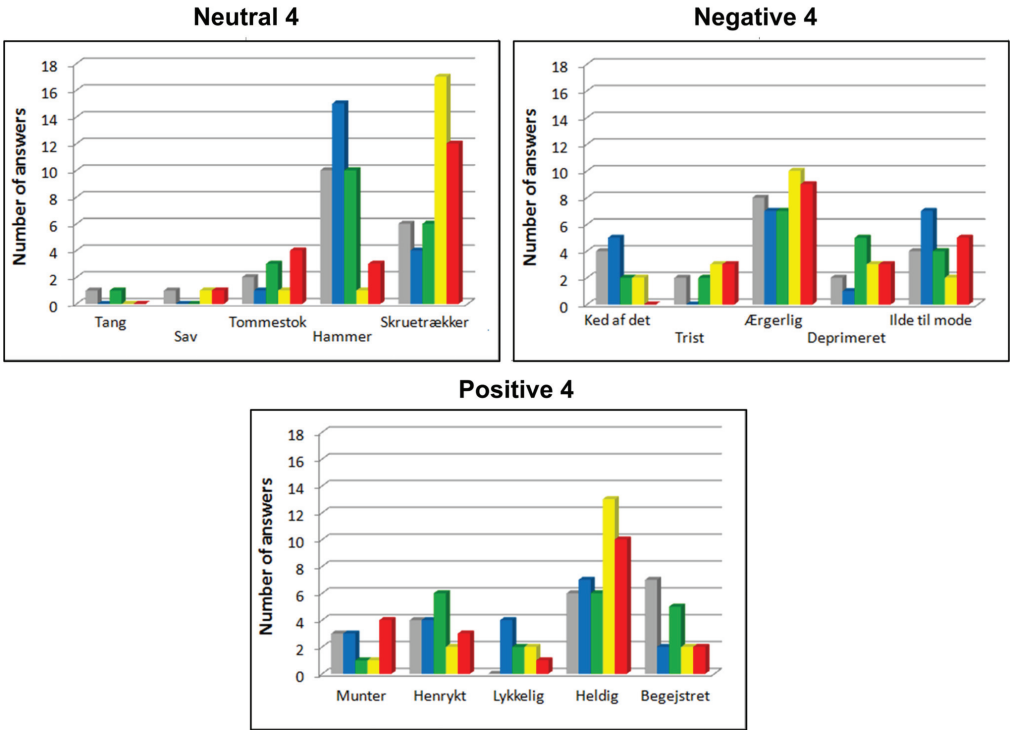


Figure 7.27: *Graphs of the three questions that exhibit significant effects of colour.*

To assure the result has not been affected by gender a glm is calculated of gender described by the colours chosen. No effect of gender is found since the lowest p-value obtained equals 0.472.

Negative 4	
Answers	P-values
(Intercept)	7.14e-16
Ked af det	0.0272
Trist	0.7782
Ærgerlig	0.7662
Ilde til mode	0.4784

Positive 4	
Answers	P-values
(Intercept)	3.05e-09
Heldig	0.0727
Henrykt	0.5174
Lykkelig	0.4022
Munter	0.3621

Neutral 4	
Answers	P-values
(Intercept)	7.04e-15
Sav	0.2559
Skruetrækker	0.0406
Skruetrækker	0.6.24e-05
Tang	0.7963
Tommestok	0.0473

Figure 7.28: *Results from three glm's of the effect of colour in question Negative 4, Positive 4 and Neutral 4.*

Discussion

The hypothesis put forward for this experiment is rejected. However, tendencies can be seen in the results and in three out of the nine questions the effect is significant (except for Positive 4 which is nearly significant, p-value=0.0727). The rejection of H.FM4.1 can be caused by several factors. The most obvious one is that colour has no framing effect and therefore no influence on decision making, and the tendencies detected might be caused by too few participants. Since the experiment is based on test subjects' subjective decisions the experiment might have had a different outcome if 100-150 test subjects had participated in each condition. However, this would have resulted in a total of 500-750 test subjects in the experiment instead of the 100 test subjects that participated.

The results might also portray that colours are able to frame test subjects' choices, but that it is not consistent within the frames provided. The significant effect found in question Neutral 4 could for instance be explained from the fact that the answers consist of objects in the form of tools. It is possible that the colour used for the handle of a screwdriver as a rule is yellow or red, whereas a hammer is the standard of tools, why this is the obvious answer chosen by significantly more test subjects when no color is attached to the answers in Experiment FM3.

Another explanation of the results could be that colours have a framing effect, however the influence of the effect might be too little to be detected in this experiment or the effect might have been camouflaged by other parameters that have not been held constant during the experiment. Also, it might be that a systematic pattern of choice based on colour cannot be found as it depends on context. Since all the answers are added the same colour for each test subject the choice is no longer between colours, as in Experiment FM3. The effect that could have been measured in this experiment would have been an effect of one colour having influenced the test subjects' decision making. This can be seen in the results of question neutral 4 subjects choose answer "Skruetrækker" when the answers are either red or yellow.

Whether it is tested in this experiment if colours have the ability to frame users choices in a certain direction, can be discussed. The answers of the questions are synonymous with each other why they do not have different direction such as negative and positive. Had the answer options e.g. consisted of a positive, a negative and a neutral answer evidence of colour framing might have been detectable. However, the main issue with the experimental setup is that it resembles more of a questionnaire than a use situation. When the results are not clear in such a setup, the applicability becomes limited, why at least one other experimental set up is necessary in order to be able to conclude if colours do seem to have an effect or not. Since the purpose is to examine if colours can act as frames in use situations it is decided to test further in another experiment where the experimental frames will be that of a use setting. In order to be able to deduce clear results the experiment will only be based on a few parameters, which will be presented in the following section.

7.5 Experiment FM5

In Experiment FM3 five colours were tested against each other to see if test subjects choice and decision making could be influenced by colour, resulting in test subjects finding an answer more attractive due to the colour of that answer. The result turned out to be very unclear and a more simple experiment was therefore set up, Experiment FM4. In this experiment the same five colours were tested, however this time each colour was tested separately to avoid any influence of colours on each other, but this meant that all five answers in each question were influenced by the same colour, why the effect might have been camouflaged. Again the results showed a few tendencies, but the hypothesis was not accepted.

In this experiment colours will again be tested against each other as in Experiment FM3 and FM4, however with fewer variables to heighten the chance of a clear result.

Experimental objective

The previous framing experiments (Experiment FM3 and FM4) have both tested if colour can affect the test subjects' choices and decision making. Some tendencies of colour framing effect have been indicated, however, no actual descriptive tendency has been detected. In this experiment it is attempted to test if certain colours can affect a user choice in a positive or negative direction.

Hypothesis

The overall hypothesis concerns whether colours can be used as a framing attribute, why the general hypothesis states that:

Colours can be used to frame users' choices

Apparatus

The test setup is made in Visual Basic 2008 Express Edition, and is run on a MSI 340x laptop with a 13.4" screen. The test is made in a group room at Aalborg University with the curtains closed. All stimulating objects, pictures etc. are removed from the room. Camtasia, a screen capture program, is used to record the screen during the test.

This experiment is executed as part of larger experiment in order to make the questions relevant in a context. The scenario is that of a plastic container factory machine which needs to be configured. The subjects are to adjust a number of parameters such as the colour of container lids, size of containers and the thickness of container plastic on an interface resembling an industrial machine setup. Following these adjustments two (yes/no) questions concerning system setup are presented. It is these two questions which are tested in this experiment.

Test subjects are divided into two groups, where one of the groups is presented with the questions where the yes/no buttons are red, whereas the other group is presented to green yes/no buttons. In figure 7.30 and 7.31 screenshots of the first and the second question are shown. One test subject will only be exposed to one colour in both questions, it is for illustration purpose that both colours are shown in the figures (see figure 7.29 for an experiment overview).

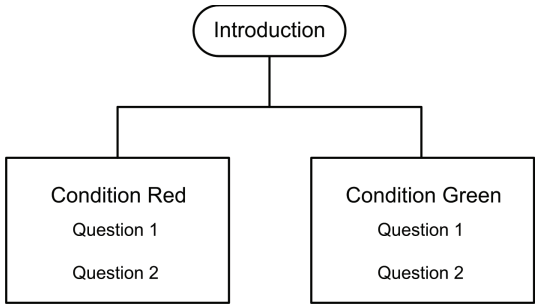


Figure 7.29: *Experimental course overview.*

The questions are considered somewhat neutral in the sense that test subjects are as likely to answer yes as they are to answer no to both questions. Only two different colours are chosen for this experiment in order to keep the experiment simple and the number of test subjects at a minimum. The two colours chosen for the buttons are green and red, since these are often used colours in user situations, where green and red are used symbolically to illustrate turn-on/accept and danger/turn-off/warning, respectively.

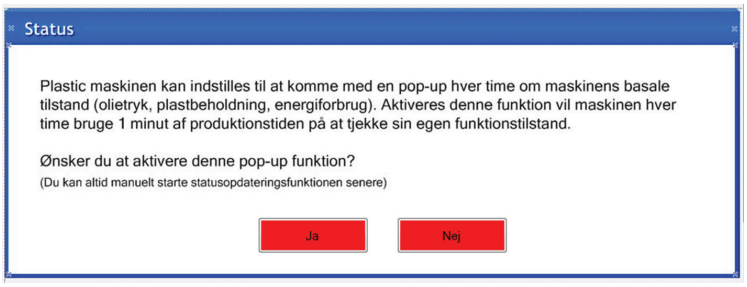


Figure 7.30: *Screenshot of the Pop-up question in the machine setup.*

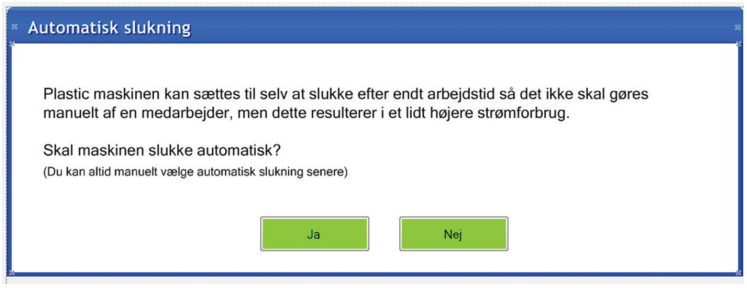


Figure 7.31: *Screenshot of the Turn-off question in the machine setup.*

Furthermore Experiment FM5 consists of only two questions in order to avoid the test subjects becoming too conscious about their decision making. It was noticed in Experiment FM 3 that test subjects commented that they noticed displaying a certain decision making pattern, such as having chosen a blue answer three times in a row. Due to the test subjects becoming aware of this pattern they deliberately broke the pattern and therefore chose e.g. a green answer as the next one even though they wanted to answer blue again.

The experimental setup is a between subjects experiment, as one group of test subjects is exposed to red buttons, whereas the other test group is exposed to green buttons. The independent variable is therefore the button colour (red and green), whereas the dependent variable is the answer registered for each of the decisions (yes/no).

For the developed program, see Attachment N and the attached DVD.

Procedure

Test subjects were placed in the group room in front of the computer. They were asked to remove their mobile phone and watch in order not to disturb the test, as well as they were told how the test would progress (see Attachment B). An introduction text was placed besides the computer (see Attachment C), which the test subjects were asked to read. The experimenter then pressed a computer button which initialised Camtasia (screen recording), without the test subject being aware.
The test subject was then left alone in the group room, and were asked to come out when having pressed the "Afslut" button.

Afterwards, test subjects were rewarded with cake and asked to sign a declaration of consent (see Attachment D) since their mouse clicks had been captured, which all of them did.

Test Hypothesis

In order to test the overall hypothesis a test hypothesis is put forward. The objective of this experiment is to test if the influence of two different colours can affect the subjects answer in a positive or negative direction, why the test hypothesis states that:

H.FM5.1 *A significant difference will be found between answers in the red condition and answers in the green condition*

N.FM5.1 *No significant difference will be found between answers in the red condition and answers in the green condition*

Only the colours are changed in the two conditions, why it can be accepted that colours influence test subjects' decisions if test hypothesis H.FM5.1 is accepted.

Participants

40 test subjects participated in the experiment; all students at Aalborg University. The mean age of the test subjects is 23.98±3.17. The gender distribution is 17 women and 23 men. 41 participated in the experiment, however, one subject did not complete the tasks completely why his results were deleted. All the subjects were screened for colour blindness, however, none of the participants reported being colour blind and all participants signed the declaration of consent.

Results

To provide an overview of the data two graphs are presented illustrating the answer distribution in the two questions between the two colour conditions (see figure 7.32).

As can be seen in the two figures the differences between the green condition and the red condition are hardly visible. However, the data is still examined in order to test the hypothesis. The data obtained is binomial non-parametric count data, why a Fisher's Exact test is applied in order to analyse the differences further. This test reveals that no significant differences in answers exists between the two colour conditions in either of the questions (Turn-off p-value = 1.00 and Pop-up p-value = 0.3416), why hypothesis H.FM5.1 is rejected and null hypothesis N.FM5.1 is accepted.

To assure no effect of gender has influenced the results a Fisher's Exact test is made. This shows that no significant effect of gender has occurred (p-value = 0.4545).

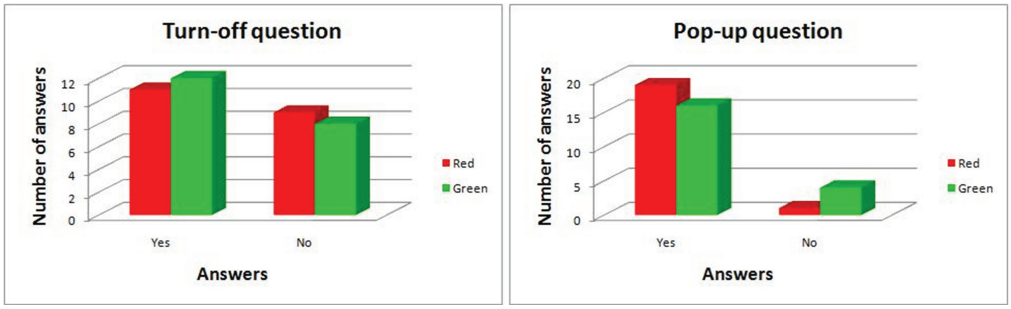


Figure 7.32: *Graph illustrating the distribution of answers in question 1 between the red and green condition.*

Discussion

Hypothesis H.FM5.1 was rejected based on the results from the experiment. No significant influence could be found of the colour of red and green on the answers. This indicates that colours cannot be used to frame users' choice, or at least the colours red and green cannot. However, another explanation could be that several nuisance variables have influenced the results and masked the influence of the colour framing. One nuisance variable could be that both buttons in each of the questions were being framed with the same colour. Adding the colour green to both the negative and the positive answer might have made both answers appear more positive instead of having the test subject choose the positive answer. This might have made the negative answer seem more attractive since it would appear less negative, and vice versa with the red colour on the positive answer.

Another, more obvious nuisance variable could be that the two questions are not neutral, which was not tested. The colours might therefore have an effect which cannot be observed due to no control condition results. Finding fully neutral questions to examine is rather difficult, as test subjects are influenced by many different parameters such as experience. One might be annoyed with pop-ups why he rejects that option, whereas another test subject likes to keep account of things in general and therefore prefer pop-ups. If all of these considerations do play an active role in this experiment, it can be argued that in order to test if colours do affect behaviour in such a situation, it should be tested in an authentic setting, where there is a direct purpose with setting the system, which is less apparent for test subjects in the experimental setting..
Therefore, based on the present experiments it can be concluded that colour framing under such circumstances does not work, however, given an authentic setup it could prove otherwise.

7.6 Framing Experiment Recapitulation

Five experiments were made in order to test if the concept of framing can be applied visually in a user interface. Two main types of framing were applied in order to examine the subject from different perspectives. The overall results and their application prospects will be discussed here as well as further ideas for examination of the subject in use settings. An overview of the hypothesis and their state of acceptance is illustrated in figure 7.33.

Hypothesis	Content	State
H.FM1.1	*A significant difference in slide bar adjustment will be observed between the two different slide bar conditions*	Rejected
H.FM2.1	*A significant difference in slide bar adjustment will be observed between the two different slide bar conditions*	Rejected
H.FM3.1	*Test subjects in the colour conditions assess the degree of difficulty of the questions significantly easier than test subjects in the control condition*	Rejected
H.FM3.2	*Answer options chosen in the colour conditions are significantly different from answer options chosen in the control condition*	Rejected
H.FM4.1	*Answer options chosen between the five different colour conditions are significantly different within each question*	Rejected (However, some effects recognised)
H.FM5.1	*A significant difference will be found between answers in the red condition and answers in the green condition*	Rejected

Figure 7.33: *Hypothesis status overview.*

The first type of framing tested was that of "physically" manipulating the frames for making a decision, directly inspired by the ice cream experiment made by [Hsee, 1998]. However, no effect of slide bar length (frame) was found when test subjects had to estimate a parameter on two differently sized slide bars, or when one parameter had to be decided on a slide bar. It seems that test subjects were perfectly capable of keeping the goal in mind instead of obeying the frames, directly opposite worded frames as shown in the Asian disease problem for instance.
The results thereby imply that users are not susceptible to frames when it concerns actions they have to perform in order to execute a task. If this is the case user interface designers are not at risk of involuntarily affecting users, at least in the case of adjusting the lengths of slide bars for i.a. aesthetic reasons.
As the results obtained from experiment FM1 and FM2 did not support the hypotheses it would be easy to conclude that no effects could be observed, but the setups in which the slide bars were tested may not represent common use situations. For experiment one the task was to estimate a given parameter using the slide bar. As mentioned in section 7.1 this experiment was directly translated from the experiment made by [Hsee, 1998] and was originally planned to be tested manipulating the maximum point of a boiler kettle, to observe if test subjects would then poor more water into it. The latter is more of an applied situation, but proved to hold some difficulties in terms of the experimental part. Nonetheless, it would be interesting to try to apply this type of framing in real world use settings as observational experiments, since one like the boiler kettle probably would also provide blurred data given an experimental setting.
Overall it can be concluded for the slide bar framing that it does not work in the settings tested in the present experiments. However, this result will in some instances benefit the user, as he cannot be manipulated to decide differently than he would otherwise have. On the

other hand it does not provide user interface designers with the ability to use the interface construction as a direct guidance, at least not in terms of slide bar settings.
Referring back to section 3.1 concerning decision making, then it seems as if test subjects act "rational" according to normative principles in use situations, as they do not make reversed judgments and decisions based on different frames, which might in fact disprove applicability of results found in setups such as the Asian Disease setup.

The second type of framing tested concerned colour framing. It was hypothesised that the colours connected to a button would affect choices made. However, the first experiment put forward did not show this effect in general, only the colour blue let test subjects to refrain from choosing the answers they chose in the no-colour condition. Also, it could be seen from the results that positive and negative questions were more affected by colour than neutral questions, however not significantly. Despite these results no other clear tendencies could be identified, which might have been caused by test subjects being aware of the colours they chose, but also the fact that not enough test subjects had been employed in relation to the amount of factors present in the experiment. This in turn made the data rather complex why tendencies can have been masked. Therefore, a second experiment was made, where the variables were kept more constant in order to diminish the complexity of the results. However, the results also did not show any general signs of colour as the number of observations for each colour was rather small, why four test subjects for example, could provide a significant result. Of course it may be that colours do not affect our choices, but the setup might as well have caused the non-result. The set up questions were designed to be neutral, but some of them showed not to be which can have caused a nuisance. Also, the setup was not use related as such, why colour potentially could provide an effect in another setting. Exactly for this reason and in order to test colours in a more use based domain a third experiment was setup where the setting was more use related since it i.a. concerned deciding upon a pop-up service, which is also seen in anti-virus software for example. This experiment was kept very simple, in order to be able to detect an effect of colour if one such exists. However, once again the hypothesis was rejected which can be caused by the questions not being neutral, test subjects being mainly technical students, not testing between colours etc.

The implications of these three colour experiments is that colours in general do not seem to affect our decisions, when posed with choice questions with multiple coloured answer options. However, concluding finally based on these experiments is not possible as they might have been affected by an interrelation effect of colours, a too much choice overload, the peculiar questions, the experimental setups etc. Furthermore, colour did not seem to affect choices in the final use test setup, however, this might also be affected by some of the factors mentioned earlier. It can be discussed whether colour is actually a mean to frame with, in the way words are. However, if assumed that colours are culturally conditioned as displayed in figure III.3 in Appendix III, and if the meaning attached to the colour generally applies, these meanings should act as reference points. In general, the results obtained in these experiments do not point to such an effect of colour however, other ways of testing this could potentially prove the effect in other situations given other circumstances. Therefore experiment FM5 was made, but it did not show any effects of colour. Instead, it could be interesting to test whether the colours tested in experiment FM5 can affect users of external hard drives for example. When an external hard drive is nearly full (20 Gb's left out of 250 Gb's for example) the progress bar in the "my computer" window shows this status. However it changes colour, and when there is still 20 Gb's left it turns red signalling that an action soon needs to be performed. It could be interesting to manipulate the colour settings of this bar so it would occur at different limits, and observe if test subjects would react solely based on the colour or on the actual memory state of the hard drive.
This situation is more directly use applied and could be interesting to test in the future. This far however, the question if colour can be used for framing remains somewhat open as different results were obtained.

CHAPTER 8

Default Experiments

In this chapter two experiments concerning defaults will be presented. The first experiment explores the effect of three different types of defaults in an unfamiliar use setting, whereas the second experiment examines two default types in a familiar use setting. The chapter concludes with an experiment recapitulation for both experiments.

8.1 Experiment DF1

The research examined concerning defaults (see chapter 4.3) found that defaults can affect decisions in multiple ways. For example when making a simple choice and one of the options are preset we seem to rely on the preset to be a standard, which makes it easier to not make the choice. Other interesting phenomenon's of defaults were found by i.a. [Samuelson and Zeckhauser, 1988] concerning allocation- and too much choice defaults. The first one covers that the "value" of a default affects the choices made to center around the original default value. The too much choice default affects choice in such a way where the amount of choices makes us even more prone to choose the default.
These different aspects of the default effect could potentially affect users' behaviour when interacting with a product. Defaults are also found in many use contexts such as on the internet on websites, when a computer is started for the first time, printer settings etc. However, if we are in fact affected by these defaults in actual use situations, meaning that we do not adjust settings simply because they are already adjusted beforehand, then defaults can act as a very powerful tool for manufacturers of all kinds of use dependent products. Therefore this experiment seeks to examine if we are affected by defaults in an actual use situation and if the different types of defaults affects us in use situations as shown earlier in questionnaire setups [Samuelson and Zeckhauser, 1988] and [Johnson et al., 2002].

Experimental objective

The objective of this experiment is to test if defaults affect use choices. The default types which are tested is the "common" default where one out of two options is preselected, the allocation default where the value of the default affects the following choice, and the too much choice default where multiple choices have shown to impair decision making.

Hypothesis

The overall hypothesis applies for all three types of defaults and states that:

Defaults affect decision patterns in a user interface.

If this hypothesis applies it will emphasise that it is important to consider which settings are preset in order to ensure the best user experience possible. Furthermore, it can be used as a tool for direct manipulation if users rely on the preset settings.

Apparatus

The test setup is made in Visual Basic 2008 Express Edition, and is run on a MSI 340x laptop with a 13.4" screen. The test is made in a group room at Aalborg University with the curtains closed. All stimulating objects, pictures etc. are removed from the room. Camtasia, a screen capture program, is used to record the screen during the test.

The default setup is made as part of the experimental setup used in Experiment ST3 (see section 6.1). The defaults are set up on an initial screen, where test subjects are to adjust settings for a plastic container factory machine, see figure 8.1. In this way defaults are tested in a use setting, while at the same time setting the scene for the sunk time experiment reported earlier.

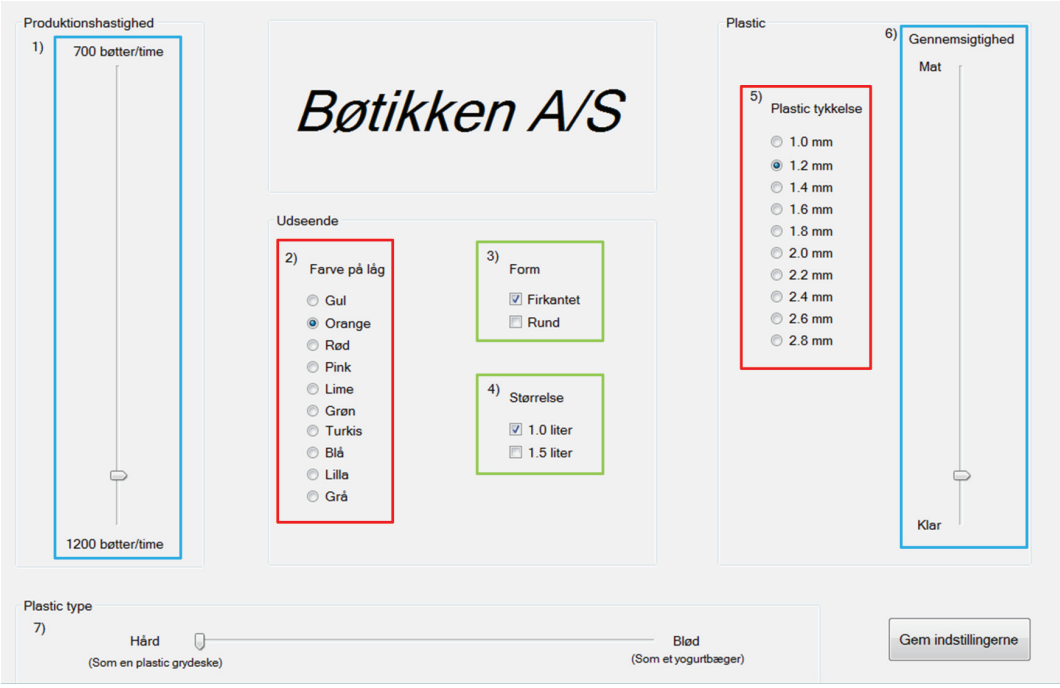


Figure 8.1: *Screenshot of the factory setting setup presented to test subjects with colour frames added for explanatory reasons.*

The test subjects are to follow the small numbers at each setting which makes it possible to control their course of action. For this reason Camtasia is utilised to record the screen in order to control that the sequence is followed by each test subject.
The different types of defaults are shown in figure 8.1, where the options highlighted with green represent the "common" defaults. The options highlighted with blue represent allocation defaults where the setting of the adjuster acts as a default value, which is quantifiable. The last highlighted options in red represent too much choice defaults where more than six options are present, as was reported as a guiding "upper" limit by [Iyengar and Lepper, 2000], see Appendix I for more on the too much choice effect.

The slide bars used are denoted "TrackBar" in Visual Basic. Each of them measures 500 pixels and is set to a value of 1000 steps. In the first condition the default value of the adjustment bar is 100, and is set to 900 in the second condition. The common defaults are made using "CheckBoxes". In the first condition the upper boxes are checked, whereas the lower boxes are checked in the second condition. Finally, the allocation defaults are made using "RadioButtons" with ten options. In the first condition radio button number two

from the top is marked, and in the second condition it is radio button number 9 from the top which is marked (see figure 8.2).

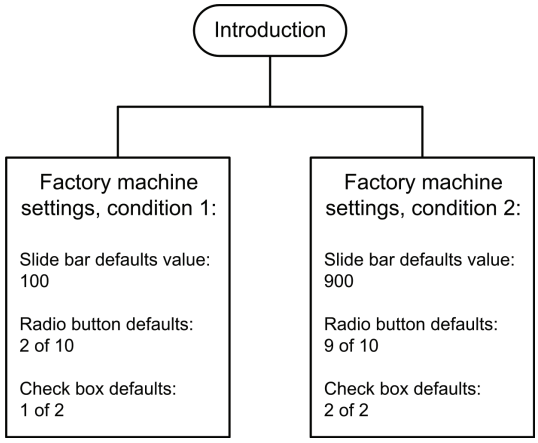


Figure 8.2: *Experimental course overview.*

The categories in which the different defaults are used are all chosen according to what would fit the type of default. As an example it would not make much sense to adjust on a slide bar which colour a lid should have, unless it was a gradient of one particular colour, since it is a continuous scale with defined steps.
Furthermore, two default setups are made for each type of default in order to be able to detect if it is a coincidence that a default is always or never chosen for a specific default setup.
The reason all default types are situated together in one screen is that test subjects should get the impression of adjusting machine settings.

The experimental setup is a between subjects experiment, as one group of test subjects are exposed to one set of default settings, whereas the other test group is exposed to another set of default settings. The independent variable is therefore the default settings (condition 1 and condition 2), whereas the dependent variable is the registered settings for each of the adjustment parameters.

For the developed program, see Attachment H and the attached DVD.

Procedure

Test subjects were placed in the group room in front of the computer. They were asked to remove their mobile phone and watch in order not to disturb the test, as well as they were told how the test would progress (see Attachment B). An introduction text was placed besides the computer (see Attachment C), which the test subjects were asked to read. The experimenter then pressed a computer button which initialised Camtasia (screen recording), without the test subject being aware.
The test subject was then left alone in the group room, and were asked to come out when having pressed the “Afslut” button.

Afterwards, the test subjects were rewarded with cake and asked to sign a declaration of consent (see Attachment D) since their mouse clicks had been captured, which all of them did.

Test hypotheses

The test hypotheses put forward relates to each of the types of defaults tested. However, the first hypothesis applies for all of the default types, and states that:

H.DF1.1 *The default will be chosen significantly more times than the non-default for each of the setups*

N.DF1.1 *The default will not be chosen significantly more times than the non-default for each of the setups*

All of the default types should affect the test subjects to just keep the default setting, why this hypothesis accounts for all. The second hypothesis is specific to the allocation default types, as they should affect those test subjects who actually defies the default effect to adjust the adjustment bars according to the default position. The hypothesis states that:

H.DF1.2 *For the incidents where the slide bars, representing the allocation default, are adjusted the default will affect the position of the adjustment bars significantly*

N.DF1.2 *For the incidents where the slide bars, representing the allocation default, are adjusted the default will not affect the position of the adjustment bars significantly*

The third hypothesis is specific to the too much choice defaults, as they are expected to impair test subjects' decision making even more than the others, as the number of options exceeds six. Besides being tested according to hypothesis H.DF1.1, it should also be able to withstand the following:

H.DF1.3 *The default in the too much choice setups will be chosen at least 95% of the time*

N.DF1.3 *The default in the too much choice setups will not be chosen 95% of the time*

As mentioned before, the implications towards manipulating and controlling user behaviour via defaults can serve as a powerful tool if people also "obey" the default phenomenon in use settings.

Participants

40 students participated in the experiment (mean age = 23.98±3.17 years). The gender distribution was 17 women and 23 men. All test subjects signed the declaration of consent.

Results

The first hypothesis H.DF1.1 concerns all of the six tested default setups, why each pair of opposite defaults will be presented in order to examine if they comply with the hypothesis. Results of the "common" defaults, which were tested using checkboxes, is non-parametric count data and shown in figure 8.3. As can be seen from the graphs the defaults do not seem

to have any impact, instead it seems that test subjects' preferences control their choice in this situation. Despite test subjects choosing the default more in half the cases the effect is eliminated as they continue to choose the same shape or size even though it is not the default. Hypothesis H.DF1.1 is therefore rejected for the common checkbox default setups.

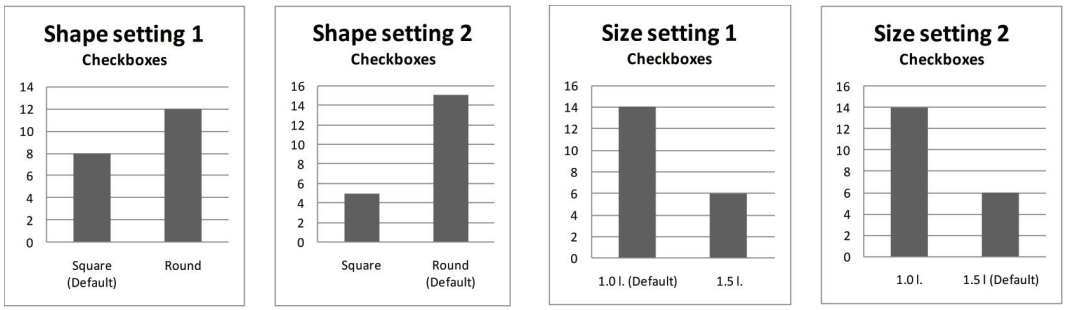


Figure 8.3: *Overview of the checkbox default setup results. 1 and 2 denotes condition 1 and 2, respectively.*

In figure 8.4 the results for the allocation default setups are presented. As can be seen, the default is only chosen in a minority of the cases, why it is easy to reject hypothesis H.DF1.1 for the allocation slide defaults as well. However, it should be noted that the purpose of this type of default is not necessarily to be chosen, but to affect the choice, why this default is tested in relation to H.DF1.2 later.

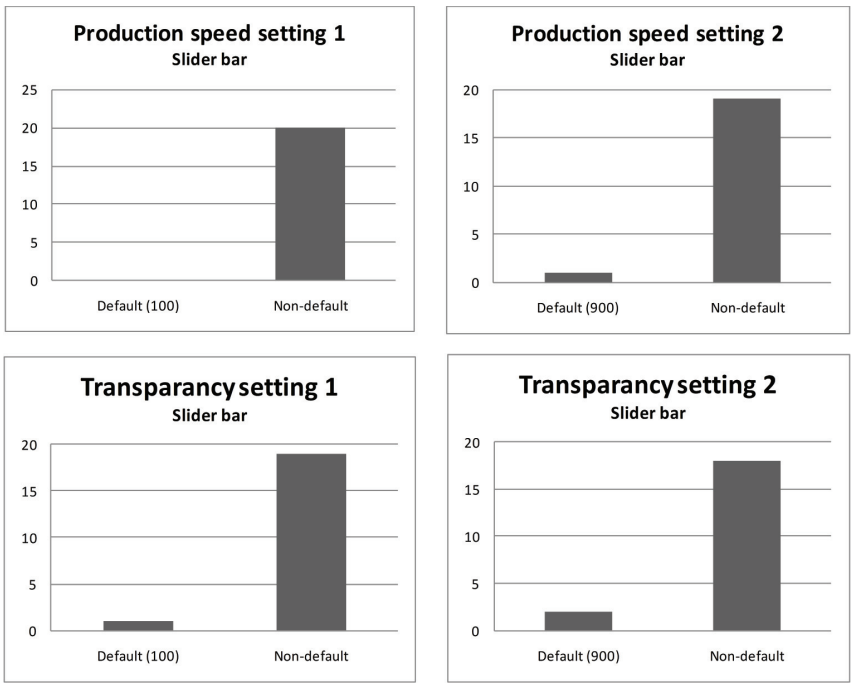


Figure 8.4: *Overview of the slide default setup results. 1 and 2 denotes condition 1 and 2, respectively.*

The final default setup to be tested is the too much choice radio button default. If this default type does not comply with hypothesis H.DF1.1, then hypothesis H.DF1.3 will also be rejected, since acceptance of H.DF1.3 requires acceptance of H.DF1.1. In figure 8.5 the results from the too much choice radio button conditions are shown. Since ten options were provided in this default the results are shown as default and non-default, in order to be

tested for compliance with hypothesis H.DF1.1. Despite this, it is easy to see that the hypothesis is also rejected for this case, as none of the test subjects choose the default colour in any of the conditions, and the thickness default is only chosen in a minority of the instances.

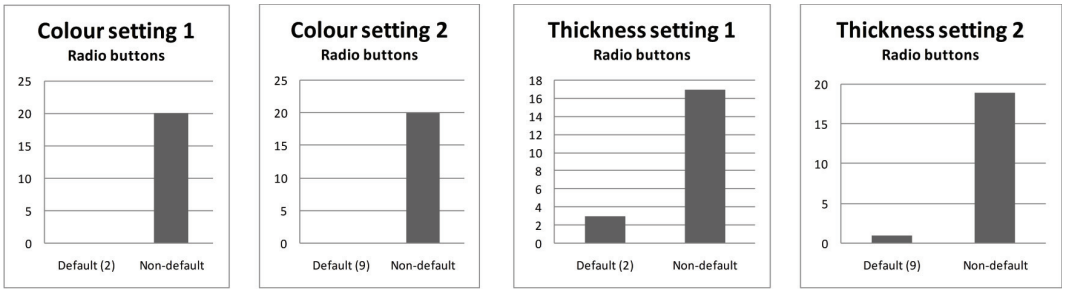


Figure 8.5: *Overview of the radio button default setup results. 1 and 2 denotes condition 1 and 2, respectively.*

Looking at the colour distribution overview in figure 8.6 it seems that some of the colours were more popular than others, and that it is this preference which has controlled the test subjects' choices. This shows that in a use situation like this, test subjects do make active choices.

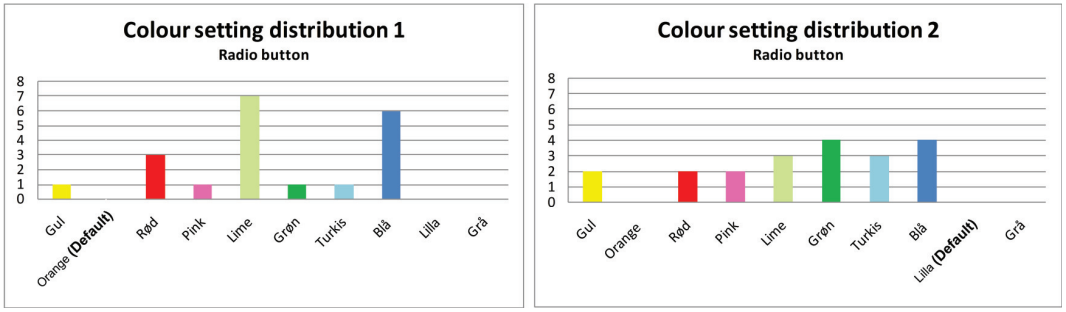


Figure 8.6: *Overview of colour distribution in the too much choice radio button default setup.*

Both hypothesis H.DF1.1 and H.DF1.3 have been rejected, why the remaining hypothesis H.DF1.2 still needs to be examined. This hypothesis concerned the allocation default, and assumes that a default allocated value will affect the value choice of the user.
The hypothesis is tested using a t.test to test for difference between the two conditions. In figure 8.7 and 8.8 qq-plots are made in order to check that data is normally distributed, which it roughly is.

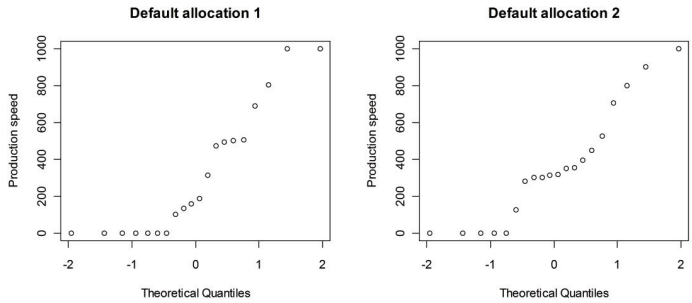


Figure 8.7: *QQ-plots for the allocation defaults condition 1 and 2 for the production speed question.*

Two t.tests are run on the respective data in order to test if test subjects for example exposed to the condition where the slide bar value was close to 1200 containers per hour choose op-

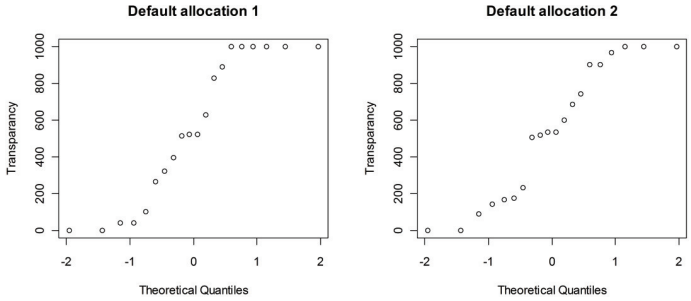


Figure 8.8: *QQ-plots for the allocation defaults condition 1 and 2 for the transparency question.*

tions nearer this end, than test subjects exposed to the condition where the slide bar value was closer to 700 containers per hour. For either of the production speed or the transparency no significant difference was found with p-values 0.7128 and 0.8728, respectively. In figure 8.9 the expected type of outcome is shown, and in figure 8.10 the actual outcome for each of the four slide bars is presented, which clearly shows no connection. Hypothesis H.DF1.2 is as a consequence also rejected, and the alternative hypothesis N.DF1.3 accepted.

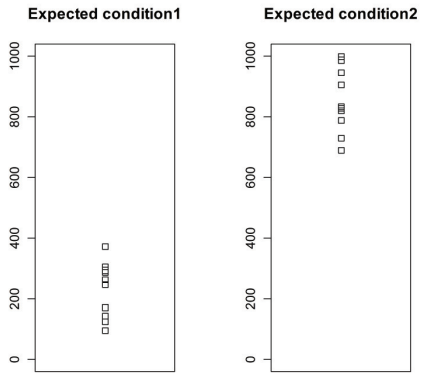


Figure 8.9: *Expected adjustment distribution pattern.*

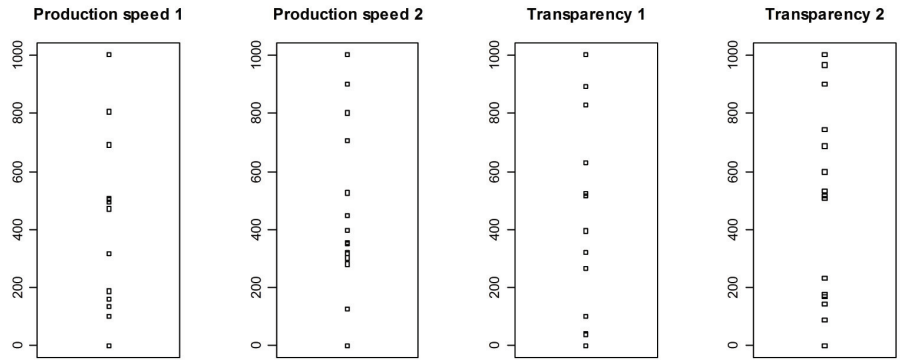


Figure 8.10: *Actual results from the slide adjustments.*

Testing for gender differences revealed that an effect of gender occurs in two out of the six default setups (p=0.04741 and p=0.03691). When it comes to choosing colour of the lid 50% of the men chose a blue lid, whereas none of the women did. For the transparency of the plastic container women in average wanted more transparent containers than men with a difference of 24%. However, even though the gender differences are interesting, they do not change the outcome since almost none of the defaults were chosen anyway, which indicates that a certain gender is not more prone to follow defaults than the other.

Discussion

None of the hypotheses were accepted why it is tempting to reject the general hypothesis that "Defaults affect decision patterns in a user interface" as well. However, many different influencing factors may have caused the observed non-result which will be discussed.

The setting of the experiment is supposed to imitate a use situation, which can be discussed since test subjects do not have a naturally motivated interest in making a plastic container. Therefore, the given setup can be critiqued in terms of direct applicability to a conventional use setting, where the user has a purpose with the interaction. To create a real motivation in an experimental setting is not easy and in this case the motivation was the award in the form of a piece of cake upon completing the experiment, why the mean to get a piece of cake is not to produce a plastic container, but to complete the experiment. Adding to the notion of motivation is the fact that it is an experimental setting, why the test subjects are not risking anything by making a "wrong" container.
Also concerning the setting of the experiment is the subject - a food container factory. The majority of the test subjects most likely did not have a clue about what plastic settings are appropriate etc., but it was assessed that most people are familiar with food containers why it is a relatively easy subject to handle. That test subjects did not know about the technical settings was somewhat indifferent to the hypotheses, since the interesting part was to observe if they would "obey" the default phenomenon observed by others. The results clearly showed that no default bias appeared in the chosen experimental setting.
That the user interface itself is unfamiliar to the test subjects may also affect them to consider each option more thoroughly, than if the interface was familiar to them. When faced with new inputs we are forced to consider the input in another way than if we are used to seeing a certain interface. In known situations we have learned how to react and what to expect why we do not necessarily need to consider the inputs once again. Therefore, for the test subjects to be exposed to an unfamiliar interface may have caused them to consider things more in depth than they would have if it was a familiar interface. On the other hand, it would be expected that test subjects then would trust the defaults even more. Therefore, it may be that exactly the experimental setting has played yet another role, as the test subjects might have felt obligated to adjust the settings, since it seems plausible that they have had the thought that it was what was expected of them given the experimental setting.

Test subjects did adjust the settings despite whatever of the mentioned reasons, and afterwards some of them mentioned their motivation for changing some of the adjustments. For example the colour setting was commented a lot on; some test subjects chose the blue lid as they associated blue with food hygiene, others chose a green lid as their IKEA food containers at home had green lids etc. These reports show that many different factors affect a decision e.g. experience and associations. Therefore it is difficult to control a decision experiment since multiple internal factors affect test subjects' decision. However, the aim with this experiment was to observe if test subjects did make an active choice or just went with a preset, which was shown despite the different reasons influencing the choices.

The three different default setups can also be discussed in terms of their effect. First is the common default setup where one of two checkboxes is checked by default. This is often seen at the end of questionnaires where such a checkbox is checked beforehand e.g. to subscribe to a newsletter or "accept terms and conditions". This was shown by [Johnson et al., 2002] in a questionnaire context, however in the context of the present experiment test subjects seem to choose according to preference, as most test subjects wanted round and 1 l. containers in both conditions.
For the allocation default almost none of the test subjects chose the default, which does not comply with the findings made by [Samuelson and Zeckhauser, 1988] in their questionnaire setup. Furthermore, linking the specific allocation setup in this experiment where slide bars

were used, they can in principle also be associated with a too much choice default, since the number of choices are “infinite” given a slide bar setup. Combined with a too much choice default effect test subjects should have been even less prone to adjust the slide bar. However, it was not the case in the present experimental setting, where test subjects made an active choice.
Finally, the too much choice default effect was also undermined given the experimental setting as test subjects did make a choice. Especially in the colour setting it was obvious that test subjects followed their own preference as none of the defaults were chosen.
Whether these result deviations from previous research made in other settings only applies because of the current experimental setting, is suspected, as multiple factors most likely have influenced the outcome. The influencers being as mentioned, the unfamiliar situation, no real purpose, different experiences etc.

During the progress of the current project another experiment was setup where two of the default types were tested. However, the experiment was discarded for multiple reasons as it was designed as a composite experiment including other decision subjects. Nonetheless, two of the default types applied in this experiment were used in that experiment in a more familiar use setting - that of a program installation process. Enough data was gathered in order to analyse the default part, why this will be treated in the next section as an independent experiment, which adds to the results found in this experiment, as it deploys another type of use setting which will supply a broader foundation for interpreting defaults in use cases.

8.2 Experiment DF2

As discussed in the previous experiment (section 8.1) the experimental setting may have influenced test subjects to make forced choices as a consequence of believing that it was the purpose of the experiment, when in fact it was to observe if people are prone to settle with defaults. On the other hand it might be an expression of different decision mindsets given the interaction setting; nonetheless, it is still believed that defaults affect us in some use settings, why this experiment seeks to test the phenomenon in a more familiar use setting.

Experimental objective

The objective of this experiment is to examine defaults in a well-known use setting where test subjects understand the consequences of their choices. However, it is still difficult to provide a motivational factor for the test subjects given the experimental setting, why a program installation setup is chosen, as the installation is a mean to use a program, why there is a motivation to follow the installation steps and install the program, as the test subjects would normally do.

Hypothesis

The general hypothesis is almost the same as in the later experiment; however it is specified a bit more in detail according to a familiar use setting.

Defaults affect decision patterns in a familiar user interface.

By the phrase "familiar" an implication is made that there is a difference between user settings and reactions to defaults in familiar and non-familiar contexts. From the results of the previous experiment it was seen that the defaults in an unfamiliar interface given an experimental setting did not support the hypotheses, why this experiment seeks to eliminate both to some extent, using a familiar installation setup as a mean (whereby degrading the experimental setting feel) to evaluate a program.

Apparatus

The test setup is made in Visual Basic 2008 Express Edition, and is run on two laptops; a MSI 340x with a 13.4" screen and a Fujitsu Siemens Amilo si1520 with a 12.1" screen. The test is made in a group room at Aalborg University with the curtains closed and no disturbing elements.

An installation setup is made based on a classic Windows Setup Wizard, see figure 8.11 for screenshots from the program. The installation is done in four steps as indicated by: A. Adjust the space the program will take up (slide bar - allocation default), B. Adjust the grade of background colour (slide bar - allocation default), C. Select which type of short-cut to create (checkboxes - common default) and D. Select installation folder (checkboxes - common default).

The slide bars (denoted TrackBar in VB) are of measures 380 pixels, has 1000 steps and the default value in condition 1 is 100, whereas it is 900 in condition 2. Three checkboxes are used where the first checkbox is checked by default in condition 1 and the third in condition 2 (see figure 8.12 for an experiment overview). Three checkboxes were chosen due to the original setup of this particular experiment, which were to contain three conditions.

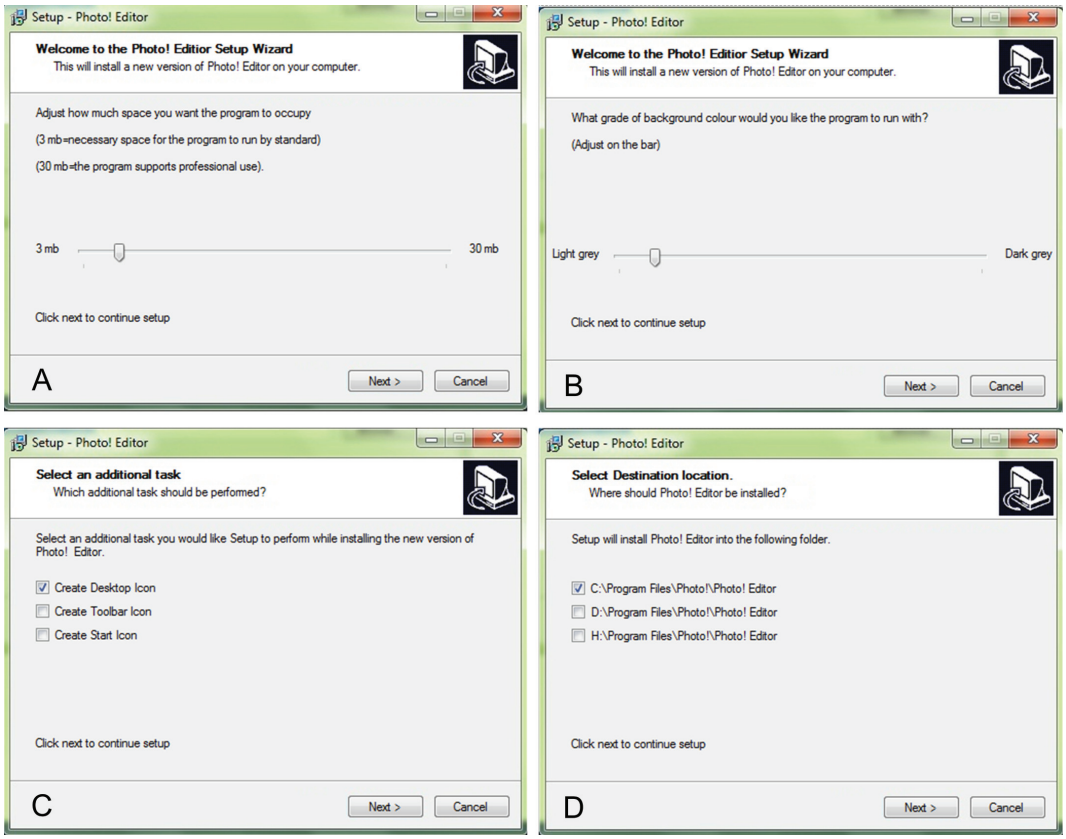


Figure 8.11: *Screenshot from the installation setup in condition 1.*

However, only two of the conditions were tested before the remaining of the experiment was discarded. The findings from the two conditions however, do still provide grounds for testing the hypothesis in this type of setting.

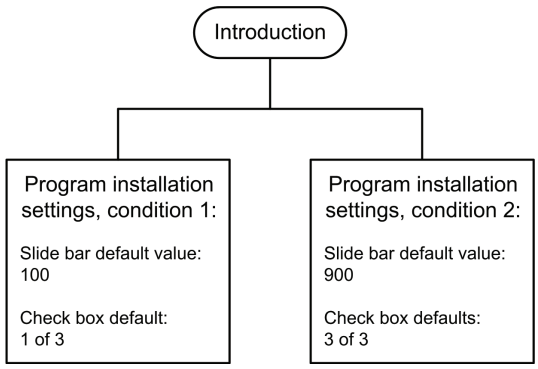


Figure 8.12: *Experimental course overview.*

The installation setup is the only scenario the test subjects have to go through. However, beforehand they will be told that they need to evaluate a photo editing program, which they have to install as well in order to be able to judge the program as a whole. Furthermore they are told that the installation together with the program are based on a new system, why the installation is a crucial part of the program feel (see Attachment B) for the introduction manuscript).
In this way the test subjects will install the program as they would do for themselves, and their suspicion towards the installation is expected to be kept at a minimum.

The experimental setup is a between subjects experiment, as one group of test subjects are exposed to one set of default settings during the installation, whereas the other test group is exposed to another set of default settings. The independent variable is therefore the default settings (condition 1 and condition 2), whereas the dependent variable is the registered settings for each of the adjustment parameters.
For the developed program, see Attachment O and the attached DVD.

Procedure

The test subjects were placed in the group room in front of the computer. The introduction found in Attachment B was given to them, where after the experiment was initialised.

Test hypotheses

The test hypotheses put forward resembles those from the previous experiment (see chapter 8.1) and relates to each of the types of defaults tested. The first hypothesis applies for all of the default types.

H.DF2.1 *The default will be chosen significantly more times than the non-default for each of the setups*

N.DF2.1 *The default will not be chosen significantly more times than the non-default for each of the setups*

All of the default types should affect the test subjects to just keep the default setting, why this hypothesis accounts for all. The second hypothesis is specific to the allocation default types, as they should affect those test subjects who defies the default effect, to adjust the adjustment bars according to the default position. The hypothesis states that:

H.DF2.2 *For the incidents where the slide bars, representing the allocation default, are adjusted the default will affect the position of the adjustment bars significantly*

N.DF2.2 *For the incidents where the slide bars, representing the allocation default, are adjusted the default will not affect the position of the adjustment bars significantly*

Participants

30 students participated in the experiment (mean age = 22.43±2.54 years). The gender distribution was 12 women and 18 men.

Results

All of the four tested default setups need to be tested according to hypothesis H.DF2.1. An overview of the results can be seen in figure 8.13 and 8.14. As can be seen for the allocation defaults (slide bars), the default is not chosen more times than the non-defaults why hypothesis H.DF2.1 does not apply for this type of default.

For the rest of the default setups (the checkboxes), the default is chosen in the majority of the cases, why they are tested to see if they comply with hypothesis H.DF2.1. The NA in "Icon setting 2" refers to two test subjects who did not choose any of the icon settings, as they unticked the default checkbox, which states an active choice opposite the default, why

Figure 8.13: *Overview of the slide bar default setup results. 1 and 2 denotes condition 1 and 2, respectively.*

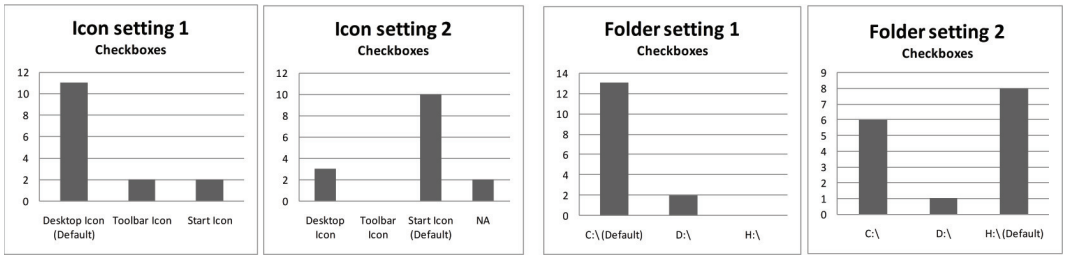


Figure 8.14: *Overview of the checkbox default setup results. 1 and 2 denotes condition 1 and 2, respectively.*

it is counted as a non-default in the later analysis. Since the data is non-parametric count data with most of the measures below the value of ten, a Fisher's Exact test is applied in order to test for statistic significant difference between the observed selections of defaults and non-defaults. An overview of the data tested in the Fisher tests is presented in figure 8.15.

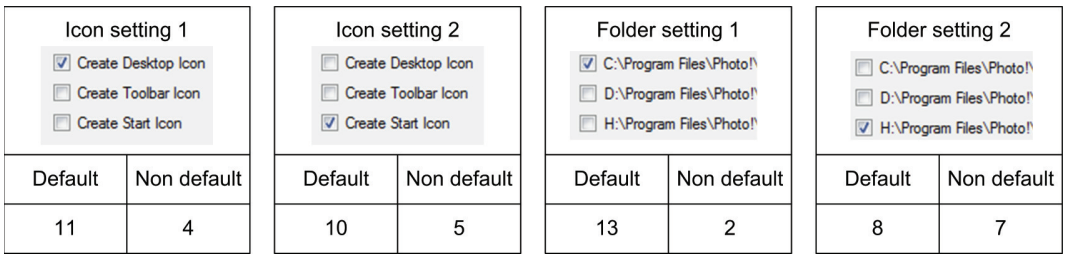


Figure 8.15: *Overview of checkbox defaults results.*

The data is tested using a Fisher's Exact test, and the results can be seen in figure 8.16. For half of the incidents the default was chosen significantly more times than the non-default. Despite all of the defaults being chosen more times than the non-defaults the small sample size makes it difficult to calculate a true p-value, since scaling up the numbers by a factor ten in the case of Icon setting 2 for example, provides a significant p-value. Furthermore, concerning the folder settings, it seems that many test subjects are aware that the C drive is the common folder for program files, why this affected the outcome of folder setting 2. The latter shows, that most test subjects trust the default, however, many do make an active choice, despite the situation being familiar.

Even though the allocation default did not comply with hypothesis H.DF2.1, it is tested to see if it complies with hypothesis H.DF2.2, which is specific for this type of default. The hypothesis is tested by sorting the slide bar data in groups of "defaults" and "non-defaults". A t.test is then performed on the non-defaults between conditions for the same choice.
Before performing the t.tests the data is checked to see if it is normally distributed, see figure 8.17. Given the amount of data the distributions are accepted. Performing the t.tests

Default	P-value	H.DF2.1
Icon setting con.1	0.02684	Accepted
Icon setting con.2	0.14310	Rejected
Folder setting con.1	0.00015	Accepted
Folder setting con.2	1.00000	Rejected

Figure 8.16: *Results of the Fisher test.*

show that hypothesis H.DF2.2 is rejected, as the p-values for the space and colour setting equaled 0.1767 and 0.7995, respectively. This is also illustrated by figure 8.18, where it is obvious that the data is spread across the range instead of within proximity of the given default value.

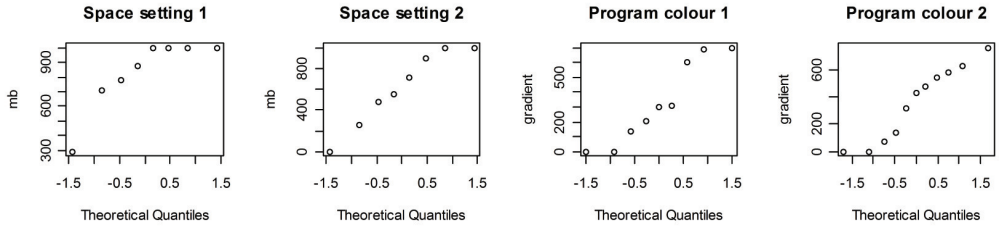


Figure 8.17: *QQ-plots of the allocation default setup results.*

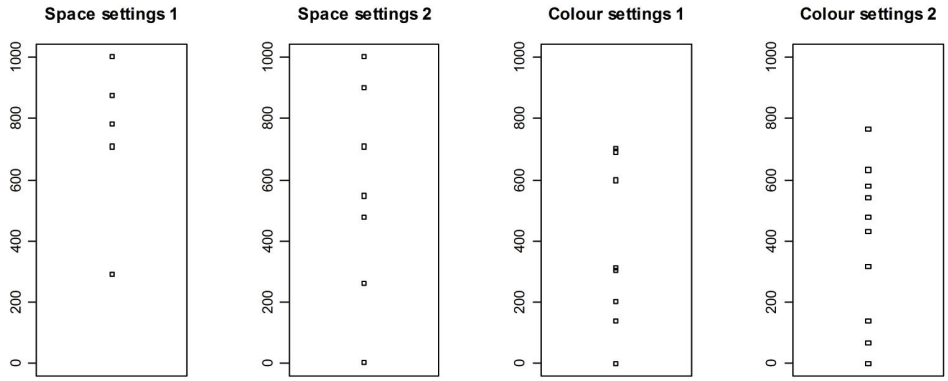


Figure 8.18: *Overview of adjustment distribution pattern (for the first condition the default value was 100, whereas the default value for condition 2 was 900).*

Discussion

The majority of the hypotheses were also rejected in this experiment; however the defaults seemed to have a larger impact in this type of setting than in the former experiment (see chapter 8.2). This could be an indication of the effect of the experimental setting as well as the familiarity with the user interface. Concerning the setting of the experiment, test subjects might not have felt as obligated to adjust the installation settings as they did the settings in the factory experiment, since they believed that it was a mean to get to the program. Therefore, test subjects most likely acted as they would when installing a program on their own computer. This shows that the defaults do play an important role, especially the check box defaults where preferences were reversed between conditions (as shown by [Johnson and Goldstein, 2003]).
It is also interesting to notice from the present results that more test subjects in general

chose the default in the checkbox setup in relation to the slide bar setup. This might be caused by the slide bar setup being unusual in an installation process, hence test subjects have put more thought into the input exposed to, as was discussed in the former factory experiment.

The allocation default phenomenon does not seem to apply in use settings, where the purpose of the use over-shadows the default value. This might be caused by the unclear default settings, at least for this and the latter experiment, where no anchors were supplied on the slide bar. Also, the allocation default slide bars function like the framing slide bars tested in experiment FM1 and FM2 (see section 7.1 and 7.2), where no effect of framing was found either. This provides even more evidence that default allocation in use is not effective as the user, when deciding upon changing a given factor, do not consider the later value. This can of course also be a consequence of the fact that when adjusting the slide bar the previous default is no longer visible, and the default value is not as easily remembered due to the diffuse values on a slide bar.

In relation to a notion of making "irrational" or "rational" decisions the picture appears blurred as only hypothesis H.DF2.1 was accepted for two out of eight incidents, and the use setting provided influenced test subjects' decision patterns. This could imply that different decision mindsets are used in different situations; in familiar settings the choices are not evaluated as thoroughly as in unfamiliar settings.

The implications of these results together with the results obtained in the previous experiment will be held together and discussed in the following recapitulation on the present empiric investigations made on the default phenomenon.

8.3 Default Experiments Recapitulation

Two experiments were made regarding defaults in use situations. In order to gain an overview of the results, the tested hypotheses with their state are presented in figure 8.19. Almost all of the hypotheses were rejected and reasons as to why have been discussed in the respective experiment discussions, why this section instead seeks to focus on the influence of the current result on user interface design, as well as possible future examination of defaults.

Hypothesis	Content	State
H.DF1.1	*The default will be chosen significantly more times than the non-default for each of the setups*	Rejected
H.DF1.2	*For the incidents where the slide bars, representing the allocation default, are adjusted the default will affect the position of the adjustment bars significantly*	Rejected
H.DF1.3	*The default in the too much choice setups will be chosen at least 95% of the time*	Rejected
H.DF2.2	*The default will be chosen significantly more times than the non-default for each of the setups*	Accepted (for checkbox 1 and 2 in condition 1) Otherwise rejected
H.DF2.2	*For the incidents where the slide bars, representing the allocation default, are adjusted the default will affect the position of the adjustment bars significantly*	Rejected

Figure 8.19: *Overview of hypothesis status for default experiments.*

In general, the results of experiment DF1 showed that default settings are not important for users. Whether this was caused partly by other influences such as the experimental setting is not apparent. The results from the second experiment showed a slightly different tendency, as defaults in some of the cases seemed to affect user decisions. The challenge is then to interpret the results in order to be able to apply them in the design of user interfaces.

The main difference of the two experiments was the familiarity of test subjects with one of the setups as opposed to the other setup. When test subjects were able to recognise the frames (in this case Windows Setup Wizard), they relied more heavily on the default settings, which was argued by Samuelson and Zeckhauser (1988) is a way to avoid the cognitive strain connected with decision making. As experiences with a choice have been positive, then it makes sense to go with that option again because one knows what to expect and that this option is sufficient, despite potentially missing out on an even better option. This again can be connected to Prospect Theory (see section 3.1) where the potential loss of switching option weigh more heavy than the potential gain. Of course, given the experimental setting it might as well be cognitive laziness, as the task in principle is to no concern of the test subject.

Contrarily, when test subjects were exposed to the unfamiliar interface they did consider all of their decisions, as they themselves had not adjusted any of these settings before, why it is suspected that they do not evaluate the preset options as critical in terms of losses and gains as they have had nothing to do with them actively yet. This seems contradictory to the findings described in section 4.3 on defaults, however all of these findings were based on questions addressing the test subject directly "... what would you do?". Posing questions in that way with the types of options provided, the choice reflects more directly back onto the test subject, hence they feel more obligated towards their answer, why the default may appear more riskless as "others" has chosen that one before. For test subjects in a setup such as the plastic container factory it is easier to look at it objectively and decide otherwise than preset, for example in the lid colour setting; whether a test subject chooses a blue or

green container lid does not reflect anything of his or her person, why the default is not as important as the choice to be made.

The results therefore suggest that upon receiving a new product users will consider the preset settings and adjust them if they do not agree, whereby not obeying the default settings. On the other hand faced with a known interface such as the setup wizard users trust the default settings, which might even extend within the same category of product. For example having bought one type of television and then buying another, the user will trust these default settings.
These statements are of course extreme based on the given results, however, imagine the first time seeing the windows setup wizard; one would have to examine each step in order to understand what is the next "right" thing to do. However, over time seeing this interface one knows what to expect and a trust towards the defaults becomes established. This would explain why most test subjects in experiment DF2 did choose the default setting - they have grown accustom to this interface. This would also explain why almost half of the test subjects chose to adjust the slide bar settings in the installation setup, simply because they are not used to seeing this kind of interactive adjuster in the setup wizard context.

In terms of actively applying these results in user interfaces the implication from the present results seem to be, that given a familiar (standard) interface users will trust defaults more than when faced with a novel user interface. If this applies then companies might not be able to take advantage of users naivety if the user interface is truly novel, however, if it is constructed to appear familiar, users will accept the default settings for the most part. Therefore, to provide an example, Microsoft could potentially preset a checkbox to accept online upgrade downloads, as users already trust this interface and accepts its defaults. In this way Microsoft could choose to either help their users even more, or benefit from manipulating users to download an extra for example. Another example is that of televisions, if televisions all came with the same (logic) channel order default users would benefit from that as they would know what to expect upon buying a new television, and not having to be cognitively distressed by choosing which channel is assigned which number.

Concerning the test subjects used, the majority were students from technical faculties why they might be more prone to be interested in installation settings as well as production speed etc. than "ordinary" users. Therefore, it would be interesting to test other segments since technical inexperience might reinforce a default bias in terms of believing that the default is the recommended settings.
Regarding the current experiments, further interesting experiments concerning defaults could be made in order to examine other aspects. For example it could be interesting to distribute a device, such as an mp3-player or similar, to multiple test subjects and tell them that they were to use it, evaluate it and then keep it. On each of the devices there should be default settings, which the test subjects should be able to adjust. Over a period of time the test subject should use the device, where after he or she will be asked to evaluate the device. In the evaluation phase the test subject should be asked to explain what changes in the settings he or she has made, if any. The interesting aspect would then be to observe if test subjects who had adjusted the default settings would in fact evaluate the product better given a set of different parameters. Results from such a type of experiment would clarify if test subjects do obey defaults or not in a realistic use situation, and if defaults diminishes the potential user experience.
Also the opt-in opt-out aspect of defaults could be examined in order to see if this effect is transferable to user interfaces. Furthermore, the types of experiments put forward in this project could be more fine-tuned by using a relevant program, in which test subjects are to solve a problem, in order to obtain a price of some sort, which then will act as a motivation. Eye tracking could potentially also reveal if test subjects do in fact register the default settings, or if they are more readily focused on the different options. This would be inter-

esting as to see if defaults do relate to cognitive laziness or if they are not registered initially.

As was shown, the effect of defaults seem to express itself differently according to context. Also, it is apparent that this specific subject, which appear rather simple can in fact be tested from multiple perspectives, why the experimental coverage of the subject in this project is far from sufficient.

CHAPTER 9

Discussion and Conclusion

9.1 Discussion

The findings in relation to the thesis statement will be discussed in this section, together with application implications of the results and finally the methodological approach taken in this project.

The project concerns decision making behaviour of users in use settings. The subject of decision making is broad and includes multiple research areas, why three subjects were selected for further examination in this project; sunk time, framing and defaults. The examination of these subjects provided an overview of which decision patterns humans seem to adopt given certain circumstances. The scope of the project is:

Can observed decision patterns be observed in user interaction scenarios, thus be used actively to predict and guide user behaviour?

Why the findings from the problem analysis were sought applied in a total of two preliminary- and nine use based main experiments. The findings applied were in the form of concrete tools such as applied waiting time, frames with different reference points, different types of default settings etc. Human behaviour according to such means has been described through classical psychological research however, this research has for the most part been based on narrow framed, rigid experimental setups why there was a need to test the theories in a wider perspective; in order to be able to identify if the decision behaviour patterns observed extend beyond rigid experimental boundaries. As a consequence the experimental setups put forward in this project were all quasi experiments which for the most part sought to set up more "realistic" use settings, in which test subjects had to make decisions given the means identified in the literature. The findings from these experiments did not point to as clear conclusions as obtained in classical psychological research studies by e.g. Tversky and Kahneman, since decision behaviour was tested in more complex situations in this project. However, more or less evidence towards sunk time-, framing- and default effects was found and will be discussed in the following.

Sunk time

The sunk time effect was the first subject examined. Indications of such an effect was found by i.a. Navarro and Fantino (2008) who suggested that wasting time lead to an effect similar to sunk cost, why people will persist in what they have already invested time into. Given the means of i.a. an ultimatum game setup, this effect was tested in the present project by mediating sunk time using two product "configuration" setups. In both experiments, test subjects were to wait an amount of time for the program to setup or configure, which then imposed a somewhat wasted waiting time. Based on the first experiment ST1, where a factory machine was configured, sunk time was discovered as test subjects who were exposed to the long-wait condition demanded more of the product than those only exposed to the short-wait condition.
This result is very interesting as it underlines that waiting times during use can have an

effect on judgment and thereby decision making. The prospect of this finding could be that in user interfaces where decisions are of a critical character, it should be tested what consequences a long wait versus a short wait will have on the specific decisions to be made, whereby it could be accounted for if it is positive to have a long waiting time or a short one. If the latter applies, the findings from experiment ST2 could be applied. It was found, almost statically significant, that perception of waiting time can be increased and diminished as a consequence of entertainment and no entertainment.
Furthermore, results from experiment ST2 revealed that assumed sunk time does not seem to affect the evaluation of a user interface upon use. Test subjects exposed to the bored condition, who estimated the waiting time to be longer than test subjects in the entertained condition, rated the user interface equally to the entertained test group. For the given amount of waiting time tested in the experiment (4 minutes) no carry-over effects on evaluation of user interface was found, which could indicate that the purpose of the use in a use situation overrules other underlying effects.

Experiment ST1 did display a sunk time effect, which is believed to be rather trustworthy as test subjects did not have a real purpose for performing the tasks and demand a higher temperature limit, other than receiving a piece of cake. Nonetheless they exhibited a significantly different behaviour between conditions, solely based on the independent variable of waiting time. However, as with all experiments and especially quasi experiments nuisance variables do appear. One which especially acts as a nuisance in terms of interpreting the results for further application, is that of test subjects being "locked" in the experimental situation; test subjects were not able to do other things while waiting. They might not do so in real life either, but they have the opportunity of doing so, maybe they would check their e-mail meanwhile the installation is in progress, go make a cup of coffee, etc. If the experimental frames are peeled off, the frames for application might look totally different.
The same applies for experiment ST2, where waiting time also was a variable. Furthermore, an actual program was also tested where a task had to be performed. Again, a real purpose for the test subjects was not present, why the motivation in the experimental setting most likely does not resemble that of users, who have bought a product in order to fulfil a purpose. As a result the experiments are only able to replicate reality to a certain extent, why implications of such experiments are only guiding in relation to design for example.
However, having established that a sunk time effect can occur and change test subjects' behaviour, then it would seem reasonable to test for in user interfaces where critical decisions are to be made upon waiting times for example. It could even be that a waiting time not directly connected to the product, however still present, could affect users' decisions. For example if a new clear power plant controller has to monitor status screens, and just have to wait until something happens, then would he react differently or not as a consequence of waiting time, or will his decisions be truly found upon training received prior to the job? Even though this example might appear extreme, research upon such an area might show some of the tendencies found in the present experiments, however, hopefully not!

Framing

The second subject examined in this project was that of framing, and how we seem to be affected by subtle changes in frames provided to us in relation to a decision. Especially reference points provided for a decision, impact the decision as was exemplified by research by Hsee (1998), where ice cream cups provided physical frames for reference. Based on such findings experiment FM1 and FM2 were put forward, where a judgment or decision needed to be made by test subjects, using a slide bar. The slide bar is a graphical object which is used in multiple graphical user interfaces as well as in some physical interfaces. The length of the bars were varied between groups to see if test subjects would decide differently based on the varied frames.
None of the experiments showed an effect of framing in terms of slide bars. As before different nuisance variables did affect the results as the first experiment concerned test subjects

assessing the length of a circle diameter by using the slide bar as adjustment parameter. For this experiment, test subjects reported that they had an inner picture of the circle, why they found it relatively easy to assess the diameter. This however, was not the main problem with the experiment. Instead it can be argued that the main problem was test subjects being asked to assess a length out of context. Despite test subjects being relatively good at performing this task, the further application of the findings could be discussed as the relevance might not appear too obvious. As explained in the experiment, it was based on the idea of having test subjects pouring water into a kettle. Given such a circumstance the objective of the experiment appears more relevant, however, when translated to a graphical user interface the utility is lacking as users seldom have to assess a geometric attribute using a slide bar. For this reason it was decided to make experiment FM2.
Experiment FM2 was applied in a user interface setting, where test subjects had to decide which value to assign plastic hardness given the slide bar. Despite being tested in a user interface no effect was found, why it can be concluded that slide bars in the tested setup cannot be used for framing. In the first setup test subjects were not provided with anything else than the slide bar, whereas in the second experiment the slide bars had labels attached to them. In the first an exterior attribute had to be estimated, which in turn does not have much to do with decision making, but more of judging, whereas in the second experiment a decision had to be made. However, it appears that test subjects were in fact able to disregard the length of the slide bar and instead utilised the labels as points of reference. Furthermore, test subjects also seemed to deploy a strategy with which they could "scale" the slide bar length according to the length and points of reference, as no significant difference between bar lengths was found. It could be that we are capable of utilising a sort of heuristic where we use the geometric aspect of the slide bar to break it down for us to be able to assess the parameter based on a symmetric distribution.

Besides testing framing in relation to slide bar lengths another visual parameter, colour, was also tested as it is always used in user interfaces. Thus, it was examined if adding a colour to a button can alter our decisions, and thereby act as a frame. This was tested through three different experiments building upon each other.
In experiment FM3 test subjects were exposed to a set of odd questions, where almost all of the answer options were equally obvious. Each option button was assigned a different colour, varied between groups, in order to observe if test subjects would choose differently according to the colour attached to an answer button. The hypothesis concerning an effect of colour on test subjects' choices was not accepted as no effect was found. However, adding the blue colour to buttons was the only colour which made test subjects deselect answers. Also, a difference between significance levels among directions was found as neutral questions were not at all affected by colour, whereas positive and negative questions approximated more towards significance. No statistical significant results were obtained though, why it was concluded that colour assigned to buttons do not frame test subjects' choices systematically, as set up in the experiment. Multiple reasons as to why were suggested, such as the colours influencing each other and test subjects being aware of which colour choices they chose, which resulted in systematic choices. For these reasons it was decided to redesign the experiment to eliminate these factors from influencing the results, why a follow-up experiment was made, where colour was kept constant for each group of test subjects.
In this experiment each test subject was only exposed to answers of one colour throughout the questions, why the variable of them being aware of a colour pattern was eliminated, as well as the colours influencing each other. This experiment however, did not result in any clear results either, but in three of nine questions some significant tendencies were found, as i.a. yellow and red seemed to affect which tool test subjects chose. This may though in turn be connected to other psychological association theories where we seem to associate certain objects internally, due to i.a. experience, why it is possible that a screwdriver is more readily associated with red and yellow due to the handle colours normally used, whereas this association has not been as clear for the other tools used as answer options. The association theory is also supported by the fact that most test subjects in the no-colour condition mainly

chose the hammer, which suggest that a hammer is associated more with the concept "tools" than a saw for example. Nonetheless, two other question results exhibited tendencies where colour did seem to affect which answer options were chosen or deselected however, due to small numbers of observations for these answer options four test subjects e.g. could provide a significant result.
Experiment FM3 and FM4 did not show any clear effects of colour framing, why it can be concluded that colours cannot be used for systematical framing, or that the experimental setup was not capable of showing such an effect. Relating to the latter, the test was not set up in a use situation and resembled more that of a questionnaire. However, originally it was supposed to be coloured buttons where test subjects had to make a more "realistic" choice in a use context, but in order to test positive, negative and neutral situations emotions were applied. Furthermore, the idea was to keep the answer options as equal as possible in order for the colour effect to be clearly shown, so test subjects might as well choose one answer over another, which in turn could mediate the colour to affect the choices. As a result of these considerations, the questions became rather odd, and resembled more of a personality test, why the unclear results obtained cannot be applied to other domains. Contrarily, had the result been clear and the hypothesis accepted, they would be more readily applicable in other domains as well. As the results did not show any clear tendencies, and since the setup did not resemble that of a use situation where there is a true purpose with the interaction, a third experiment where colour framing was the object too, was put forward.
Experiment FM5 was constructed to resemble a use situation as well as diminishing the complexity of the experimental data in order to examine colours intensely. The coloured buttons were added in a use setting, rather than in a questionnaire type of setup, where test subjects had to decide on system settings provided with yes/no buttons of different colour. However, the results from this experiment did not show an effect of colour, since test subjects did not change their answers to the system settings according to colour. As with the other experiments, nuisances were present such as the questions not being neutral, test subjects being mainly technical students, not testing between colours etc.
In general for the colour framing experiments, it seems that colours cannot frame our choices, at least not systematically, despite the notion of deselecting blue colour options in experiment FM3. Of course the settings of the three reported experiments might not have tested the phenomenon thoroughly, and in order to be able to conclude more generally, more experiments should be made. For example, the colour framing could be examined further in use settings, other colours could be tested, different nuances of coulours etc. However, based on the present findings colours do not seem to interact with our choice behaviour patterns, besides tendencies observed for blue in experiment FM3, which is good news to interface designers who are concerned of aesthetics; applying certain colours to different buttons does not affect choice behaviour, why the aesthetic use of colours can be favoured.

Defaults

The default theory concerns how people tend to rely on default settings in many aspects of their lives, from subscribing to a newsletter to accepting or declining organ donation. The effect has been documented in many studies, and this subject was considered of high relevance to examine in this project as products more or less always come with default settings. Two experiments were made, where the first one (DF1) tested three different kinds of defaults; a common default, an allocation default, and a too much choice default. However, no effect of any of the defaults was found as. This could imply that we are not susceptible to defaults in user interfaces, but actually make active choices. On the other hand nuisances can have affected the setup such as it being an unfamiliar user interface for the test subjects, which might have caused them to invest more effort into making a decision. Furthermore, the premises of the setup were experimental, why test subjects might have felt obliged to adjust the settings. In fact, none of the test subjects chose the default in the lid colour setting for example, which could imply that they would not choose the default since it was

a test. Of course it might also be that the colours set by default in general were disliked!
Based on the premise of familiarity another experiment was made where defaults were tested in a familiar setting; a windows setup wizard setting. Here test subjects did display the default effect, as choices were in some instances reversed between conditions. However, the slide bar was once again utilised as a setting parameter. This time however, the lengths were not varied, but the default setting of the slide adjuster. Again test subjects' decisions when using the slide bar as an adjustment tool were not affected between conditions. For this case of defaults, it can be explained by the slide bar used as a setting adjuster in the setup wizard, which is not commonly used in such a setup, why the familiarity hypothesis posted earlier might apply. However, setting different defaults can also be perceived as framing between groups, why test subjects not deciding differently in this scenario supports the findings made in experiment FM1 and FM2.
In general defaults were found not to have an effect, but the circumstances seem to have influenced these results a great deal. For some of the defaults however, test subjects did react as predicted where the context was familiar. This points to an effect of habit and cognitive laziness as discussed by Samuelson and Zeckhauser (1988); when having made a decision which worked previously, then we are prone to be satisfied with that decision and not risk to make a new one, as we have no guarantees of it being better. According to Prospect Theory the risk of the decision not being better outweighs the chance of the decision being better, why we prefer to stay with the decision that does work for sure.
From this and the results obtained in the present project it can be concluded that we seem more prone to act according to defaults when an interface appears familiar than if it appears unfamiliar. The implications of this conclusion points out that affecting users' choices in a user interface is most easily obtained if the user interface appears familiar, however to what degree is not defined in this project. It could be that within the range of products such as mp3 players for example, users expect a certain setup, and when they encounter one containing common controls such as a play button, a volume bar etc. they perceive this interface as familiar and therefore more readily "trust" the default settings.

Application

Many of the results obtained in this report are not directly applicable as they need to be examined from more perspectives to shed a light on their limits and prospects. However, the findings still point to aspects to consider when designing a user interface. The sunk time effect was found and if a user interface contains waiting times in the form of i.a. loading, configuration, initialisation, setup etc. the choices following such waiting times should be considered, as they were found to be affected, at least in the case where a choice is made based on an assessment. On the other hand, waiting time also seems to engage the users more in the product and makes them demand more of it, which might in turn create a greater loyalty towards the product, if it meets the expectations.

In terms of framing, the slide bar lengths were not found to affect test subjects' choices as they seemed to be able to abstract from the length and instead focus on translating the geometric attributes of the bar into the decision parameter. Even though it does not seem as if user behaviour can be guided using such interaction objects, the good news is that the length of such a slide bar is of no relevance, why it can be designed to fit aesthetic design requirements instead.
For colour framing it was found that test subjects were not affected systematically. However, buttons added the colour blue were found to be de-selected significantly more times in experiment FM3, as well as certain answers were chosen more systematically in the negative and positive questions than in the neutral ones. If the blue colour effect is genuine cannot be determined finally as the setup of the experiment might have caused the effect, and it could not be supported by results from experiment FM4, as one fifth of the test subjects in that experiment had to choose a blue option each time. Nonetheless, it is worth noting that colours may play a role and providing different buttons with different colours may be

affected if blue is one of the colours; but as stated it cannot finally be concluded. In general, it can be argued that the applicability from these questionnaire setups onto a user interface is limited as user interfaces rarely contain positive and negative loaded words, but are more command based. Nonetheless, the different direction questions (pos/neg/neu) did seem to be affected by the colours attached why upon more scrutiny there could potentially be an effect. Application-wise, emotional design is more and more integrated into user interfaces in avatars, jumping icons, sounds etc., why colour used actively could be able to act as a mean for altering user behaviour if further exploration could elaborate on the effects approximated for these experiments.

Defaults appear in all user interfaces, and they do seem to have an effect in familiar user interfaces. However, the test setup might have masked an even bigger effect, or the test subjects themselves, as technical inexperienced test subjects would potentially be more inclined to choose the default settings in computer programs for example. Nonetheless, the results imply that defaults can be used to affect users' decisions since in some of the instances observed, the preferences were reversed. Therefore, upon releasing a product the designers need to be aware of which settings are default in relation to guiding or manipulating the user.

Methodology

The project is based on findings made in classical psychological research, which focuses on choices in isolation. This way of testing decision making provides clear answers, however, not founded upon a solid base from which it is applicable to other domains. The experimental setups which have provided the findings reported in the problem analysis, encompass far from all aspects which a decision is based upon. Decisions include experience, context, social influences etc. why decisions made in isolation where only two predefined choice options are available cannot be used to explain decision making processes.
This was also found in experiment STA and STB, when classic decision making problems were replicated in order to examine sunk time. By adding comment boxes to each of the questions test subjects expressed their reason for making a certain choice, which revealed more than them being "irrational". For example, test subjects who did continue with developing the operating system despite a superior competitor entering the market, explained how they would make SWOT analyses and market for specific segments etc. Without having had these comment boxes the only thing observable would be that test subjects chose to persist with the project, whereby sunk time would be found, and the choice would be labeled "irrational", despite there being a reason for the choice. The word "reason" itself emphasises how the narrow framed classical setups disregard reason and might in fact mask reason as being irrational.

In order to be able to apply findings from this project more broadly, a quasi experimental approach was used, where experimental settings were made to resemble use settings. However, it can be discussed if a user interface like the narrow framed experiments does also create narrow frames, as user interfaces often only consists of a limited set of options, whereby limiting the users' decision space. On the other hand user interfaces will always be limited in terms of choices to a certain degree, why this should not have affected the applicability of the results obtained. In fact, the way user interfaces have been tested in this project resembles how they would be used in real use settings, as test subjects have not been limited in which choices they wanted to make, however, they were only able to choose among options within the frame of the user interface.

The experimental methods used in this project sought to go beyond the narrow framed experimental model and engage the user, to test if the findings made by i.a. Tversky and Kahneman could be observed in a broader use based context. Some of the experiments put forward did however, resemble a more narrow framed experiment, such as those of colour

framing where test subjects were provided with five options, which they were to choose from. Test subjects were even told to make the choice which they could reconcile most with, if there was not an obvious answer. However, the option they would choose cannot be labelled irrational as all of the choices were similar, instead it was more of an explorative framework, where the main purpose was observation, not as much judging the outcome. Other setups such as the installation or factory machine setups sought to utilise a user interaction frame in which decisions could be measured. Again, no right or wrong answers could be given, why the object tested is more broadly applicable.

Despite trying to go beyond experimental boundaries in the present project, an experiment will always entail certain nuisances, why the experiments made in this project have been categorised as quasi experiments. As mentioned before, decision making is founded upon such individual incentives that it is more or less impossible to account for all variables included, as they differ among test subjects and contexts. Furthermore, a small sample of test subjects were used in the experiments, why it is hard to generalise as the statistical power is limited. Also, the test subjects were mainly university students who are not a representative sample for the broader user segments. Therefore, the findings made in these experiments are not directly applicable, but do mark some areas where design considerations can be of importance.

Summary

Generally seen, the research made in classical psychology has presented a set of observations concerning human decision behaviour in specific situations. Despite the (questionable) experimental frames of these, their findings have shown that based on given means we make "irrational" decisions, which can be predicted for certain domains. The means with which this "irrational" behaviour has been observed, was sought applied in a use domain in this project. Given that the decision patterns observed in narrow framed experiments also appear in use based experiments, interface designers should account for it.

Some of the observed patterns were also found in the present experiments, however not labelled "irrational" decisions, since it is assumed that reasons for acting in certain ways are present. For example, if we act according to sunk time, the underlying reason could be that we do not want to feel that we have wasted an amount of time we otherwise could have used differently. In relation to framing, we act according to what is provided to us as it would be too cumbersome to always gather most possible data in order to make a "rational" decision. The same applies for defaults; it would be cumbersome to always focus on which choices to make if one is found to work either by oneself or others, in that way we save time and can focus on other more important matters. People and users will almost always have a good reason for their actions why it is not important to categorise them as rational or not; what matters is to observe the behaviour in order to accommodate users in the best possible way. As covered in this project sunk time can alter our behaviour, as well as framing and default settings to some extent, why these parameters need to be accounted for in design for human beings.

9.2 Conclusion

The project is founded upon decision making research findings within the psychological field, where certain decision behaviour patterns have been observed. Originally it was suggested that humans always make rational decisions described by a set of axioms. This theory however was disproven through several decision making experiments, where it was found that given certain circumstances people act "irrationally". However, the experimental paradigm utilised in revealing these decision patterns does not leave much room for interpretation of actual decision making in more complex settings, such as use settings. This lead to the formulation of the initial problem:

Can observed decision patterns be observed in user interaction scenarios, thus be used actively to predict and guide user behaviour?

In terms of observing the same behaviour, as observed in classic psychological experiments, in the experiments set up in this project, some of the observations complied with previous findings. Sunk time has been observed in classic psychological experiments and was also found to be present in a use setting experiment, as test subjects displayed changed behaviour as a consequence of a time influence. The framing experiments however, only displayed few significant tendencies of colours framing test subjects answers and defaults did in some instances reverse choices.
Therefore, applying these elements into a user interface needs consideration as they can potentially affect user behaviour. In some instances such as with defaults they might in fact be utilised actively in order to aide and guide users in making a decision. The prospects of these findings are therefore substantial, especially if further research is made on the area, as decision making is a complex subject which cannot be covered completely, however be elucidated from different perspectives in order to put forward design directions to keep in mind when designing for human product interaction.

9.3 Future Works

The results obtained in this project only cover a small part of the subject of decision making, and in order to be able to apply the results more specifically further research has to be made on the area; not only within sunk time, framing and defaults, but within all aspects of decision making e.g. too much choice, compromise effects, expertise etc. However, in this section ideas for future works relating to the three examined subjects only will be presented. In general, research of a more ecological character where experiments are executed over longer time periods and in more natural use based settings could provide novel, interesting results as they could shed light on the subject from other perspectives. Potential experimental setups are specified for each area in the following.

Sunk time

Due to the aspect of time in sunk time, it would be interesting to set up experiments with durations of several hours, days or weeks, to fully test the influence of sunk time on decision making.
Sunk time was found to affect decisions where a minimum acceptance requirement had to be made based on time invested. However, if test subjects did perceive the time as wasted, or if they felt having invested time is not clear from the results. If test subjects felt that they were wasting time when installing the program, they would probably find something else to do meanwhile in a real use situation, whereas a feeling of investing time might evoke other types of behaviour and relations to the product.
It could therefore be further examined if time has to be fully wasted in order for users to experience a sunk time effect, or if investing time can act in the same way. Sunk time was only tested in terms of waiting times in this project, which can be classified as wasted time. Therefore, it would be interesting to observe users investing time in a product and see if the invested time will create a sunk time effect.
Such a test could be mediated by providing two groups of test subjects with a product such as an mp3-player or mobile phone. Both groups should be given the product and be told to use it as they would under normal circumstances. However, the first group should be given the product for one month, whereas the other group should only use the product for two days. When the testing period is over both group would be interviewed concerning the use of the product etc. Upon finishing the interview they should be told of a competing slightly better product which is really easy to use and holds all of the same features. Having been presented with the competitive product test subjects should be told that as a gift for having participated, they can choose one of the products to keep. The results of how many on each time condition chooses to switch to the new better device will reveal if a sunk time can be based on more than waiting time, but on a time investment as well. In terms of customer loyalty such a result would prove very apprehensive for products which do demand timely investments.

Another aspect to research further considering sunk time is if perceived time correlates with sunk time; if time is perceived to be longer than it actually is, will it then enhance sunk time, and vice versa? In the present project perceived time was tested in experiment ST2, but based on the assumption that the sunk time effect observed in experiment ST1 also applied in the settings of experiment ST2. Therefore, this correlation could be explored further by setting up an experiment similar to that of ST1 combined with the bored/entertained conditions, and then add the time evaluation parameter. In this way it could be found if sunk time is affected by perceived time, since test subjects in the entertained condition should not react as demandingly as test subjects exposed to the bored condition.

Further experiments could also be made based on test subjects evaluating different products upon having wasted and invested time in them, since only one program was tested in this

way in experiment ST2. The latter did not include that much product interaction, why this could be made more extensive as other judgment parameters could also be tested.
It could also be interesting to test sunk time in the scenario mentioned previously where nuclear power plant controllers wait with and without "entertainment", and then simulate a situation where critical choices would have to be made, and examine if they are affected by the sunk time or not to make different decisions.

Sunk time is an extensive area which could be examined further in relation to user interaction; however, it might require greater resources as time would need to be invested from test subjects as well as the experimenters.

Framing

Ambiguous findings in the area of visual framing were obtained from the experiments made in the project. It seems as if visual framing using slide bars do not act as a mean to frame with, as we are able to dissect the visual input and translate it into equal pieces on which we can base our decision.
For colour framing only a few tendencies were found in terms of colours being able to affect our decisions, but no directly observable patterns were found. For the most part of the colour framing experiments they were made based on questionnaire types of questions, why colour framing could as well be tested in more ecological user settings.
For example, it could be tested how we react to colour framing in real use situations. An external hard drive could be distributed to test subjects, who should then use it, as mentioned earlier. Test subjects should be divided into two groups where one of the groups' external hard drive would show a red colour on the progress bar, illustrating the amount of space used on the drive, when 20 GB's are left, whereas the others would not show the red colour until 2 GB's were left. It would then be interesting to observe if a difference between groups could be observed in terms of when they would clean up the hard drive.

Applying word based framing to a user interface could also be examined as labels are often attached to buttons for example. As mentioned in the framing recapitulation section (section 4.4), two groups of test subjects could be exposed to a buying situation in an interface where the variable varying between groups is the wording of the buy button. In one design the buy button could have the label "Invest your money" attached, whereas the other could have the label "Spent your money" attached. Would this difference of investing or spending money affect the users' decision of whether or not to buy the product? An extension of the experiment could be to test if applying different colours to the two different buttons would have an effect on the buying rate in the two groups.

In terms of the few colour tendencies found in this project, it could be tested further given another mean; the text itself. The experimental setup could be made as a different version of Experiment FM3 and FM4 where the buttons are kept in a neutral colour (e.g. white), but the questions are added different colours. In this way the question is the object of framing by applying e.g. grey or red as background colour to the question. This might frame the question to be perceived more negatively by test subjects in which case they would maybe choose a negative answer. Applying yellow or green might affect test subjects to perceive the question more positively, and therefore lead them to choose a positive answer. Each question could have answers of both negative, neutral or positive character. However, instead of different answers a slide bar could be used upon which test subjects should assess how negative or positive they perceive the question to be. In this way smaller effects of colour framing could be measured since the slide bar has many more options for the test subjects to choose between. Such an experiment would reveal how colours are perceived, as well as if colours can be used as frames (not buttons) in a graphical sense.

Framing is also an extensive subject which could be tested in multiple domains, as sound and shape could potentially also hold the ability to frame. Furthermore, testing visual framing in real use based settings could as well be interesting as external input is often sorted from experimental setups, why colour framing might not prove as effective in more complex settings.

Default

The defaults tested in the project did not convincingly show the anticipated effect, however, the experimental frames are suspected to have played a great role, why testing this phenomenon in other domains could prove that an effect exists, beyond the installation setup. Therefore, testing the different types of defaults in more ecological situations such as in actual installation scenarios would be interesting, as test subjects would then be motivated to adjust settings, since a program would be installed on their personal computer. Also, as mentioned earlier, a more versatile test group could prove to provide other results than found in the present experiment.

Another aspect of default settings which was not tested in the present project was that of defaults' influence on the perceived user experience. It could be that users who do actually adjust settings do evaluate the use of a product better than those who just use the default settings. An experiment could be set up by distributing a product which comes with default settings to a number of test subjects. The product should be purposeful and test subjects should utilise it as if it was their own property. Following a period of use, test subjects should be interviewed and asked to evaluate the use of the product, and if they have changed any of the default settings. Based on these data it would be interesting to see if test subjects who did alter the defaults settings did evaluate the user experience better than test subjects keeping default settings. It is however hard to distinguish if the potentially better ratings are mainly caused by the fact that users actively have chosen to adjust the settings, or that the settings by adjusting them fit the needs of the user better.
Examining this aspect of defaults could reveal if manufactures should aim at producing default settings which fulfil users' needs; making it unnecessary for the users to change the settings, or if actively changing the settings increases users' evaluation of the product, why the aim should be to force the user to actively adjust the settings.

A more low-fi experiment concerning finding a threshold for defaults could be to ask test subjects to take a short bike ride. The independent variable of the experiment would be the height of the seat, which would be varied corresponding to the users height in an attempt to find a limit for how low or high the seat can be before the test subject would want to adjust it. A different version of this experiment would be to keep the seat at one specific height in relation to each test subject's height and then vary the distance they are to cycle and thereby find a threshold for when test subjects in the specific setting are willing to adjust the seat, and not just accept the default. The result of the experiment however, would be rather difficult to apply in general.

Testing the default effect in an everyday use setting could also be interesting as we encounter defaults everywhere. A microwave oven could serve as an experimental object as test subjects' motivation for using it would be to get a meal. The interesting part would be to observe if test subjects would actively make settings in order to cook the food, or just go with the default settings?

The specific defaults tested in the experiments in the project could as well be examined further, as they hold different properties, which might work given other circumstances than the ones tested. As mentioned the experimental context might have caused test subjects to

feel obligated to change the settings, why an effect of the defaults was diminished. Therefore, the different types of defaults could prove to be effective in other settings.

The possibilities of applying the present decision making theories in use settings are almost infinite, since decisions are being made during interaction with almost all products. This also emphasises the importance of the knowledge research within this area can provide, and to what extent it can prove useful. The field of decision making including all of its sub subjects should be examined further given use contexts in order to be able to account for human decision behaviour in use settings.

Bibliography

Adobe [n.d.], 'Technical guides: Color models', `http://dba.med.sc.edu/price/irf/Adobe_tg/models/rgbcmy.html`. Downloaded 14.04.2011.

Allais, M. [1953], 'Le comportement de l'homme rationnel devant de le risque, critique des postulates et axioms de l'ecole americaine', *Econometrica* .

Ariely, D. [2009], *Predictably Irrationality - The hidden forces that shape our decisions*, 1. edn, HarperCollinsPublishers, London.

Arkes, H. and Blumer, C. [1985], 'The psychology of sunk cost', *Organizational Behavior and Human Decision Precesses* .

Ayton, P., Pott, A. and Elwakil, N. [2007], 'Affective forecast: Why can't people predict their emotions?', *Thinking and Reasoning* .

Bernoulli, D. [1738], 'Exposition of a new theory on the measurement of risk', *Econometrica* . Published in Econometria January 1954.

Bertini, M., Ofek, E. and Ariely, D. [2008], 'The impact of add-on features on consumer product evaluations', *Journal of consumer research* .

Blumenthal, J. A. [2009], 'Affective forecasting and capital sentencing: reducing the effect of victim impact statements.', *American Criminal Law Review* .

Boor, S. and Russo, P. [1993], 'How fluent is your interface? designing for international users', *INTERCHI 1993* .

Damasio, A. [1996], 'The somatic marker hypothesis and the possible functions of the prefrontal cortex', *Philosophical Transactions: Biological Sciences* .

Ellsberg, D. [1961], 'Risk, ambiguity, and the savage axiom', *The Quarterly Journal of Economics* .

Ganzach, Y. and Karsahi, N. [1995], 'Message framing and buying behavior: A field experiment', *Journal of Business Research* .

Gigerenzer, G. [2000], *Adaptive Thinking: Rationality in the real world*, Oxford University Press, New York.

Gigerenzer, G. [2008], *Rationality for Mortals - How People Cope With Uncertainty*, Oxford University Press, New York.

Goldstein, D. and Gigerenzer, G. [2002], 'Models of ecological reationality: The recognition heuristics', *Psychological Review* .

Hartmann, J., Angeli, A. D. and Sutcliffe, A. [2008], 'Framing the user experience: Information biases on website quality judgement', *CHI '08 Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* .

Heft, H. [1993], 'A methodological note on overestimates of reaching distance: Distinguishing between perceptual and analytical judgments', *ECOLOGICAL PSYCHOLOGY* .

Hsee, C. [1998], 'Less is better: When low-value options are valued more highly than high-value options', *Journal of Behavioral Decision Making* .

Iyengar, S. and Lepper, M. [2000], 'When choice is demotivating: Can one desire too much of a good thing?', *Journal of Personality and Social Psychology* .

Johnson, E. and Goldstein, D. [2003], 'Do defaults save lives?', *Science* .

Johnson, E. J., Bellman, S. and Lohse, G. L. [2002], 'Defaults, framing and privacy; why opting in-opting out', *Marketing Letters* .

Jowett, B. [n.d.], 'Phaedrus, the internet classics archive', `http://classics.mit.edu/Plato/phaedrus.htm`. Downloaded 14.03.2011.

Kahneman, D., Knetsch, J. L. and Thaler, R. H. [1991], 'Anomalies: The endowment effect, loss aversion and status quo bias', *The Journal of Economic Perspectives* .

Kivetz, R., Netzer, O. and Srinivasan, V. [2004], 'Alternative models for capturing the compromise effect', *Journal of Marketing Research* .

Klein, G. [2009], *Sources of power: How people make decisions*, MIT Press, Massachusetts.

Linn, M. and Petersen, A. [1985], 'Emergence and characterization of gender differences in spatial abilities: A meta-analysis', *Child Development* .

MD, R. [n.d.], 'Rational md, what is a cataract?', `http://rationalmd.com/cataract/`. Downloaded 14.04.2011.

Moeller, A. R. [2003], *Sensory Systems: Anatomy and Physiology*, 1. edn, Academic Press, London.

Navarro, A. and Fantino, E. [2008], 'The sunk-time effect: An exploration', *Journal of Behavioral Decision Making* .

Oxoby, R. and Bischak, D. [2005], 'Passing the time: Other-regarding behavior and the sunk cost of time', `http://econ.ucalgary.ca/fac-files/rjo/timedgug%20wp.pdf`. Downloaded: 12.04.2011.

Ritov, I. and Baron, J. [2002], 'Outcome knowledge, regret, and omission bias', *Organizational Behavior and Human Decision Processes* .

Samuelson, W. and Zeckhauser, R. [1988], 'Status quo bias in decision making?', *Journal of Risk and Uncertainty* .

Savage, L. [1954], *The Foundation of Statistics*, Wiley. New York.

Schwartz, B., Ward, A., Lyubomirsky, S., Monteresso, J., White, K. and Lehman, D. [2002], 'Maximizing versus satisficing: Happiness is a matter of choice', *Journal of Personality and Social Psychology* .

Shafir, E. [1993], 'Choosing versus rejecting: why some options are both better and worse than others', *Memory and Cognition* .

Simonson, I. and Tversky, A. [1992], 'Choice in context: Tradeoff contrast and extremeness aversion', *Journal of Marketing Research* .

Slovic, P. and Tversky, A. [1974], 'Who accepts savage's axiom?', *Behavioural Science* .

Soman, D. [2001], 'The mental account of sunk tomes costs: Why time is not like money', *Journal of Behavioral Decision Making* .

Starmer, C. [2000], 'Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk', *Journal of Economic Literature* .

Tetlock, P. E. [1985], 'Accountability: The neglected social context of judgment and choice', *Research in Organizational Behaviour* .

Thibodeau, P. and Boroditsky, L. [2011], 'Metaphors we think with: The role of metaphor in reasoning', *PLoS ONE* . e16782. doi:10.1371/journal.pone.0016782.

Tversky, A. [1969], 'Intransitivity of preferences', *Psychological Review* .

Tversky, A. and Kahneman, D. [1973], 'On the psychology of prediction', *Psychological Review* .

Tversky, A. and Kahneman, D. [1974], 'Judgment under uncertainty: Heuristics and biases', *Science* .

Tversky, A. and Kahneman, D. [1979], 'Prospect theory: An analysis of decision under risk', *Ecomonica* .

Tversky, A. and Kahneman, D. [1981], 'The framing of decisions and the psychology of choice, science', *Science* .

Tversky, A. and Kahneman, D. [1986], 'Rational choice and the framing of decisions', *Journal of Business* .

Tversky, A., Sattath, S. and Slovic, P. [1988], 'Contingent weighting in judgment and choice', *Psychological Review* .

Zeelenberg, M., Dijk, E. V., Bos, K. V. D. and Pieters, R. [2002], 'The inaction effect in the psychology of regret', *Journal of Personality and Social Psychology* .

Zimmerman, C. [1994], 'Damasio's proposition: A review of damasio', *Canadian Journal of Experimental Psychology* .

Appendix

APPENDIX **I**

The Too Much Choice Effect

This appendix provides two observed examples on the too much choice effect. The appendix is meant as a further elaboration on the too much choice effect, which is mentioned multiple times in the main report.

Too Much Choice Research

In the western world choices are often associated with freedom and the more choices the better. People often prefer to be able to choose for themselves, and have a lot to choose from. However, several studies have shown that the amount of choices available to the individual can have an effect on the individual´s performance and what he or she chooses.

In 2000 Iyengar and Lepper executed three experiments on this subject testing, the influence the amount of choices have on test subjects' decision making. They were interested in finding out if an extensive amount of choices had an effect on the choices we make compared to a limited amount of choices.

Their first experiment was a field experiment taking place in a supermarket. This supermarket was especially chosen since it is known for having a broad variety of groceries. The experiment was set up to look like a tasting booth of jam including two different test scenarios changing every hour. One tasting both had a variety of 6 jams and the other 24 different kinds of jam. The jams were all the same brand and a series of selection had been done prior to the experiment to remove the most popular kinds of jams due to biasing. The 6 jams in the limited selection scenario were exchanged regularly to ensure that all 24 kinds of jam from the other scenario were present in the limited selection scenario as well.

This experiment had several interesting results to show including a difference in initial attraction in the two scenarios. Of customers passing the booth with the extensive selection display of jams 60 % stopped to have a look. In addition only 40 % of the customer stopped when passing the booth with the limited selection of 6 jams. The customers were allowed to taste as many jams as they liked, but in both scenarios no customer actually tasted more than 2 different kinds of jam subordinate to the two scenarios. [Iyengar and Lepper, 2000]

In regards to whether the amount of choices had an effect on the choice of the customer to buy a jam very significant results were found. Even though significantly more people stopped at the booth with the extensive selection of jam only 3 % of these customers ended up actually buying a jam of that brand from the store. In contrast 30 % of the customers who stopped at the limited selection booth ended up buying a jam. This could imply that a broad range of choices is found more attractive initially. However, it seems that the task of choosing one out of the selection is a challenge. This led on to an observational experiment concerning how extensive selections of choices effect students´ intrinsic motivation

Students were told that they could score 2 extra credits by writing an assignment concerning a movie they had seen earlier. The students were informed that the quality of the assignment was subordinate and only one to two pages should be handed in. The students were divided in two different scenarios. In one scenario 123 students were given a list of 30 different essay

tittles to write about. In the other scenario 70 students were given a list of only 6 essay titles. These essay tittles differed from student to student making sure that all 30 essay titles from the first scenario was also used in the second scenario. Two measurements were made in this experiment; the quality of the essays and the percentage of student who chose to write the essay.

Furthermore all the essays were graded by two independent graduate students and the result showed that the essays from the students in the limited selection scenario were significantly better than the essays written on the basis of the choice between 30 essay titles. The results from this experiment thereby imply that extensive selections of choices effect our motivation.

The third experiment executed by Iyengar and Lepper (2000) was executed in a controlled laboratory environment and concerned choosing chocolates. This experiments was set up to find out if people who encounter extensive choices experience more frustration and find the decision making task more difficult than people encountering limited choices.

134 students participated in the experiment divided into three groups who tried three individual scenarios. In one scenario the participants had to choose a chocolate from an extensive selection of 30 different chocolate bars whereas in scenario two a chocolate was to be selected from a limited selection of only 6 chocolates. The last scenario was the control group and these participants were simply not provided with a choice, they were just given a specific chocolate. This control group was added to the experiment to measure if the motivation in the extensive and limited selection groups was a result of increase or decrease in motivation.

The result of this experiment showed that when asked, the participants who had encountered 30 choices reported that this was too much choice compared to the participants who encountered 6 chocolates to choose between. These participants reported that 6 chocolates was the right amount of choices. Furthermore no significant difference was found between the two groups of participants when they rated how sure they were that the chocolate they had chosen would satisfy them. The amount of choices the participants encountered had no effect on the participants' belief of the chocolates' ability to satisfy them. When asked if they believed that the chocolate they had chosen was the very best one there was still no significant difference between the answers of the two groups. It was found in the experiment that participants in the extensive choice condition found the decision making process more enjoyable than participants in the limited choice condition. However, the participants in the extensive choice condition also found the decision making process to be more difficult and frustrating than the participants in the limited choice condition.

Application

The too much choice effect manifests itself when people seem to be overwhelmed with a large selection of choices. This could potentially also apply in user interfaces, as a large number of choices would inhibit the user in making a decision. However, in order to observe this, it would need to be tested further, which is not done in this project even though it could be interesting. Therefore this appendix just provided a brief overview of how the effect manifests itself, in order to be able to understand it in the contexts mentioned in the main report.

APPENDIX **II**

Preliminary Sunk Time and -Effort Experiments

This appendix contains two experiments made within the field of sunk time and -effort. The experiments were made based on a curiosity towards sunk cost versus sunk time. Therefore, previous experiments on sunk cost and -time, which disproved sunk time, were modified in order to examine if alterations of these experimental setups could change the outcome and support the notion of sunk time.
These two experiments however, do not directly comply with the scope of the current project why the documentation is presented in this appendix to read for those interested. The setups used are based on a narrow framed imaginary paradigm. From the results it was found that applying this kind of setup does not provide a descriptive basis for human decision making behaviour, as this was revealed by combining narrow framed experimental setups with comment boxes.
It should be noted that the sunk time effect was only supported to some extent, and no conclusions towards user interface design is made based on the following two experimental results.

II.1 Experiment STA

Concerning the subject of sunk cost much research has shown that people seem to display a type of behaviour where monetary spending outweighs other aspects of a decision (see chapter 4.1). This effect has also been examined in terms of time and effort ([Navarro and Fantino, 2008], [Soman, 2001], [Oxoby and Bischak, 2005]), but with different results.

The way that sunk time and effort has been examined is through set up hypothetical questions where test subjects have to imagine themselves being in a certain situation. As argued by Soman (2001), this seems to work in sunk cost because it is possible to "measure" money objectively, and what is lost can be earned again. This does not apply for time and effort. Nonetheless, the sunk time and sunk effort experiments have been constructed in the same way as the sunk cost experiments.
The results obtained from these experiments are not in all cases convincing, even though it seems plausible that there is a sunk time and -effort effect. The reason it is believed to be plausible is due to own experiences, where time and effort has played a role. As an example the following scenario can be considered; during a project much research is made and texts are written, however, at the end of most projects the scope becomes more evident and as a consequence previous work, which does not fit the project frames any more, is discarded. The process of discarding a text one has spent hours reading for, writing on, restructuring, reading some more for etc. can be almost painful for the person who has spent his or her time on the subject, because of the time and effort invested in it.

Also, the time and effort spent on a semester project would make it harder to shift subject at the end of a term , even if guaranties were given of passing with the new project. Based on i.a. these two examples it seems likely that a sunk time and -effort effect exists, why it is believed that the setups of previous experiments in this area cause unnecessary disturbances towards revealing a sunk time and -effort effect, thus they are sought replicated with minor changes. If sunk time can be observed by minor alterations of the decision questions it will support the notion of sunk time and -effort.

Experimental objective

The experiment is made in order to examine if a sunk time and -effort effect can be observed in imaginary test setups, as this would strengthen the conception of sunk time. As sunk cost can be displayed using simple imaginary test setups sunk time is as well believed to exist why it should be possible to display it using imaginary setups as well. Sunk time might apply in use cases as well why the subject is of interest to examine further. The experiment presented is used only for the purpose of providing grounds for the sunk time effect.

Hypothesis

An overall hypothesis is put forward to set the scope of the experiment. The overall objective is to test whether a sunk time and -effort effect can be observed, since a use situation is more often connected to time and effort than cost. Therefore the following hypothesis is put forward:

Time and effort can influence choices due to sunk time and -effort

The general hypothesis is sought measured through the following experiment and "predicts" that a sunk time and -effort effect can be observed. This observation will be based on different decision questions, why actual test hypotheses are specified for each question in the following.

Apparatus

This experiment is made as a questionnaire survey based on four questions. These four questions are replicated from Arkes and Blumer (1985), Navarro and Fantino (2008), and Soman (2001) however, altered to fit a sunk time examination frame. The questionnaire format is based on the format used in the experiments from which the questions are replicated. A questionnaire survey is efficient in terms of gathering data, which fits the purpose of providing basic grounds.

Below the four decision questions chosen for use in this questionnaire are presented. An overview of the original questions are found in Attachment P together with the altered versions of the questions.

The first decision question is from Arkes and Blumer (1985) and originally concerns sunk cost. The question is a two-part hypothetical question concerning investments in an airline company. In the first part of the question, the test subjects are to imagine that they are the president of an airline company and that they are building an air plane, which is funded by $10 million. The test subjects are then to imagine that the plane is 90% finished and a competitor releases a similar, better plane; why a decision of whether they should

stop the project and save the last million or continue, must be made.
The other part of the question involves deciding if a similar project with the same competitor conditions estimated to cost 1 million dollars should be initiated.
Thereby, the amount of money and the external circumstances are the same in both scenarios, but nonetheless test subjects answer differently (None of the test subjects are exposed to both questions.)
In the altered version used in this project the airline company is replaced by a computer company, and the monetary incentives are replaced with time incentives. The airline company is replaced as it is assumed that test subjects used for the present experiment will more likely be able to imagine themselves in that context. The amounts of money is replaced by time, where in part one 4 years and 6 months have been spent on a project with 6 months remaining, whereas in the second part 6 months are in question.
Test subjects' answers should be affected by the fact that much time has already been spent in the first scenario whereas no time has been spent in the second scenario. The setup is a between subject setup, why no test subjects will be presented with two versions of the same question.

The second decision question is from Soman (2001), and concerns sunk time originally by working an amount of time, which is rewarded with tickets. Test subjects are to imagine that they have worked 5 hours for one professor and 15 hours for another. As payment they are given a rock concert ticket and a theatre ticket, respectively. Furthermore the test subjects are told that they expect to enjoy the rock concert more (worked 5 hours for). Both events are at the same night, why the test subjects have to choose which event to attend.
In this setup, test subjects chose the rock concert despite having worked longer for the other ticket. Therefore, the amount of time worked is altered for the present experiment, as it can be discussed if the time difference of 10 hours is adequate to measure a sunk time effect based on test subjects' answers. Instead of 5 and 15 hours, test subjects are to imagine having worked 5 hours for the rock concert ticket and 50 hours for the theatre ticket. By making the difference more extreme an effect of sunk time is expected. It was decided to use this question since it is easy to relate to and the direct measure is the time used.

The third decision question is from Navarro and Fantino (2008) and concerns sunk time mediated by sunk effort. The setup focuses on time and effort spent on a dig with the goal of finding copper. Two different scenarios are provided; in the first scenario 60 days have been used on the dig and it has been easy, whereas in the second scenario the 60 days have been hard. This question measures both time and effort, but it did not show any effects of effort, only time in the original setup.
Therefore, these questions are altered to make the effort difference more extreme to provoke the sunk effort effect. The adjectives "hard, difficult and a lot of effort" are replaced by "very hard, complicated and a lot of trouble and effort" in the high effort condition, whereas the adjectives "soft and easy" are replaced by "soft and effortless" in the low effort condition. The test setup is made as a between setup, why no test subject will be presented with both versions of the same question.

The fourth decision question is from Arkes and Blumer (1985), and concerns sunk cost. The setup concerns money spent on a tv dinner meal. The test subjects are asked to imagine having bought a discounted tv dinner meal. they then call a friend and asks him to come, why they need to go and buy a second tv dinner meal of the same type, however this meal is not discounted. When they return home and prepare the meals the friend calls and cancels. The question is then which of the two tv dinner meals test subjects will eat. Sunk cost was shown in this experiment, as test subjects chose the non-discounted meal in the original setup.
In this experiment the monetary aspect is replaced with time and effort, why the prices will remain the same for both meals. The difference between the two meals is then how they

are obtained; the first one is picked up on the way home, whereas the other one needs to be collected by going to the store again. In order to differentiate the latter scenario from the former, difficulties with a punctured bicycle tire, forgotten keys etc. are added. A "no preference" option was available in the original experiment why it is kept. Furthermore, the notion of a "tv dinner" is replaced by a "ready-made meal" notion, as tv dinner is considered slang, which might not be understood by non-English test subject.

All of the questions are posed in two different questionnaires, so that one test subject will not encounter the same question in two versions. Furthermore, the questions are in English in order to keep the wording as close to the original questions as possible. Keeping the questionnaire in English also provides a broader participant foundation since international students will also be able to understand and answer the questions fully, however, it was considered that some Danish student might find it more difficult to answer the questions in English. This consideration was nonetheless discarded since the participants are all students and a reasonable level of English skills is assumed.

Whenever a test subject has answered one of the questions they are asked to explain their choice, which will help to reveal if they in fact consider economic aspects of time, such as in the digging and computer company questions. If they mention that cost has been a motivation for the decision they have made, the answer will be discarded. It is not made mandatory to answer the "explain your choice" questions, since some test subjects might have answered on an instinctual level, and forcing them to explain their choice might push them to justify their "gut feeling" with economics for example.
In figure II.1, the combination of questions in the questionnaires is shown.

Questionnaire 1	Questionnaire 2
Computer company (spend 6 months additional to 4.5 years)	Computer company (Spend 6 months on a project)
Theatre tickets vs. Rock concert tickets (50 hours vs. 5 hours)	Theatre tickets vs. Rock concert tickets (50 hours vs. 5 hours)
Digging at Shady Creek (High effort)	Digging at Shady Creek (Low effort)
Ready-made meal (High effort)	Ready-made meal (High effort)

Figure II.1: *This table presents the combination of questions in the two questionnaires.*

The questionnaires are made using Google spreadsheets and are distributed by email to students at different institutes at Aalborg University. However, due to an unsatisfying level of replies each questionnaire was printed, copied and distributed by hand at Aalborg University.

Procedure

As mentioned the main part of the questionnaires were distributed by hand to students at Aalborg University. Participants were allowed to fill out the questionnaires in their group rooms, but were told not to discuss their answers (see Attachment B for the introduction).

Test hypotheses

In this section the actual test hypotheses for the four decision questions are put forward. Each of the hypotheses is accompanied by a null hypothesis, which is going to be tested. The first test hypothesis concerns the computer company decision question, where it is hypothesised that:

H.STA.1 Significantly more test subjects exhibit a preference for answering "yes" in questionnaire 1 compared to questionnaire 2

N.STA.1 Test subjects do not exhibit a significant preference for answering "yes" in questionnaire 1 compared to questionnaire 2

If test subjects answer "yes" in questionnaire 1, they do so on the basis of the 4.5 years already spent on the project, which is sunk time unless they mention economics, that is. If they answer "no" in the second questionnaire, it will show that the time spent in questionnaire 1 has had an impact, since the amount of time to spend further for both scenarios is 6 months.

The second test hypothesis concern the scenario of working a number of hours for a ticket. For this scenario it is hypothesised that:

H.STA.2 Significantly more test subjects will choose to attend event 1 than event 2

N.STA.2 Insignificantly more test subjects will choose to attend event 1 than event 2

Test subjects are to imagine that they have worked ten times more for the theatre ticket, why they should choose this ticket according to the sunk time theory, as more time has been spent on earning this ticket.

The next hypotheses are put forward on the basis of the third question concerning digging for copper at Shady creek. In both scenarios 60 days of work have been spent and 10 more days are needed. Two versions are made, one where the past 60 digging days have been hard and required a lot of effort, and one where the past 60 days have been effortless. Therefore it is hypothesised that:

H.STA.3 Significantly more test subjects will choose to continue the dig in both versions of the question

N.STA.3 Insignificantly more test subjects will choose to continue the dig in both versions of the question

H.STA.4 Significantly more test subjects will choose to continue the dig in questionnaire 1 compared to questionnaire 2

N.STA.4 Insignificantly more test subjects will choose to continue the dig in questionnaire 1 compared to questionnaire 2

In both scenarios time has been spent, why both questions should result in a preference of continuation of the dig. However, because of the one dig being harder, sunk effort is as well applied, why test subjects exposed to this scenario are anticipated to have a further incentive to continue the dig.

For the last decision question concerning the ready-made meal, sunk effort should be the main driver, since the one meal was easy to obtain whereas the other required a lot of effort. It is therefore hypothesised that:

H.STA.5 Significantly more test subjects will choose "the meal bought for your friend" compared to the amount choosing "the meal bought for you"

N.STA.5 Insignificantly more test subjects will choose "the meal bought for your friend" compared to the amount choosing "the meal bought for you"

The meal which requires a lot of effort is the one picked up for the friend. In the original decision experiment where the buyer's meal was on sale, and the friend's was not, test subjects chose to eat the friend's meal because it had been more expensive. Therefore the same findings with effort are assumed.

Participants

Before examining the results a screening was made in order to discard replies were test subjects obviously did not grasp that they were to imagine the scenarios and options as the only ones possible, why they made extra boxes etc. 100 students participated, but were reduced to 61 participants due to the reasons just mentioned. The remaining answers were used and cover 59 students and 2 university staff (mean age = 24.31±5.66 years). The gender distribution was 29 women and 32 men.
For each question further screening was made according to the answers in the "explain your choice" boxes, where replies influenced by economical thoughts were discarded.

Results

In the following each of the hypotheses are examined in terms of the results obtained. Hypothesis H.STA.1 is tested using Fisher's Exact test to test if the results deviate from the null hypothesis, since the data is count data and the number of observations is limited. The data are presented in figure II.2.

	4.5 years spent	0 years spent
Continue project	18	5
Discontinue project	5	16

Figure II.2: *The table presents the count data from the first decision making question.*

The result from the Fisher's Exact test shows that significantly more test subjects answer "yes" in the first questionnaire, where 4.5 years have been spent (p=0.0006911), why hypothesis H.STA.1 is accepted. In the second questionnaire the majority of test subjects chose "no", which shows that the sunk time effect applies. An effect of sunk time was found; when 4.5 years had already been invested test subjects wanted to continue; opposite starting fresh on a new project taking the same time as the remaining time of the first mentioned project.

The second hypothesis H.STA.2 states that significantly more test subjects will choose the theatre ticket, since they have invested more time obtaining it. The data is presented in figure II.3, and a Fisher's Exact test is applied.

Again there is a significant difference (p=0.0001703), but in favour of the rock concert ticket which is apparent from the data. According to the comments from the test subjects this was

	Theatre ticket	Rock concert ticket
Yes	12	30
No	30	12

Figure II.3: *This table presents the count data from the second decision making question.*

due to the fact that the question said "I expect to enjoy the rock concert more". Therefore test subjects seemed to follow what they would enjoy rather than what they have worked more for. However, many of the test subjects mentioned the time wasted and wanted to sell or give away their ticket (see Attachment R), implicating that the ticket still represent some form of value to them, because they have worked for it. Furthermore these statements from the test subjects indicate that they want to avoid wasting any of the tickets, which complies with Arkes and Blumers' theory of us wanting to avoid wasting time, thus sunk time seems to exist. The conclusion however, is that hypothesis H.STA.2 cannot be accepted.

The third and fourth hypotheses concern whether test subjects would continue to dig at Shady Creek, and if more test subjects would choose to do so when the effort has been higher. Hypothesis H.STA.3 is accepted, since the majority in both scenarios chose to continue the dig (52% and 84%). However, it is not convincing for the "high effort" group, why this result is considered somewhat erratic.

	High effort	Low effort
Continue the dig	13	21
Abandon the dig	12	4

Figure II.4: *This table presents the count data from the third decision making question.*

To test hypothesis H.STA.4 a Fisher's Exact test is applied on the data in figure II.4. The results show that there is a significant difference between the two question versions in terms of who wants to continue the dig (p=0.03216). However, it is in a reversed fashion than predicted, since test subjects who imagined not to have worked hard in the 60 days are more prone to continue than test subjects who worked hard the 60 days. This might be caused by the fact that the latter group thought the next ten days of digging would be easy as well, even though the text says the same for both groups in relation to the next ten digging days. The sunk effort is therefore not apparent in this case and hypothesis H.STA.4 is rejected, however, sunk time seems to play a role since the majority of all test subjects wanted to continue the dig the extra 10 days.

The last hypothesis concerns the ready-made meal decision question, where it is hypothesised that an effect similar to the sunk time effect observed will appear with sunk effort. In figure II.5 the data obtained is shown, and a Fisher's Exact test is run on the "own meal" and "friends' meal" columns.

	Own meal	Friend's meal	No preference
Yes	6	3	49
No	52	55	9

Figure II.5: *This table presents the count data from the fourth decision making question.*

The "no preference" answers were removed since they are of no interest to the hypothesis. The results however, showed no significant difference between choosing the first or the second bought meal (p=0.49). However, from the core results it is easy to see that the vast majority did not care which meal to eat, as many of them commented "it is the same food". One of

the test subjects who wanted to eat his own meal stated that "he would rather throw the other meal away since it is filled with bad luck" (see Attachment R). This comment shows that he is aware of the effort put into getting it, and that there is a "difference" between the meals in that perspective. Test subjects who chose the friends' meal mentioned having used more energy or effort on getting this meal. This shows that some of the test subjects were in fact affected by the sunk effort, however a minority. Hypothesis H.STA.5 is therefore rejected, and the null hypothesis is accepted.

Discussion

Some of the results implied that a sunk time and to some extent a sunk effort effect does exist. However, only the first decision question showed the effect with statistical significance. As mentioned an extensive part of the gathered data was discarded due to test subjects expressing an economic justification of their choice, which undermines the sunk time effect. In the original studies from where the questions are obtained an "explain your choice" box is not reported. It is believed that due to this addition the data becomes more valid, since test subjects who explained themselves by "having wasted time or effort" in fact did show that a sunk time effect can be found.

On the other hand, those answers discarded, showed a tendency to evaluate time in terms of cost, which is not insignificant. However, all of the decision questions posed in this survey are set up as imaginary questions, which means they lack authenticity in the sense that test subjects are in fact in the situation. If test subjects had worked 50 hours for one ticket, but only 5 hours for another, and assuming they would enjoy the least worked for performance, it is doubted that it would be as easy to deny the performance worked for the longest. Also, in the present decision question a rock concert versus a theatre performance was mentioned. Since the majority of participants in this questionnaire were students, the rock concert might have seemed more appealing to them in general, why this is considered a nuisance. It is therefore decided to rewrite the question to not focus on what type of event is earned for the work hours (see Attachment Q).

For the first decision question concerning the computer company many of the test subjects reported on other areas where they would develop the operating system further. All in all they put a lot of effort into answering the question and give reasons for their choices on a detailed level, disregarding that the text said that the other company's system would be better. Therefore, it is decided to emphasise that there is no way of optimising the system, and that test subjects need to consider the facts in the text and not relate it to something outside the described, why the wording is changed to emphasise the message within the question (see next section).

In the second decision question the hypothesis was rejected, however, several of the subjects commented that they would not throw away the ticket they did not use, even though this was the scenario they were presented with. They commented that they would go to the rock concert, but then give the theatre ticket to a friend to avoid wasting it. This statement indicates the existence of sunk time since the test subjects revealed an eager to avoid wasting the time they spent on the theatre ticket, by giving it to a friend.

For the third decision concerning the dig at Shady Creek some of the test subjects noted that they would continue, since the alternative was to go home instead of searching for another place to dig, which they thought made it a strange choice (see Attachment R). Therefore it was decided to alter one of the choice options in this decision question from *Abandon the "Shady Creek" mine* to *Give up the dig at "Shady Creek" and search for another place to dig.* The change may make it more realistic to the test subjects imagining that they are a leader

of a copper mining group, but in terms of the answers it should not change the choices made.

The fourth decision question concerning the ready-made meal seemed to be too exaggerated in the high effort story. One of the test subjects even mentioned "I would find it easier to get rid of the one that I just bought because it's filled with bad luck", referring to the meal bought for the friend. Making this option very negative might have affected the test subjects' choice why the high effort scenario is toned down.

Based on the observations made and the comments made by test subjects it is decided to reformulate some of the questions and distribute them to other students, and see if other results are obtained and other explanations given, which will be described in the following, together with a thorough discussion of the results and the questionnaire decision questions.

II.2 Experiment STB

This experiment is a follow-up experiment for experiment STA. This experiment is used to reduce nuisances discovered in the previous experiment. They are corrected in order to accommodate for possible misleading answers found in the first part of the experiment.

Experimental objective

The objective of this experiment is to account for the nuisances discovered in the previous experiment, in order to be able to observe a sunk time and effort effect, as it was only supported in half of the experiment as well as test subjects' comments indicated nuisances.

Hypothesis

The overall hypothesis remains the same for this experiment as it resembles the previous with only a few parameters changed, why the same hypothesis is put forward:

Time and effort can influence choices due to sunk time and -effort

The general hypothesis is sought measured through the following experiment and "predicts" that a sunk time and -effort effect can be observed. This observation will be based on different decision questions, why actual test hypotheses are specified for each question in the following.

Apparatus

The four scenarios with modifications can be found in Attachment Q.

The first decision question is changed in order to emphasise that no other changes or improvements can be done to the computer system, since several of the test subjects chose to continue with the research project, but with the argument that they would just develop the system to be better than the other company's. The "rational" choice in this scenario given the options provided should be to discontinue the project. Since test subjects in Experiment STA are of the belief that they can improve the system it shows that they have not grasped the full idea, why it is attempted to emphasise this, by pointing out that the competitive company's system is much better than the one the test subjects' company can build.

The second decision question had two rewards in terms of tickets for a theatre performance and a ticket for a rock concert. This can be considered a nuisance since students in general seem to be more attracted to rock concerts than theatre performances, which in turn can affect the results. Therefore the reward for working is replaced by an event 1 and an event 2. In this way the test subjects' preference for one of the events should be evened out, hence the focus will be on the hours they spent on earning each ticket.
By not focusing on the type of event, this should not be able to interfere with the choices made. As formulated now it is either time or the mere fact that test subjects expect to like event 2 more that are in play.

For **the third** decision question it is only one of the choice options which is changed, in order to make it more realistic, since a copper-mining group most likely would just not abandon a mine in order to go home. The change is from:

Abandon the "Shady Creek" mine
to
Give up the dig at "Shady Creek" and search for another place to dig

This is done since it was obvious from the test subjects' comments that they did not choose to give up, since the alternative option included just going home.

The fourth decision question is modified so the effort put into getting the second meal is not as extreme and negative, but just annoying. Furthermore, this question is placed at the beginning of the questionnaire so it is not as affected by the other questions, since test subjects in general commented more negatively on this question. Of course it can be caused by the nature of the question, but it is considered that there might be an effect as this is the last question in a row of peculiar questions.

All of the questions are once again posed in two different questionnaires, in order for one test subject not to encounter the same question with moderations twice. The "explain your choice" boxes are also utilised in this experiment. In figure II.6, the combination of questions in the questionnaires is shown.

Questionnaire 1	Questionnaire 2
Ready-made meal (Reduced effort)	Ready-made meal (Reduced effort)
Computer company (spend 6 months additional to 4.5 years)	Computer company (Spend 6 months on a project)
Event 1 vs. Event 2 (50 hours vs. 5 hours)	Event 1 vs. Event 2 (50 hours vs. 5 hours)
Digging at Shady Creek (High effort)	Digging at Shady Creek (Low effort)

Figure II.6: *Questionnaire overview.*

These questionnaires are only distributed in hard copies to students at different institutes at Aalborg University, who have not already answered the former questionnaire.

Procedure

The questionnaires were distributed by hand to students at Aalborg University. Participants were allowed to fill out the questionnaires in their group rooms, but were told not to discuss their answers, see Attachment B for the introduction given to the test subjects.

Test hypotheses

In this section the test hypotheses for the four decision questions are put forward. The hypotheses are the same as tested in experiment STA, since the object of both experiments is the same. The hypotheses are presented in order to distinguish these from the previous if another accept/reject pattern occurs.

H.STB.1 Significantly more test subjects exhibit a preference for answering "yes" in questionnaire 1 compared to questionnaire 2

N.STB.1 Test subjects do not exhibit a significant preference for answering "yes" in questionnaire 1 compared to questionnaire 2

As the question is stating more clearly that the project cannot be developed to be better than the competing product, more test subjects are expected to answer "yes" in the first questionnaire, as they will be more affected by sunk time, whereas more will answer "no" in the second questionnaire.
The second test hypothesis concerns the scenario of working a number of hours for a ticket. For this scenario it is hypothesised that:

H.STB.2 Significantly more test subjects will choose to attend event 1 than event 2

N.STB.2 Insignificantly more test subjects will choose to attend event 1 than event 2

This time test subjects are not able to choose based on type of event, only on the mere fact of time and what event they expect to like the most.
The next hypotheses are put forward on the basis of the third question concerning digging for copper at Shady creek:

H.STB.3 Significantly more test subjects will choose to continue the dig in both versions of the question

N.STB.3 Insignificantly more test subjects will choose to continue the dig in both versions of the question

H.STB.4 Significantly more test subjects will choose to continue the dig in questionnaire 1 compared to questionnaire 2

N.STB.4 Insignificantly more test subjects will choose to continue the dig in questionnaire 1 compared to questionnaire 2

In both scenarios there is sunk time, why both questions should result in a preference of continuation of the dig. However, because of the added notion of "abandoning the Shady Creek and look for another place to dig" a bigger difference in answers is expected; test subjects in the high effort condition should still continue to choose to continue the dig, whereas the low-effort group is expected to be more inclined to choose the other option.

For the last decision question concerning the ready-made meal, the effort notion is toned down, in order not to bias test subjects negatively. The hypothesis states that:

H.STB.5 Significantly more test subjects will choose "the meal bought for your friend" compared to the amount choosing "the meal bought for you"

N.STB.5 Insignificantly more test subjects will choose "the meal bought for your friend" compared to the amount choosing "the meal bought for you"

Participants

Before examining the results a screening was made in order to discard replies were test subjects obviously did not grasp that they were to imagine the scenarios and options as the only ones possible, why they made extra boxes etc. 100 students participated in total, but for the reasons mentioned only 52 of the replies were usable. For the 52 students answered (mean age = 23.35±4.24 years), the gender distribution was 12 women and 40 men.
For each question further screening was made according to the answers in the "explain your choice" boxes, where replies influenced by economical thoughts were discarded.

Results

In the following each of the decision question results are examined in the same way as explained in Experiment STA Results and is compared to the hypotheses.

Hypothesis H.STB.1 is tested using a Fisher's Exact test to test if the results deviate from the null hypothesis, since the data is count data and a limited number of observations have been made. The data for question 1 is presented in figure II.7.

	4.5 years spent	0 years spent
Continue project	10	4
Discontinue project	10	18

Figure II.7: *The table presents the count data from the first decision making question.*

Hypothesis H.STB.1 is accepted and the null hypothesis is rejected, since significantly more test subjects chose to continue the project in the first part than in the second part of the question (p=0.04878). Also, as in the previous experiment, more test subjects did not continue in the second part, which supports the notion of a framing effect. The results for this setup display the effect of the changed wording in relation to the operating system, which make the test subjects understand the question better. This is observed by them not commenting as extensively on the scenario.

The second hypothesis states that significantly more test subjects will choose to attend event 1, since they have invested more time in it. The data is presented in figure II.8, and a Fisher's Exact test is once again applied.

	Event 1	Event 2
Yes	5	42
No	42	5

Figure II.8: *The table presents the count data from the second decision making question.*

Again there is a significant difference (p=2.934e-15), but still in favour of the second event. This time no differences between the events were announced, and judging by most comments made by the test subjects they would choose event 2 because "I expect to like event 2 the most" (see Attachment R). This time no personal preference could interfere with the results, why in this case no sunk time was found. However, many of the test subjects further commented that "I want to go to event 2 the most. And I might sell ticket 1, which must be worth more". By comments like these, it shows that test subjects still ascribe a high value to the ticket they worked for the most, but nonetheless most made the "rational" choice of attending the event they would expect to enjoy more.

The next hypotheses concerned whether test subjects would continue to dig at Shady Creek, and if more test subjects would choose to do so when the effort has been higher. Hypothesis H.STB.3 is accepted, since the majority in both scenarios chose to continue the dig (64% and 57%). However, the small numbers of test subjects make the results fairly volatile (see figure II.9).

	High effort	Low effort
Continue the dig	9	8
Abandon the dig	5	6

Figure II.9: *The table presents the count data from the third decision making question.*

To test hypothesis H.STB.4 a Fisher's Exact test is applied on the data in figure II.8.

	Own meal	Friend's meal	No preference
Yes	5	1	29
No	30	34	6

Figure II.10: *The table presents the count data from the fourth decision making question.*

Based on the small numbers the p-value approximates 1, which indicates no difference ($p \approx 1$). Hypothesis H.STB.4 is therefore rejected. Making the choice options more described and to some extent realistic has made more test subjects prepared to abandon the dig.

The last hypothesis concerns the ready-made meal decision question, where it is hypothesised that an effect similar to the sunk time effect observed will appear with sunk effort. In figure II.10 the data obtained is shown and a Fisher's Exact test is run on the "own meal" and "friend's meal" columns.
The "no preference" answers were removed since they are of no interest to the hypothesis. The results however, showed no significant difference between choosing the first or the second bought meal (p=0.1981) why hypothesis H.STB.5 is rejected. As in the first questionnaire, test subjects mainly did not care what meal to eat. However, a misconception among test subjects seem to have arisen, since many test subjects commented that the second bought meal was more fresh why they would eat that. Replies of this kind were of course discarded. This tendency seems to be an effect of placing this question as the first question in the questionnaire. It might seem strange to test subjects that they are to answer a questionnaire of this kind, why they try to make sense out of the question by "adding" to the story. In the previous questionnaire, it was posed as the last question, why test subjects at this point had most likely found the questions to be odd, why they were not trying as hard to make sense out of it.

Discussion

In general the results did not prove statistically significant that a sunk time and -effort effect exists. Only one of the decision questions gave a significant result benefiting sunk time in both versions - the computer company question. However, due to the "explain your choice" box additional data was registered, some of which is redundant, but some of it indicated the differences between how people make decisions. It seems that the majority is more "rational" when it comes to time and effort, whereas others act as in sunk cost examples. However, it is important to keep in mind that all of these questions are imaginary questions, why the results also might reflect who is good at imagining and who is bad at it.

Another aspect of the results from this experiment compared to sunk cost experiments is that people are used to measure money, and that they are easy to count and grasp. Time on the other hand is more ambiguous, for example; using 50 hours to earn a ticket might sound as much, but what would the alternative be? That the person would watch TV the 50 hours? make homework? etc. Earning 50 kr. on the other hand is easier to quantify, since there is certain things one can buy for 50 kr. Therefore, transforming these sunk cost questions into sunk time might not be all fair, as they are based on imagination.

In terms of the questions posed where there is time ahead to finish a project, e.g. 6 months for the computer project and 10 digging days, test subjects tended to want to continue when time had already been invested. In this case it can be discussed if the test subjects continue the projects because of the time already spent or simply because the remaining period of time is short. In order to determine if the forthcoming time is short one would need a frame of reference to decide from, why the period already used is an obvious reference frame. The same argument can be used when it comes to sunk cost. It is therefore considered sunk time when test subjects decide to continue a project, since the prospective time is sort in relation to the already used time. This was especially seen in the computer company question, where the framing reversed the preferences. Six months might have been considered much time, when six months was the only period discussed, whereas it might have been considered short in relation to the 4.5 years mentioned.

So the question remains somewhat open; if sunk time and sunk effort do exist. The results from the present study showed that sunk time to some extent can be found, whereas sunk effort was not shown under any of the present circumstances, almost the opposite was found in the digging question. So does this mean that no sunk time and effort exist? One may argue that time and money are always connected, but money and time are not, why it is difficult to discern the two in order to focus only on time.
However, the key problem with identifying sunk time and -effort by these imaginary setups might also be caused by the objectives discussed in chapter 5 on experimental paradigms. The results of these setups rely on the test subjects' ability to imagine themselves in a rigid situation where only two choice options are present. However, as opposed to just presenting test subjects with a question and a set of answer options, they were also presented with a comment box in the present experiments. This has extended the data and provided a broader basis for understanding test subjects' choices in the given situations. And it clearly revealed how many of the test subjects did imagine themselves in the scenario, but came up with far more complex solutions to them, than just "yes" and "no". This type of artificial setups made the test subjects react by making up other answer options as they were not capable of choosing between the constructed choices. Of course it can be perceived as an annoying factor for some researchers, but nonetheless it reveals a great deal of this type of setups; they do not encompass authentic human behaviour. For this reason the sunk time and -effort effect may have been even harder to observe as time and effort is not as easily imagined, since time and effort can take on multiple different "appearances" where time can be perceived as fun or boring, fast or slow, and effort can be high, low and the perception

of it might depend on the goal to be obtained. Contrarily, money is accountable and might more easily be tested in a rigid setup such as those tested in the present experiment. Therefore, it can be argued that in order to observe a sunk time and -effort effect, the test setups need to be more authentic and less restrained, as human behaviour is versatile and it seems that it cannot be generalised from experiments such as the present one.

APPENDIX **III**

Colour

Using colours in a graphic user interface involves considerations towards their meaning as well as the use on a screen.

Colour Perception

On a general level the eye receives a light input, which it sends towards the brain. The information is encoded into chains of electronic impulses (neural activity). The pattern of these impulses are then decoded in the brain and the seen object is represented in the Visual Cortex.
Light is essential for human vision and further perception. Light is part of the electromagnetic spectrum, which also contains i.a. x-rays, micro waves, infrared radiation etc. However, it is only a small part of the electromagnetic spectrum which we are able to perceive, as the eye's receptors are only susceptible to wavelengths between 400 nm and 700 nm. In this interval humans are able to perceive colour.
Physiologically, the eye is constructed as seen in figure III.1. The light first hits the Cornea, then it continues through the hole in the Iris, where after it is bent by the Lens. Finally, the image from the visual field is projected onto the Retina. The Retina consists of two types of brain-like cells called photo receptors, which again consist of three different types of rods and three different types of cones. The function of the rods is to register light, whereas the role of the cones is to register colour, which demands a certain amount of light.

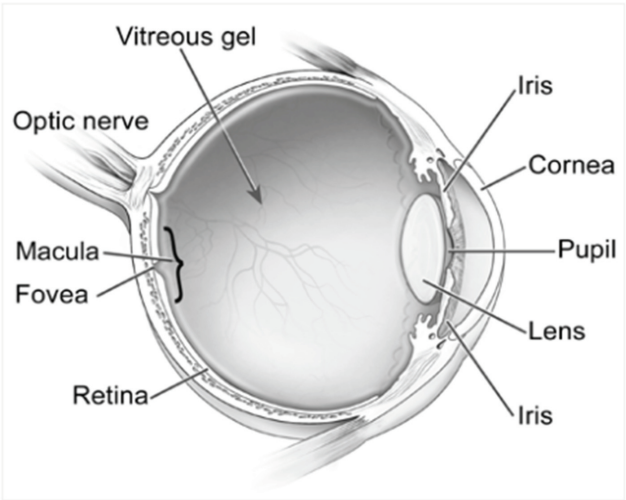


Figure III.1: *Eye model in cross section [MD, n.d.].*

From the visual sensory perception in the eye, the information from the light waves are transported further in the system via the optic nerve and through to the brain. The visual sensory perception is then processed in the Primary Visual Cortex.[Moeller, 2003]

Display Colour Format

The RGB-model is used for the colour resolution on computer screens. The RGB stands for Red, Green, Blue which are the three colours whose wave lengths are combined to create other colours. The use of RGB colours rely on an additive model, where adding two colours provide a third colour which is lighter than the two original colours. White is obtained by mixing all three colours (see figure III.2).
It is not by coincidence that the model is based on red, green and blue, since the retina contains the light sensitive receptors, where especially cones contain photo pigments which are sensitive to exactly the wavelengths humans interpret as red, green and blue. [Adobe, n.d.]

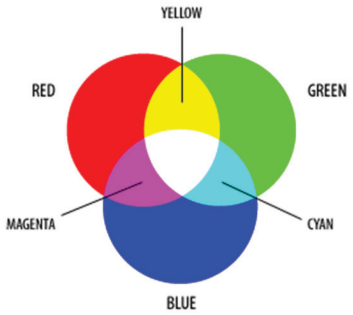


Figure III.2: *Additive colour model [Adobe, n.d.].*

Colour Interpretation

Besides from the wavelengths of the light, the perception and understanding of the colours are even more important for the current project, as the different colours are ascribed certain values. In figure III.3 a cut-out of the colour culture chart made by [Boor and Russo, 1993] is shown, to reflect what is associated with the different colours in the Western world.

Country	Red	Blue	Green	Yellow	White
United States	Danger	Masculine	Safety	Cowardice	Purity

Figure III.3: *Reduced colour culture chart [Boor and Russo, 1993].*

These values assigned might prove to be an influencing factor in the colour framing experiments, as all of these colours are used.

APPENDIX **IV**

Website Colouring

In order to decide which colours to utilise in experiment FM3 (section 7.3), it was decided to visit 20 randomly chosen websites, and count the colours used actively (meaning that colours used in e.g. pictures are not counted). In this way the colours more often used are obtained, and can be examined in the experiment.

The websites are randomly chosen, by writing a letter in the Google search engine, and picking the first subject it suggests and the first of these websites. Below the websites are listed, together with the colours observed on the site. Commercials, pictures, logos and movie clips were disregarded, why only background colour and frames are counted.

1. http://www.aalborgkommune.dk/Sider/Forside.aspx
2. http://www.bt.dk/
3. http://cdon.dk/
4. http://www.dmi.dk/dmi/index/
5. http://www.elgiganten.dk/
6. http://www.facebook.com/
7. http://www.guloggratis.dk/
8. http://www.hotmail.com/
9. http://www.ikea.com/dk/
10. https://job.jobnet.dk/CV/frontpage.aspx
11. http://www.krak.dk/
12. http://www.lego.com/da-dk/
13. http://dk.msn.com/
14. http://www.nordjyllandstrafikselskab.dk/default.aspx
15. http://www.dr.dk/b/Oline/forside.htm
16. http://www.dr.dk/p3/
17. http://www.qxl.dk/
18. http://www.rejseplanen.dk/
19. http://www.skat.dk/
20. http://tv2.dk/

Colour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Purple																					1
Red																					5
Black																					5
Orange																					6
Green																					7
Yellow																					8
Grey																					8
Blue																					14
White	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	19

Figure IV.1: *Overview presenting the results of colour counting on 20 websites.*

The five colours for the experiment was selected to be the five colours used most frequently on the 20 websites, since it can be argued that there is a bigger need to examine colours more frequently used as opposed to colours less frequently used. However, the colour white was most often used (19 out of the 20 cases), but mostly as a background colour. It was therefore decided to disregard white, as the other colours are specifically applied in areas of the screen used for button and frames for instance.

Attachments

Program for Experiment ST1

For experiment ST1 the configuration times were varied between groups. In condition 1 test subjects wait 10 seconds while configuring, whereas test subjects in condition 2 waits 2.5 minutes, see figure A.1.

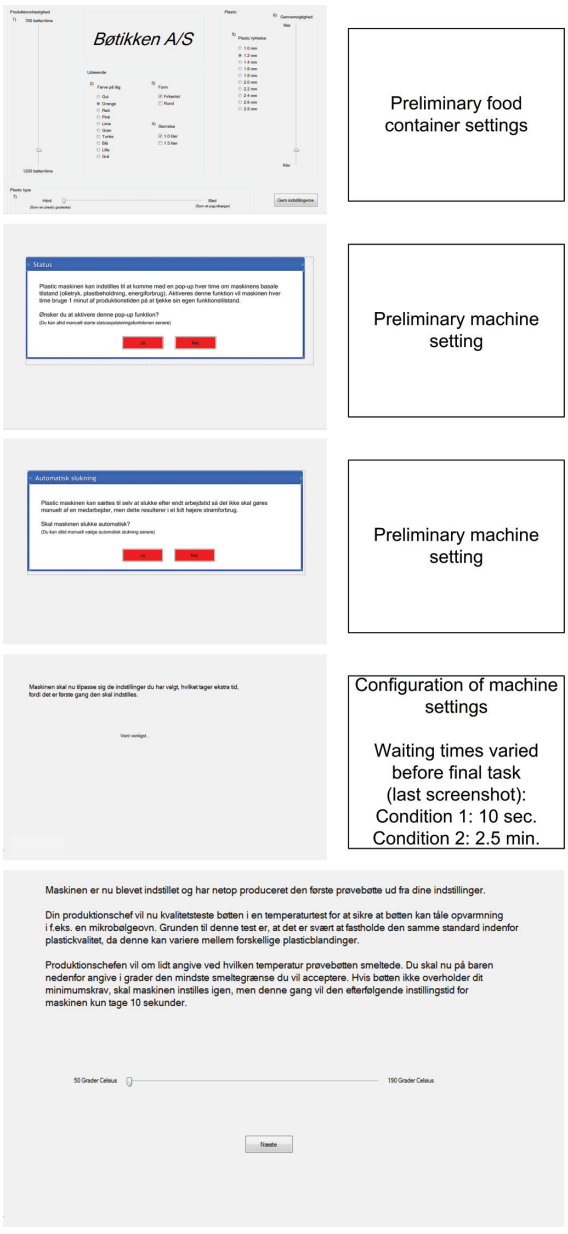


Figure A.1: *ST1 Experiment program, the two screens at the bottom are of particular importance for this experiment.*

ATTACHMENT **B**

Introduction to the Experiments

In the following all of the introduction manuscripts used by the experimenter are presented.

Experiment STA and STB

Hej, og mange tak fordi I vil udfylde dette spørgeskema.
Når I skal svare på spørgsmålene vil vi gerne bede jer om, at have in mente at I skal forestille jer de forskellige situationer der opstilles, og hvad I ville gøre i den givne situation. Da spørgeskemaerne henvender sig til jer enkeltvis, beder vi jer om ikke at diskutere jeres svar indbyrdes før I har afleveret spørgeskemaerne tilbage til os.

Experiment ST2

Hej, og mange tak fordi du vil deltage i dette forsøg.
Først vil jeg gerne bede dig om at sætte din mobiltelefon på lydløs og lægge den i kassen her sammen med dit ur.

Opgaven du nu skal i gang med består i, at du skal vurdere et nyt program, og for at vi kan få dig til at vurdere det som en helhed, skal du også installere programmet og løse en lille opgave i det. For at du kan vurdere programmet som en helhed, beder vi dig også installere det, da det skal med i din endelige vurdering.
Programmet er klar til at blive installeret på computeren foran dig, og du skal bare følge installationen skridt for skridt, som du ellers ville gøre det privat. Når installationen er færdig skal du udføre opgaven som ligger på bordet ved siden af computeren. Når du har udført den, bedes du markere, hvorefter forsøgslederen vil komme og lukke programmet ned, hvorefter du bedes besvare nogle få spørgsmål, og så er du færdig.

Undervejs vil vi gerne bede dig om ikke at forlade det aktuelle program på computeren. Når du er færdig, må du gerne forlade lokalet.

Experiment ST1, FM2, FM5 and DF1

Hej, og mange tak fordi du vil deltage i dette forsøg.
Først skal jeg lige forsikre mig, at du ikke er farveblind. Dernæst vil jeg bede dig om at sætte din mobiltelefon på lydløs og lægge den i kassen her sammen med dit ur.

Du skal nu teste programmet på computeren. Programmet er lavet sådan, at det selv vil følge dig igennem nogle forskellige trin. Du ved du er færdig med testen, når computeren beder dig om demografisk data som alder og køn, hvorefter du skal trykke på “Afslut” knappen.
Undervejs vil vi gerne bede dig om ikke at forlade det aktuelle program på computeren. Når du er færdig, må du gerne forlade lokalet.

Experiment FM1 and FM3

Hej, og mange tak fordi du vil deltage i dette forsøg.
Først skal jeg lige forsikre mig, at du ikke er farveblind. Dernæst vil jeg bede dig om at sætte din mobiltelefon på lydløs.

Du skal nu svare på en række spørgsmål i programmet på computeren foran dig. Hvis ikke du synes en af svarmulighederne passer helt præcist, så vælg den der kommer tættest på.
Efterfølgende spørgsmålene kommer der en anden lille opgave, hvorefter forsøget er færdigt.

Du bedes om ikke at forlade programmet med spørgsmålene under forsøget.

Experiment FM4

Hej, og mange tak fordi I vil udfylde dette spørgeskema.
Først skal jeg lige forsikre mig, at der ikke er nogen af jer der er farveblinde?

I skal nu svare på en række spørgsmål som kan forekomme lidt sære, men hvor I hver især bedes give et svar per spørgsmål. Hvis ikke I synes en af svarmulighederne passer helt præcist, så vælg den der kommer tættest på. Imens I udfylder spørgsmålene, vil vi gerne frabede os at I taler sammen, men I er selvfølgelig velkomne til at diskutere spørgsmålene efterfølgende.

Experiment DF2

Hej, og mange tak fordi du vil deltage i dette forsøg.

Inden vi sætter forsøget i gang, vil jeg gerne bede dig om at sætte din mobiltelefon på lydløs og lægge den i kassen her, sammen med dit ur.
Derudover vil jeg også gøre opmærksom på, at du undervejs i forsøget ikke må forlade programmet du er i gang med at teste.

Opgaven består i at du skal vurdere et nyt program, og for at vi kan få dig til at vurdere det som en helhed skal du også installere programmet, da installeringen foregår lidt anderledes og skal med i din endelige vurdering.
Programmet er sat op, så du skal egentlig bare følge programmet skridt for skridt. Når du kommer til et skærmbillede hvor du bedes om alder, køn m.m. er du færdig, når du har udfyldt disse data og trykket på knappen “Afslut”.

Har du nogen spørgsmål? Ellers må du gerne gå i gang.

Introduction sheet for Experiment ST1, FM2, FM5 and DF1

Hej

Og tak fordi du vil deltage i vores eksperiment!

Du er direktøren i firmaet Bøtikken A/S, som producerer plasticbøtter i alle former og størrelser til opbevaring af fødevarer. Du har netop købt en ny produktionsmaskine, der kan producere runde og firkantede bøtter i mange forskellige farver og plasttyper. Den skal nu indstilles for første gang så en produktion af plasticbøtter kan sættes i gang. Du kan indstille maskinen på skærmen her fra dit kontor.

Bøtikken A/S sælger og producerer plasticbøtter af alle kombinationerne muligt på maskinen, da maskinen er blevet speciallavet til formålet.
Din opgave går nu ud på at indstille maskinen til at producere den første produktion af plasticbøtter, og du bestemmer som direktøren hvilken type plasticbøtter maskinen skal starte med. Det er vigtigt, at du indstiller de forskellige parametre i den rigtige rækkefølge, som er angivet med tal 1) til 7).

Når du har indstillet maskinen helt (når du har trykket "afslut") bedes du forlade lokalet.

Declaration of Consent

Samtykkeerklæring

Produkt- og designpsykologi Gruppe 1075

Dato: __. maj 2011

Jeg bekræfter hermed som forsøgsperson at:

- Video- og taleoptagelser samt observationer må blive brugt i skriftligt materiale i forbindelse med projektet.
- Jeg giver tilladelse til at video- og taleoptagelser samt observationer fra forsøget må videregives til vejleder og censorer i forbindelse med projekteksamen og anvendes hertil. Ydermere at denne information må videregives til tredjepart efter skriftlig tilladelse fra projektgruppen.

Dit navn vil blive hemmeligholdt overfor både offentligheden og Aalborg Universitet.

Forsøgspersonens navn: _____

Underskrift: _____

Forsøgslederens navn: _____

Underskrift: _____

Tak for din deltagelse!

Program for Experiment ST2

In general the program consists of: four installation setup wizard screens, the installation, the Photo! Editor program and then an evaluation part. However, two programs are made since two conditions are tested. In the first program test subjects have to wait four minutes for the installation to complete, while only looking at the progress bar. In the second program test subjects are presented with a tutorial video during the installation period. In the figure below the progress of the program is shown. Test subjects are exposed to all of the steps, however, in figure E.2 the difference between the conditions is illustrated.

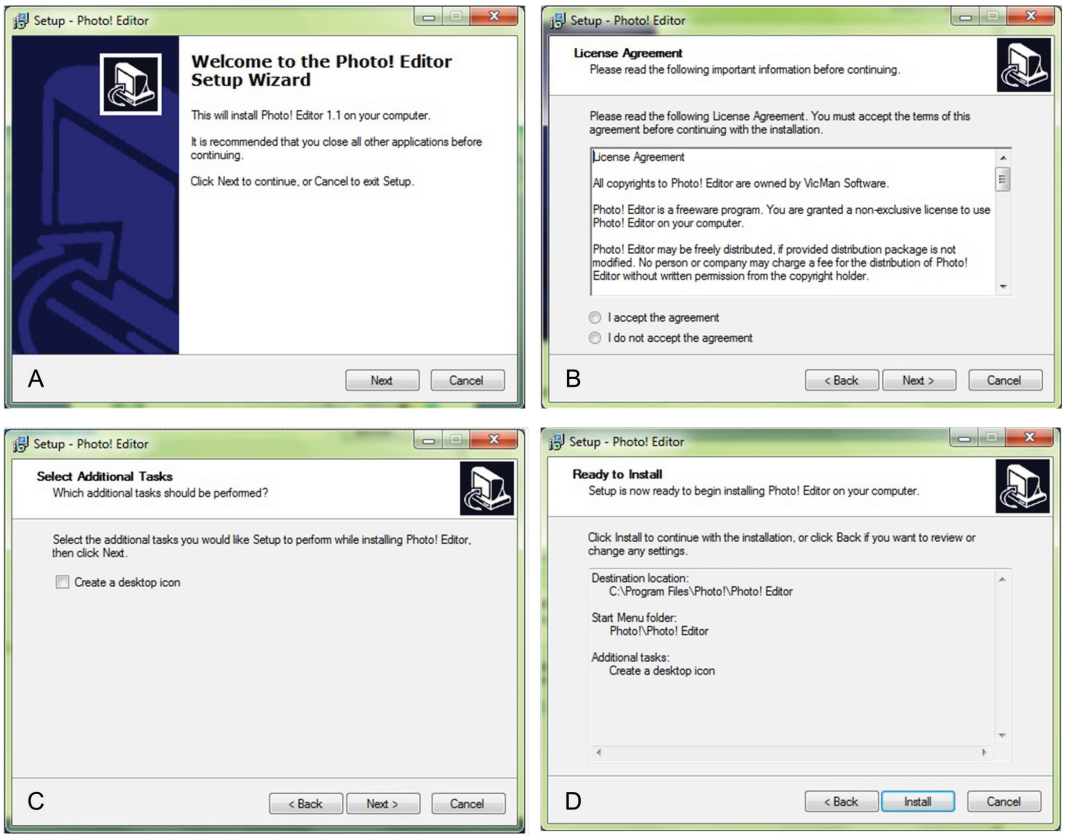


Figure E.1: *Photo! Editor setup wizard.*

In condition 1 the screen below was shown for 4 minutes.

In condition 2 the screen below was shown for 15 seconds, whereafter the tutorial popped as shown in the screen image to the right

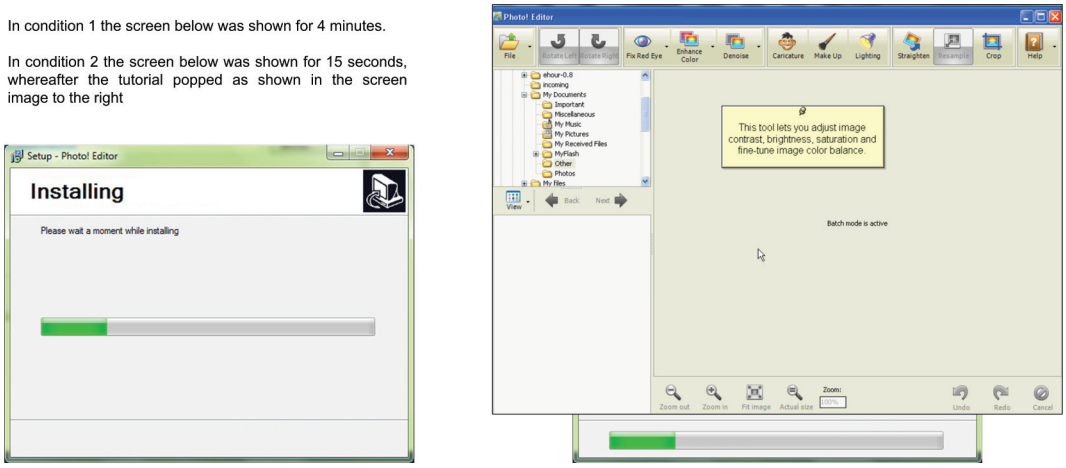


Figure E.2: *The installation of the program. Left: Condition 1 where test subjects are only presented with the progress bar. Right: Condition 2 where test subjects are presented with a tutorial video during the wait.*

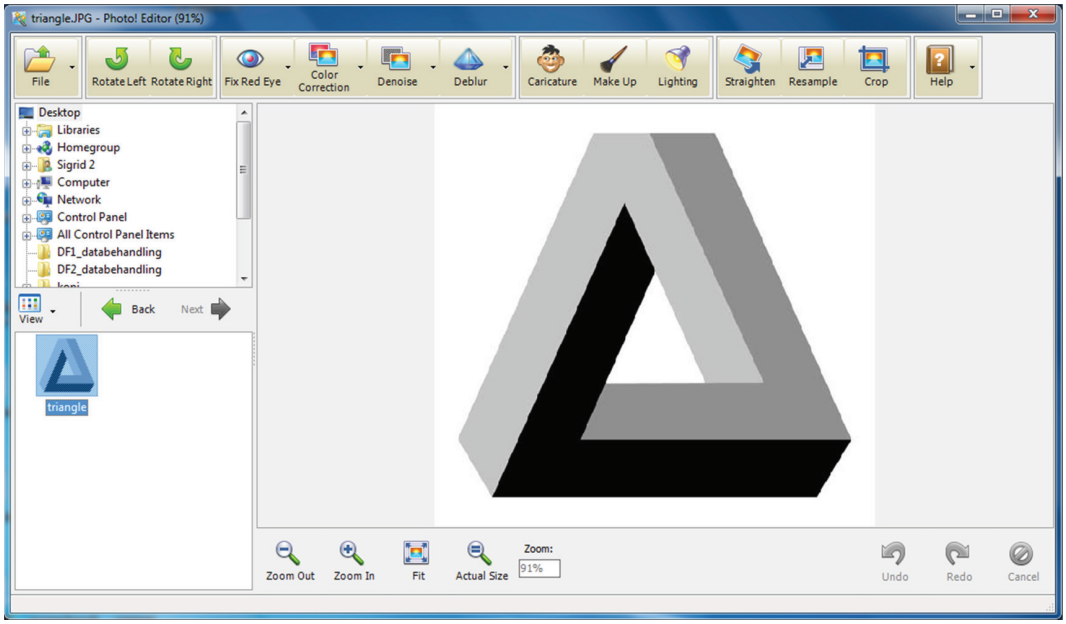


Figure E.3: *Photo! Editor is launched upon installation, and the test subjects edit the triangle photo.*

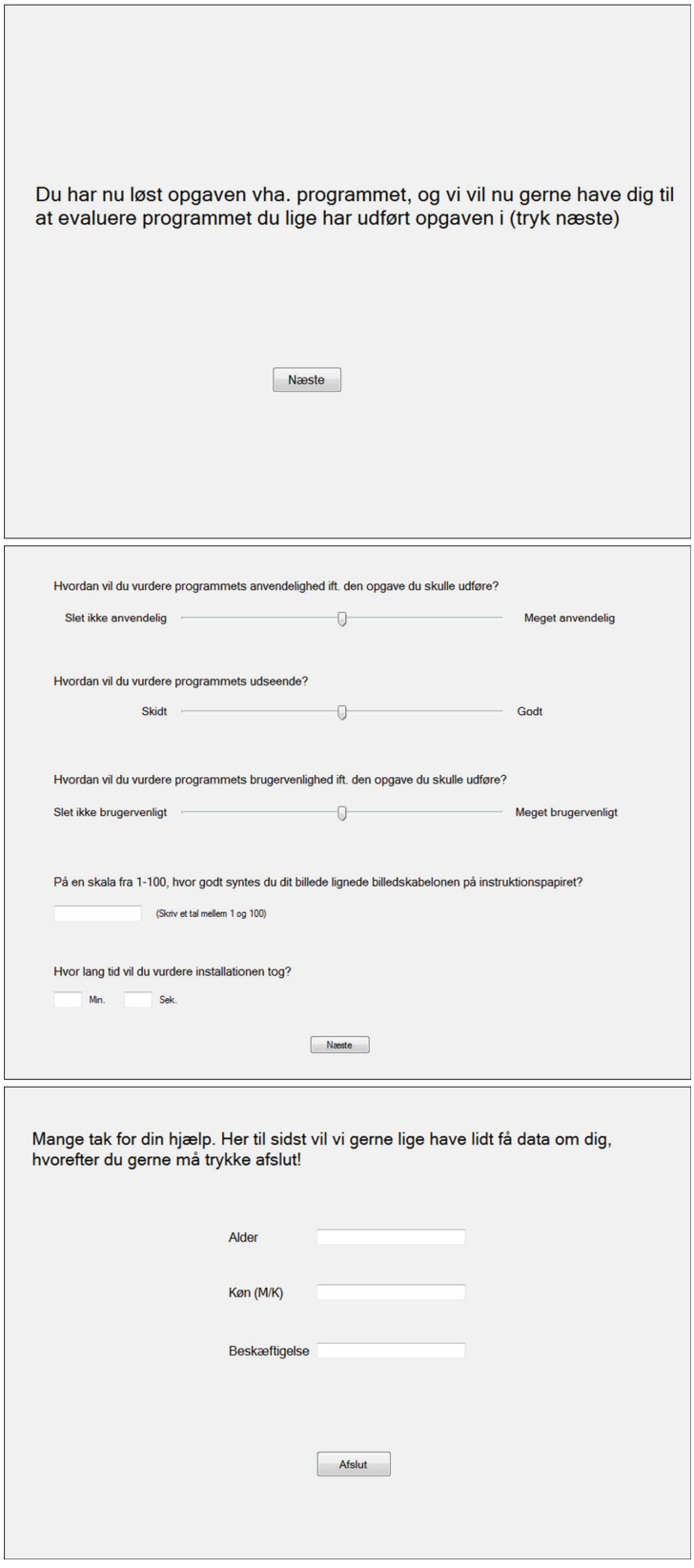


Figure E.4: *Evaluation of the program upon having solved the editing task.*

Introduction sheet to Experiment ST2

Hej og tak fordi du vil deltage i forsøget.

Du skal nu afprøve og vurdere et fotoredigeringsprogram, hvorfor du først skal installere programmet og dernæst udføre en fotoredigeringsopgave i programmet. Installationen er taget med i forsøget, da programmet skal vurderes som en helhed - du skal derfor installere programmet på samme vis som du ville have gjort det privat.
Når du er klar, går du bare i gang med installationen af programmet på computeren og udfører nedenstående opgave i programmet.

Opgave
Du skal nu klikke 1 gang på billedet af trekanten, i venstre side af programmet. Din opgave går ud på at få billedet af trekanten til at ligne det nedenstående billede mest muligt.
Når du føler du har løst opgaven skal du blot kalde på forsøgslederen og derefter udfylde et lille spørgeskema.

Hints:

- Beskær billedet
- Rediger lysindstillingen
- Billedet behøver ikke ligne 100%, så bare markér når du er tilfreds

Program for Experiment FM1

For experiment FM1 slide bar lenghts were varied between two conditions as seen in figure G.1, where only one of the two bottom screenshot were shown to test subjects in each condition: long and short.

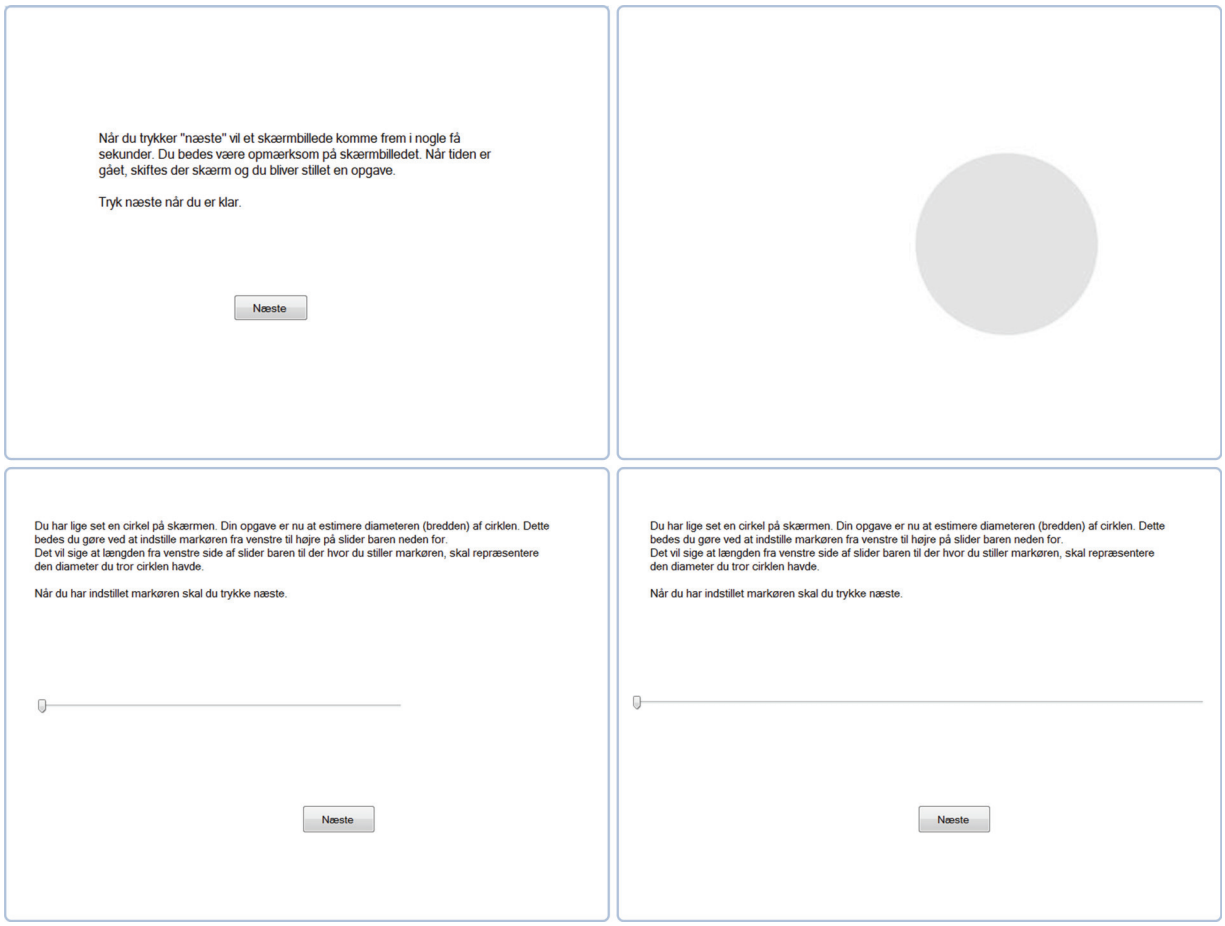


Figure G.1: *Four screenshot from the program used in Experiment FM1 in the order presented to test subjects, however only one of the two bottom screenshots are shown in each of the two conditions.*

Program for Experiment FM2 and DF1

The two screenshots present the program screens shown to the test subjects in experiment FM2 and DF1. For both experiments two conditions were made; for the defaults the preset option was changed between programs, and the length of the framing slider was varied between conditions (no. 7), see figure H.1 and H.2.

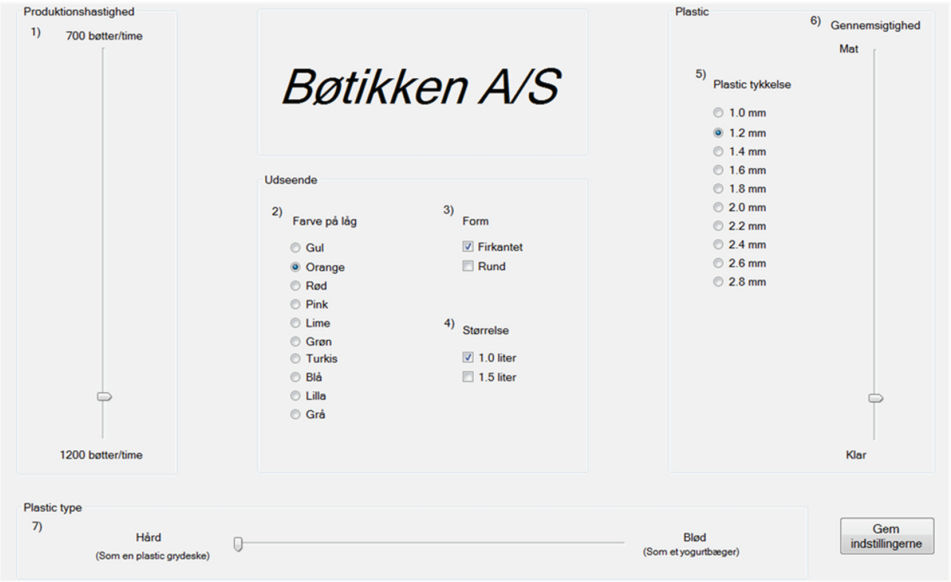


Figure H.1: *Condition 1.*

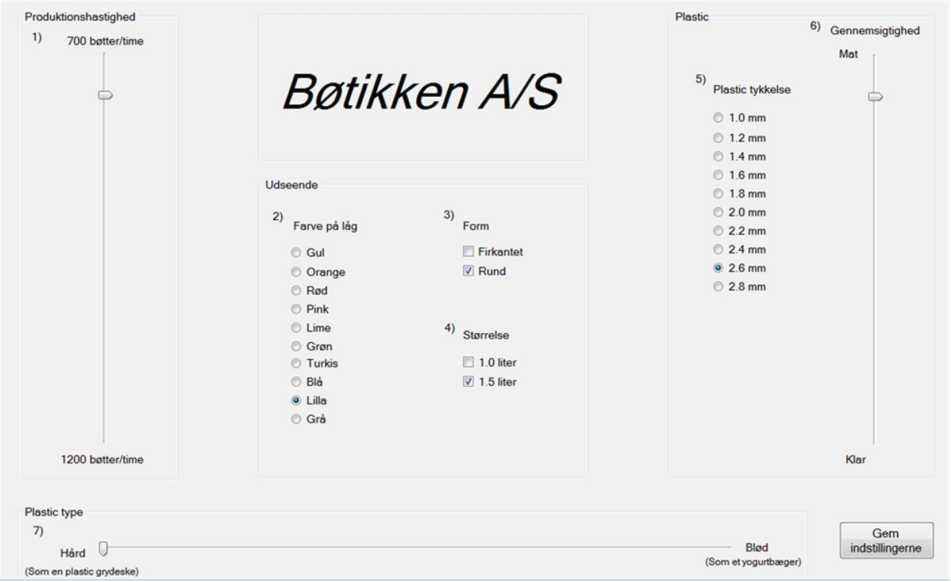


Figure H.2: *Condition 2.*

Experiment FM3 and FM4 Questions

The 16 questions used in experiment FM2 are all presented here. The questions are based on the fact that the test subjects are all students at Aalborg University, why the questions cover subjects like: group work, partying, exams and computers. The questions are categorised into negative, positive and neutral questions. Three of the questions (one in each of the three categories of questions) are special since these questions are constructed with the purpose that significantly more test subjects should choose a specific answer. This answer in each of the three questions is marked with bold in the following.

Negative 1

Du har netop lagt sidste hånd på et afsnit til et studieprojekt, som du har arbejde på hele dagen (godt 8 timer), men pludselig fryser skærmen og da du endelig får gang i computeren igen opdager du, at dit afsnit er blevet slettet. Du ved godt, at man altid skal tage backup af sine filer, men det har du ikke gjort i dette tilfælde. Hvad er din reaktion?

A Du er ærgerlig

B Du er irriteret

C Du fortryder

D Du er bitter

E Du er indebrændt

Negative 2

Du har netop været til matematikeksamen og du føler du har gjort dit bedste. Da du kommer tilbage ind i eksaminationslokalet får du besked om, at din præstation ikke var nok til at bestå. Hvordan reagerer du?

A Du er vred

B Du er hidsig

C Du er skuffet

D Du er gal

E Du er irriteret

Negative 3

This question was originally constructed like this:
Du er i gang med dit semesterprojekt og har lige afsluttet et af de afsnit du skulle skrive. Du har gjort dig umage og synes det er godt. Du sender det rundt til dine gruppemedlemmer, og dagen efter skal I sammen diskutere det. Alle dine gruppemedlemmer kritiserer dog dit afsnit, og siger rent ud, at de synes du skal skrive det helt om. Hvad er din reaktion?

However, in the pilot experiment several test subjects experienced many difficulties and expressed that they would not react as negative as the five answers they had to choose between. Due to these difficulties the question was changed and the final question wording can be seen in the following.

Du skal have nogle venner over i aften til lidt hygge. Du køber derfor en lækker kage ved bageren og sætter den i køleskabet. Du bor på et kollegium med delekøkken, hvorfor køleskabet også er et delekøleskab. Om aftenen da dine venner kommer, går du ud i køkkenet for at hente kagen. Det viser sig, at nogen har spist næsten halvdelen af kagen. Hvordan reagerer du?

A Du er ophidset

B Du er irriteret

C Du er vred

D Du er sur

E Du er gal

Negative 4

Du skal til forelæsning, hvor du skal holde et oplæg for de andre på din årgang. Du føler dig overhovedet ikke sikker i det du skal fremlægge om, men du ved du har gjort dit bedste. Du fremlægger, men under fremlæggelsen kan du høre, at dine medstuderende begynder at fnise og hviske, og se meget mærkeligt på dig. Da du er færdig med at fremlægge kommenterer forelæseren og dine medstuderende på din fremlæggelse, og det viser sig at du fuldstændig har misforstået emnet, og dummet dig meget. Hvordan føler du det i situationen?

A Du bliver ked af det

B Du bliver trist

C Du bliver ærgerlig

D Du bliver deprimeret

E Du bliver ilde til mode

Negative 5

Du er den i din projektgruppe, der har skrevet det meste af projektet, og har brugt mest tid på det generelt. Men til projekteksamen får du den dårligste karakter i gruppen, hvordan reagerer du?

A Du bliver harm

B Du bliver oprørt

C Du bliver bitter

D Du bliver fortørnet

E Du bliver forarget

Positive 1

Du har fødselsdag og bliver hentet af en go ven. Planen er, at I skal hjem til vennen og fejre din fødselsdag med go mad og lidt at drikke. Da du træder ind af døren går det dog op for

dig at din ven har arrangeret et surprise party for dig i stedet for, og alle dine venner er der for at fejre dig. Hvordan reagerer du?

A Du er glad

B Du er lykkelig

C Du er overstadig

D Du er ellevild

E Du er ekstatisk

Positive 2

Du har længe været på udkig efter en ny sommerjakke og har også fundet én du rigtig gerne vil have. Den er dog alt for dyr til dit budget. Du har derfor opgivet at få denne jakke og er på udkig efter en anden, da du ser jakken på udsalg til under halv pris. Du har nu netop råd til at købe jakken, hvilket du gør. Hvad føler du bagefter?

A Du føler dig opløftet

B Du føler dig heldig

C Du føler dig glad

D Du føler dig begejstret

E Du føler dig lykkelig

Positive 3

Du har været i byen hele natten og da du vågner næste morgen kan du ikke finde din pung. Du indser til sidst, at den må være blevet stjålet eller smidt væk i løbet af natten og du ærgrer dig rigtig meget. Da du senere på dagen tømmer din postkasse, finder du din pung deri sammen med en note om, at din pung er blevet fundet på dansegulvet af det diskotek du festede på. Hvordan reagerer du?

A Du er ellevild

B Du er himmelhenrykt

C Du er taknemmelig

D Du er begejstret

E Du er ekstatisk

Positive 4

Du har købt et skrabelod til 20 kr. og efter at have skrabet alle felterne indser du, at der er 3 felter med 5.000 kr. Du skynder dig tilbage til kiosken, hvor du har købt det, og de udbetaler dig gevinsten på 5.000 kr. Hvordan reagerer du?

A Du føler dig munter

B Du føler dig henrykt

C Du føler dig lykkelig

D Du føler dig heldig

E Du føler dig begejstret

Positive 5

Du går en tur igennem gågaden med din cykel i hånden. Du går og kigger lidt på nogle butiksruder, solen skinner og du går og hygger dig. Pludselig er der én der prikker dig på skulderen, og du stopper op og vender dig om. Det er en ældre herre, der står med din pose som du havde sat på bagagebæreren. Han forklarer, at den faldt af for lidt siden, og at han lige ville gøre dig opmærksom på det, så du ikke mistede den. Hvilken følelse fortsætter du din gåtur med?

A Du er henrykt

B Du er taknemmelig

C Du er glad

D Du er fornøjet

E Du er opstemt

Neutral 1

Du går en tur igennem byen, da en kvinde stopper dig. Hun spørger dig, om du kan fortælle hende vejen til Irma. Du kender faktisk ikke vejen til Irma, og siger:

A Jeg ved ikke hvor det er

B Jeg er ikke kendt her i området

C Jeg kan ikke hjælpe dig

D Jeg kender Irma, men ikke vejen

E Jeg kender ikke vejen dertil

Neutral 2

Du vinder 10 millioner kroner i Lotto, og af ren og skær glæde, vælger du at donere 100.000 kr. til et godgørende formål i Afrika, hvilket vælger du:

A Børn i Kenya

B Børn i Namibia

C Børn i Marokko

D Børn i Tanzania

E Børn i Ghana

Neutral 3

Du er på en ny type date, som egentlig er en blind date, dvs. du ikke kan se din date, men tale med ham/hende. Hvad ville du først spørge ham/hende om?

A Skostørrelse

B Hårfarve

C Øjenfarve

D Alder

E Højde

Neutral 4

Det er din vens fødselsdag, og du ved han intet værktøj har, men han ønsker sig en masse basisværktøj. Hvad giver du ham?

A En tang

B En sav

C En tommestok

D En hammer

E En skruetrækker

Neutral 5

Du er taget til Tyskland for at købe ind til en fest. Ud over øl, vand, spiritus og mad skal du også købe noget chokolade til kaffen. Butikken har lige den chokolade du gerne vil have, men i fem forskellige varianter. Dvs. chokoladen i æskerne er helt den samme, men har forskellig form. Hvilken form chokolade vælger du?

A Sekskantet

B Rund

C Firkantet

D Stænger

E Kugleformet

Neutral 6

Du skal i byen og skal sammen med din ven finde ud af hvilken bar I skal starte på. Der er 5 barer i byen, og din ven har ingen bestemt præference. Umiddelbart er der ikke den store forskel på de 5 barer; de ligger alle i centrum og de er lige populære. Hvilken bar synes du I skal tage på først.

A Lommelærken

B Kælderstuen

C Svinget

D Bjælkehytten

E Lanternen

Control Program for Experiment FM3

For experiment FM3 conditions of colour and no colour were varied between groups. The following is screenshots of the program developed in visual basic 2008 Express Edition for control condition A1. The program was made in the four conditions A1, A2, B1 and B2, all with no colour differing between questions and orders of questions and answers.

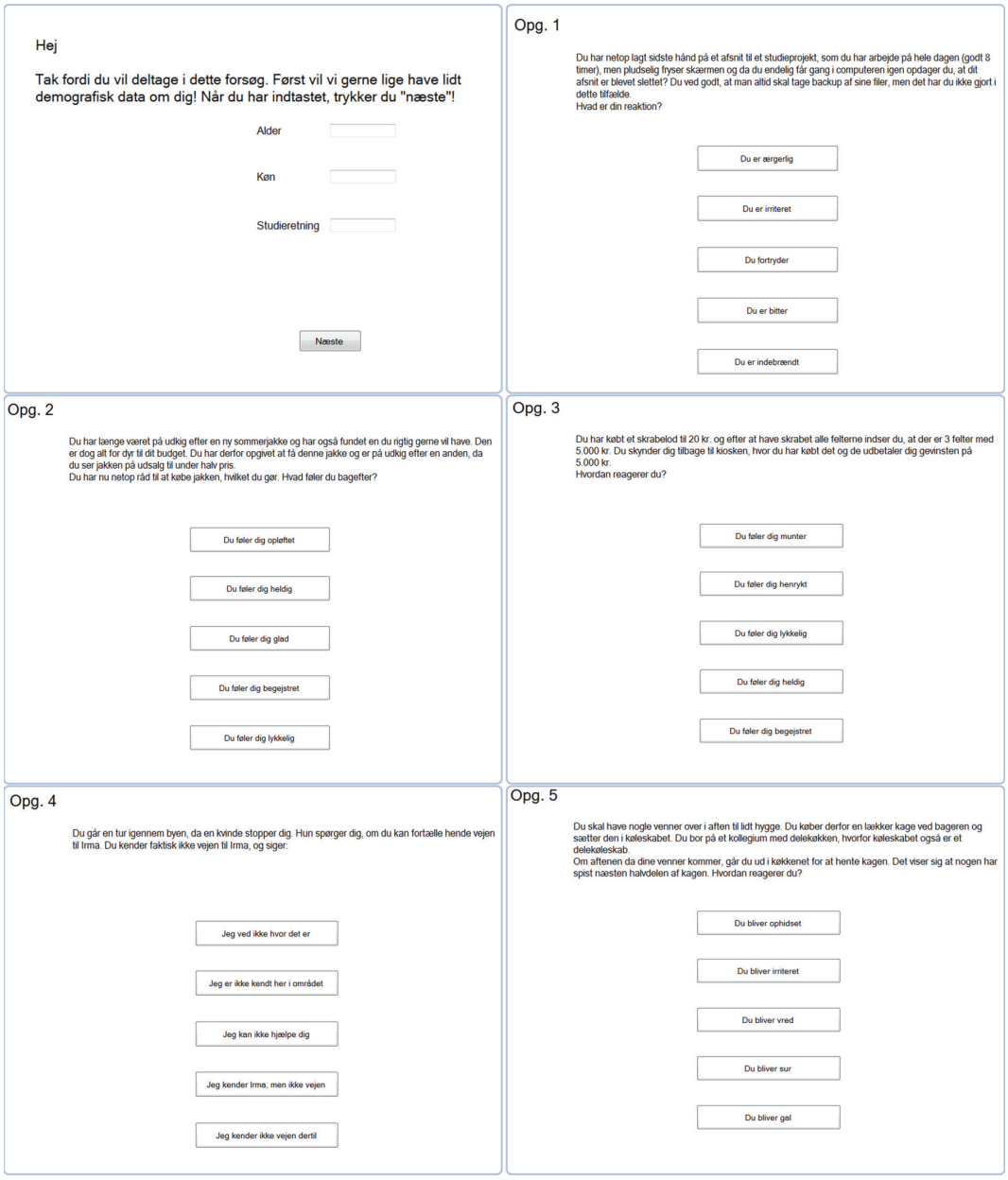


Figure J.1: *Six screenshots in the order presented to test subjects from the program used in Experiment FM3 in control condition A1.*

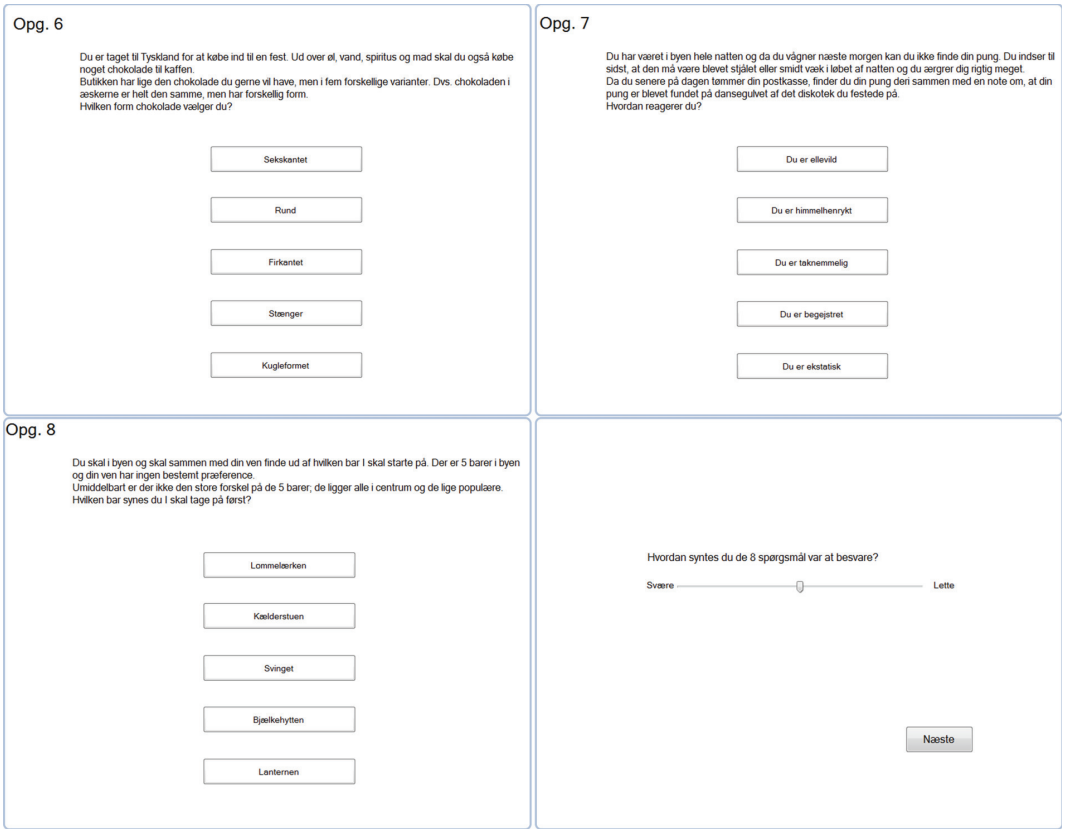


Figure J.2: *The last four screenshots from the from the program used in Experiment FM3 in control condition A1.*

Colour Program for Experiment FM3

For experiment FM3 conditions of colour and no colour were varied between groups. The following is screenshots from the program developed in visual basic 2008 Express Edition for colour condition FB1. The program was made in the four conditions FA1, FA2, FB1 and FB2, all with colours attached to the answer buttons and the conditions differing between questions and orders of questions and answers.

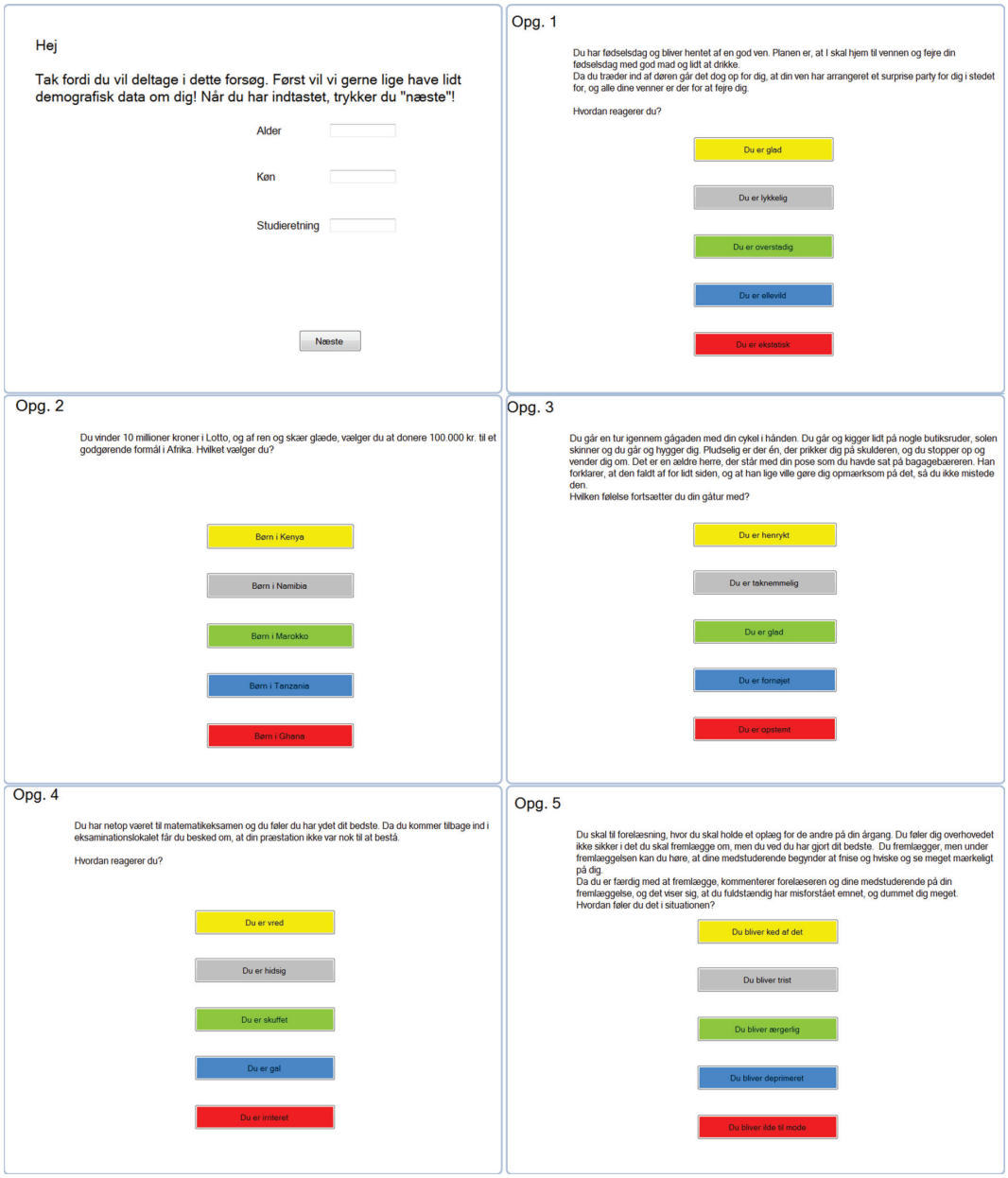


Figure K.1: *Six screenshot in the order presented to test subjects from the program used in Experiment FM3 in colour condition FB1.*

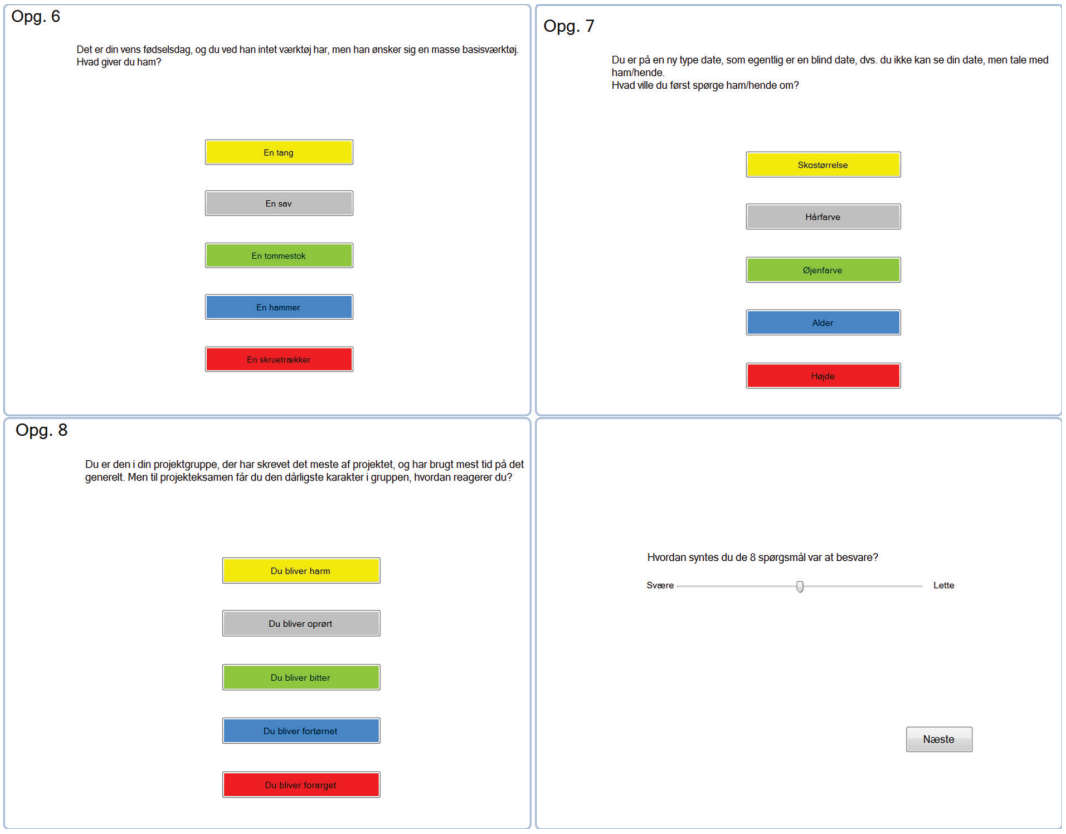


Figure K.2: *The last four screenshot from the from the program used in Experiment FM3 in colour condition FB1.*

Questionnaire for Experiment FM4

For experiment FM4 colours were varied between five conditions. The following screenshots were made into a questionnaire and divided to test subjects. The questionnaire was made in the five conditions: blue, green, yellow, red and grey.

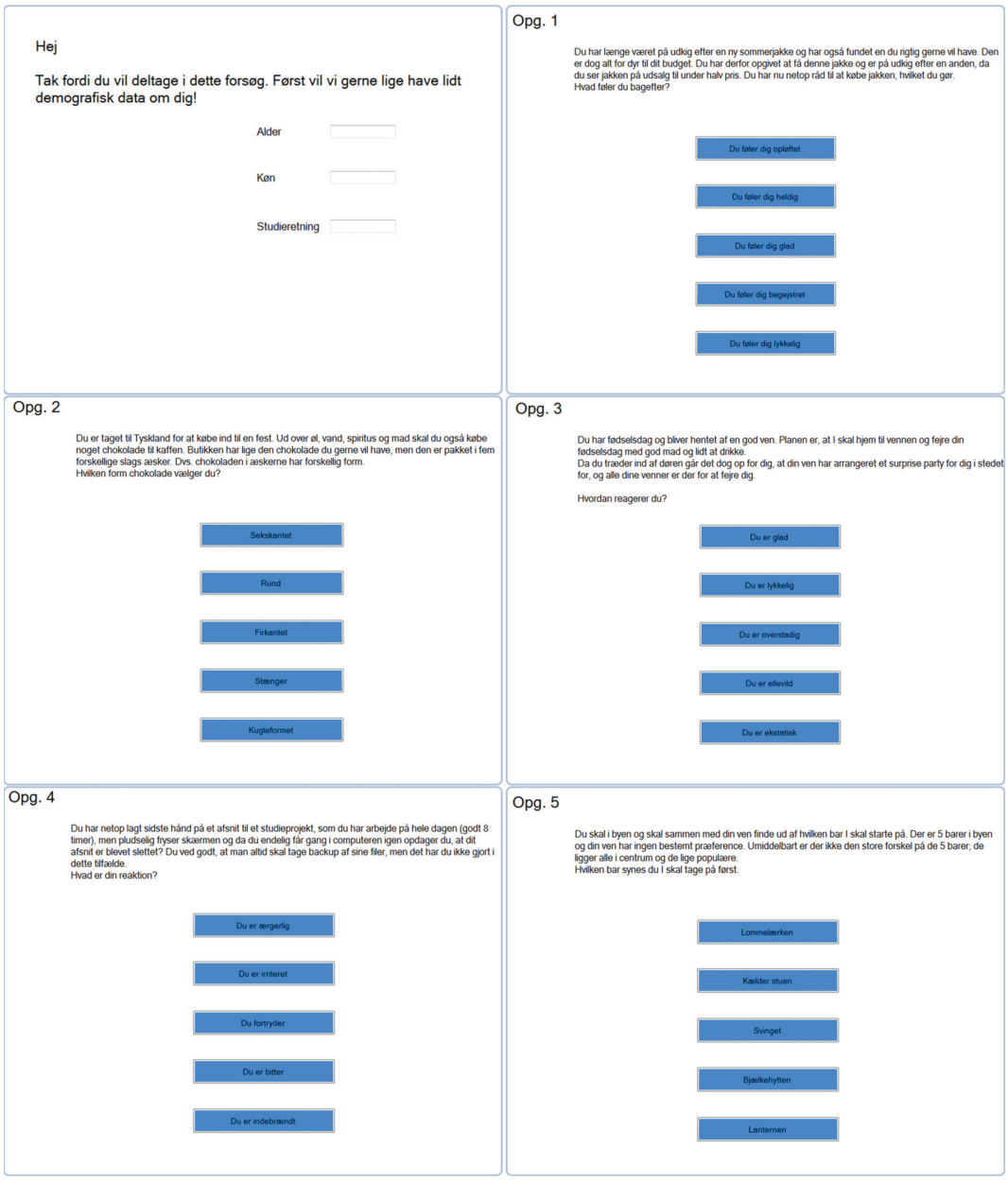


Figure L.1: *Six screenshots in the order presented to test subjects from the program used in Experiment FM4 condition blue.*

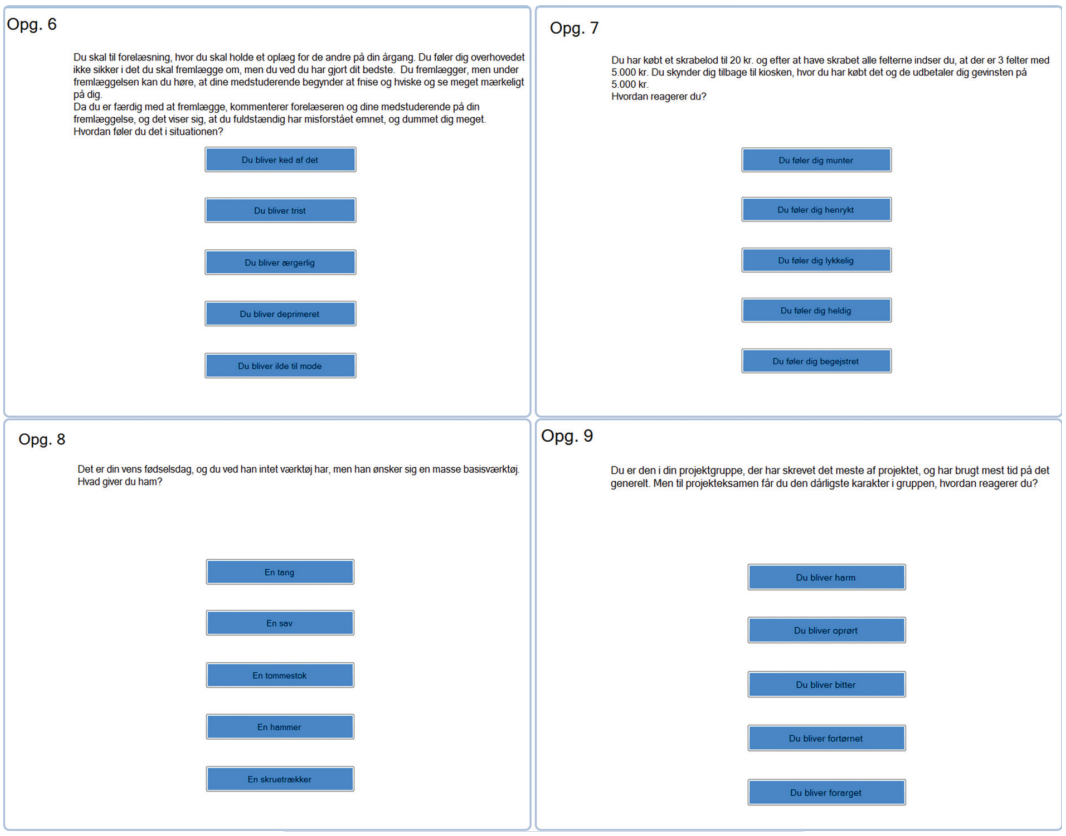


Figure L.2: *The last five screenshots from the from the program used in Experiment FM4 in condition blue.*

Graphical Results from Experiment FM4

A graphical overview of the answers made in each of the nine questions in Experiment FM4.

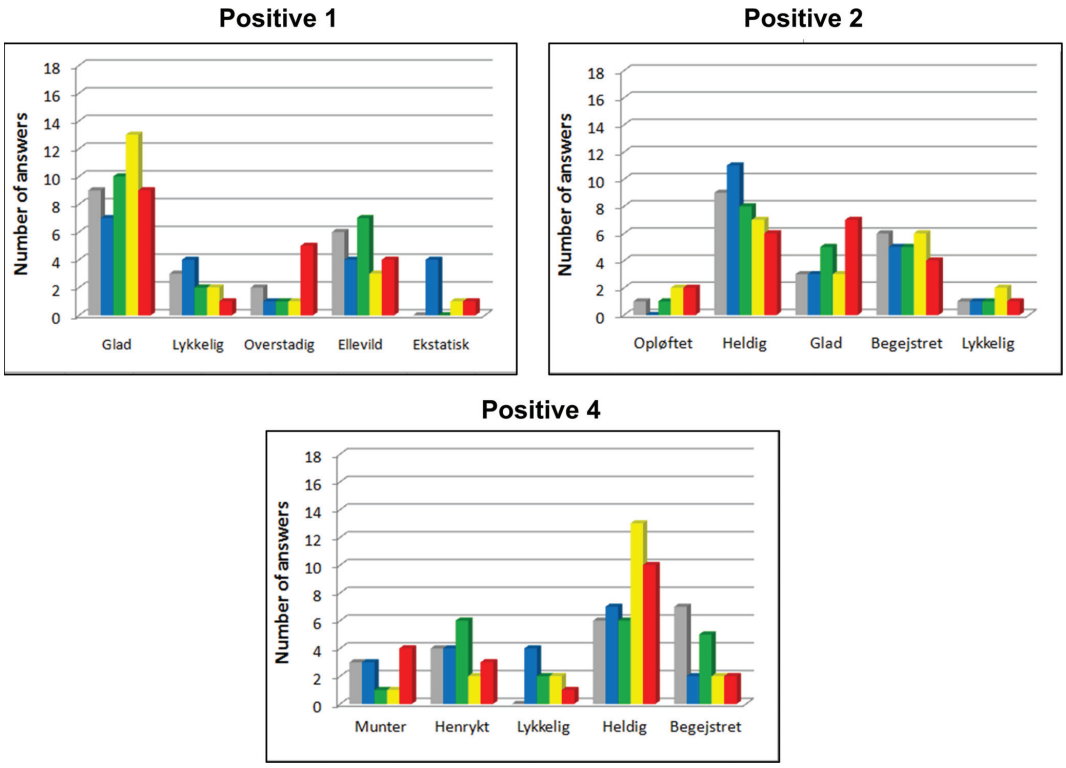


Figure M.1: *The graphs presents the answers made in each of the positive questions in Experiment FM4.*

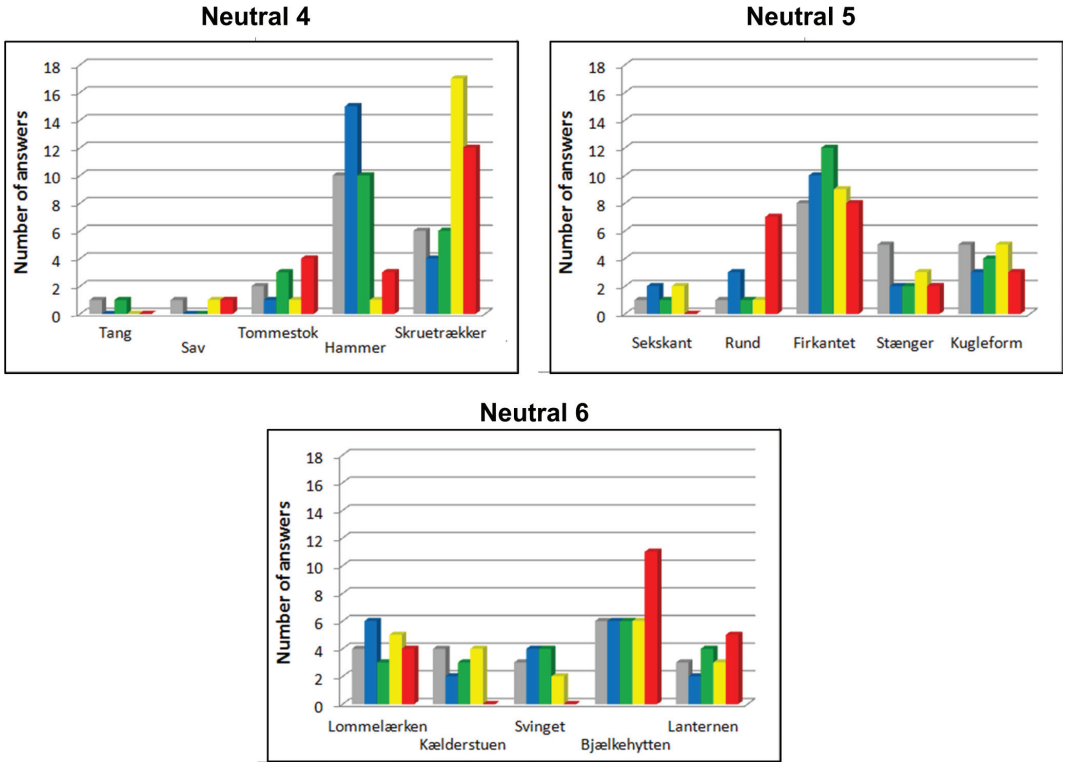


Figure M.2: *The three graphs presents the answers made in the three neutral questions in Experiment FM4.*

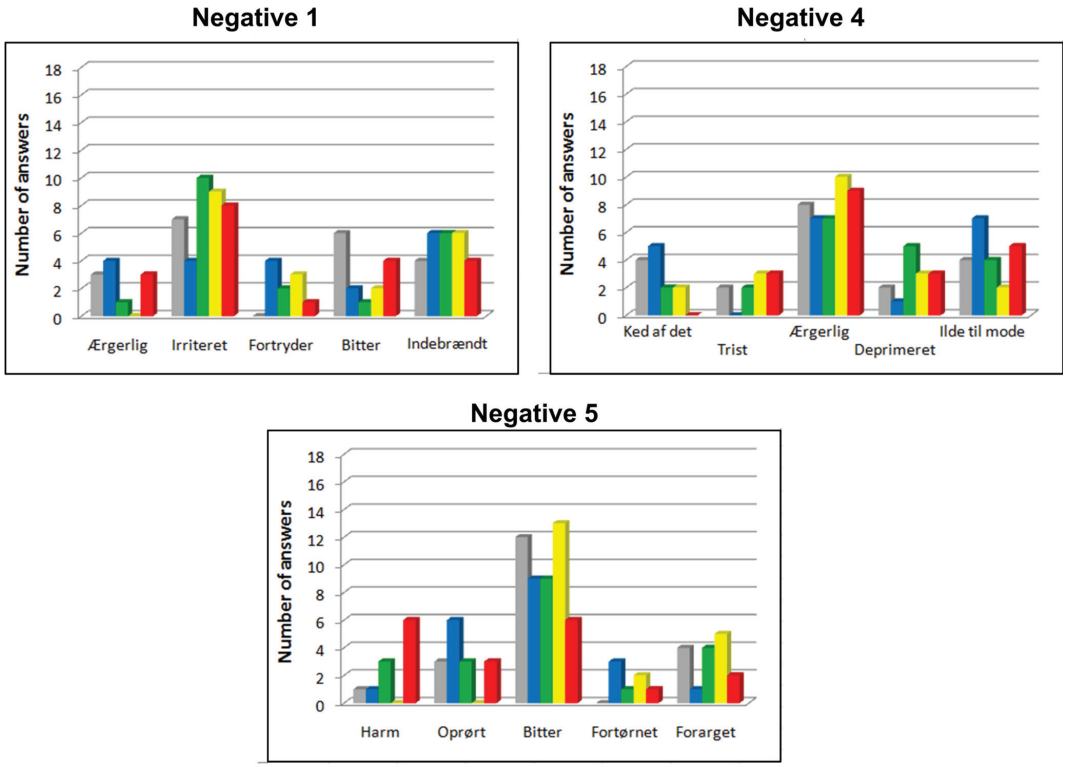


Figure M.3: *Three graphs presenting the answers made in each of the three negative questions in Experiment FM4.*

Program for Experiment FM5

For experiment FM5 where two colours (red and green) were varied between two conditions. The following is screenshots of the program developed in visual basic 2008 Express Edition for both conditions.

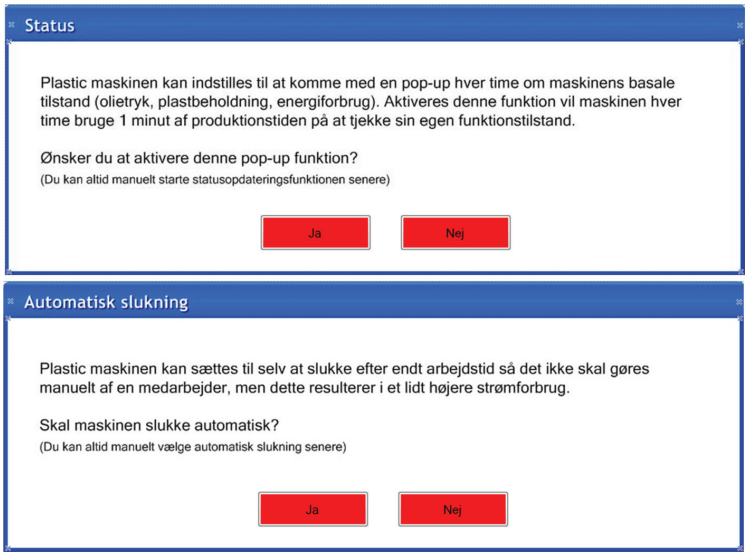


Figure N.1: *Two screenshots in the order presented to test subjects from the program used in Experiment FM5 condition red.*

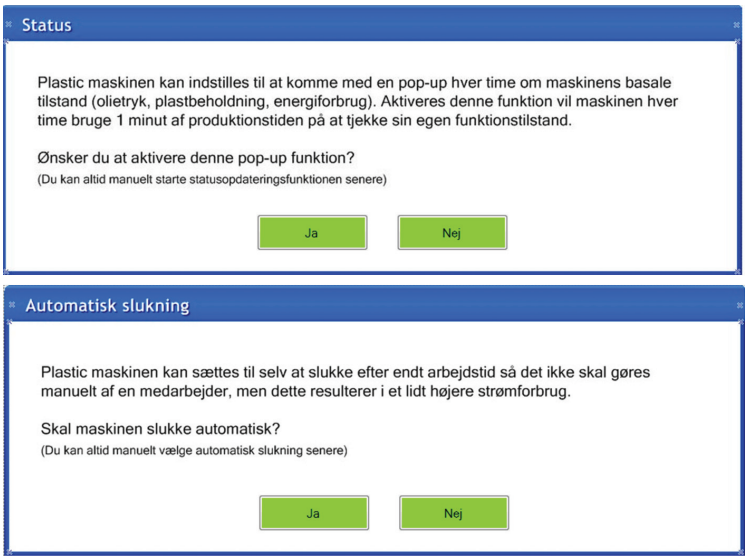


Figure N.2: *Two screenshots in the order presented to test subjects from the program used in Experiment FM5 condition green.*

Program for Experiment DF2

For experiment DF2 two versions of the same program were used with different default settings, as shown in figure O.1.

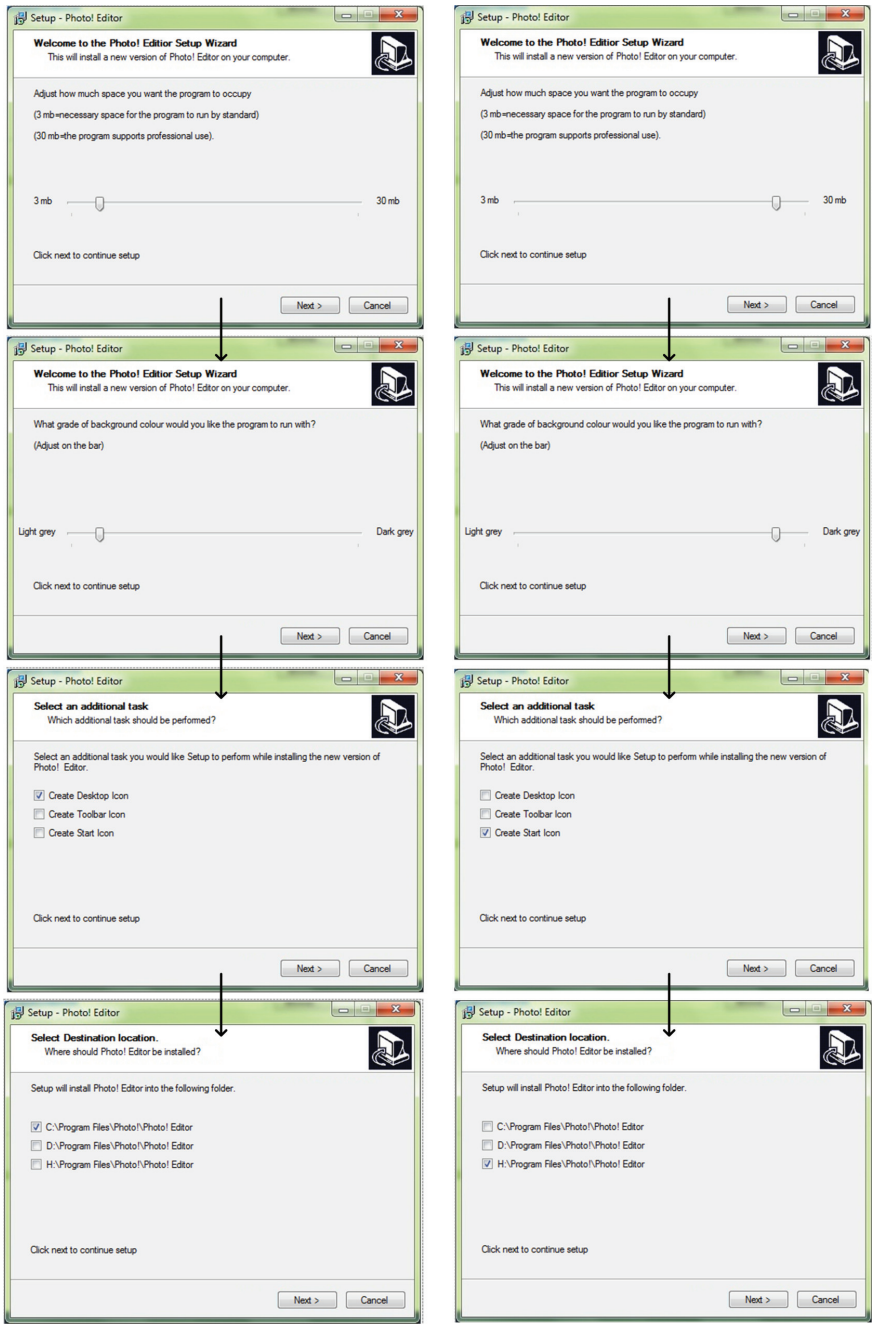


Figure O.1: *DF2 Program overview. Right: Condition 1 and left: Condition 2.*

ATTACHMENT **P**

Questionnaire STA

In this section the original decision problems tested by Arkes and Blumer (1985), Navarro and Fantino (2008) and Soman (2001) are presented. These four questions are altered to fit the purpose of Experiment STA, why the alterations made are shown below each of the original questions.

Original Arkes and Blumer

Part 1: As the president of an airline company, you have invested 10 million dollars of the company's money into a research project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed, another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company is building. The question is:

Should you invest the last 10% of the research funds to finish your radar-blank plane?

-Yes
-No

Part 2: As president of an airline company, you have received a suggestion from one of your employees. The suggestion is to use the last 1 million dollars of your research funds to develop a plain that would not be detected by conventional radar, in other words, a radar-blank plane. However, another firm has just begun marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company could build. The question is:

Should you invest the last million dollars of your research funds to build the radar-blank plane proposed by your employee?

-Yes
-No

Altered Arkes and Blumer

Part 1: As the president of a ***computer company, you have devoted 5 years of the research department´s time*** *in a research project. The purpose was to build* ***a new kind of operating system that offers a special new technology to the user****. When the project* ***has been running for 4 years and 6 months*** *another firm begins marketing* ***an operating system using the special new technology****. Also, it is apparent that their system is much faster and far more economical than the* ***system*** *your company is building. The question is:*

Should you invest the ***last 6 months of the research department's time*** *on finishing the* ***operating system****?*

-Yes
-No

Part 2: As the president of ***a computer company, you are considering to devote 6 months of the research department´s time in a research project. The purpose of the research is to build a new kind of operating system that offers a special new technology to the user****. However, another firm has just begun marketing* ***an operating system using the special new technology****. Also, it is apparent that their* ***system*** *is much faster and far more economical than the* ***system*** *your company is building. The question is:*

Should you invest ***the 6 months of the research department's time*** *on developing the* ***operating system?***

-Yes
-No

Original Soman

Imagine that you recently saw an advertisement on the bulletin board. A literature professor was looking for a research assistant to work for about 15 hours. The payment was in the form of a front row seat to a professional theatre performance. On the same bulletin board, a music professor was also looking for a research assistant to work for about 5 hours and this assistant would be paid with a ticket (in a good section) to a rock concert. You think you would like to see both these events, although you expect to like the rock concert more. You work for both professors (15 hours for literature and 5 hours for music) and get paid with the two tickets (theatre and rock concert respectively). As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are non-refundable, nor can they be exchanged. You can use only one of the tickets and not the other.

Which ticket will you use?

-Theatre performance ticket
-Rock concert ticket

Altered Soman

Imagine that you recently saw an advertisement on the bulletin board. A literature professor was looking for a research assistant to work for about ***50*** *hours. The payment was in the form of a front row seat to a professional theatre performance. On the same bulletin board, a music professor was also looking for a research assistant to work for about 5 hours and this assistant would be paid with a ticket (in a good section) to a rock concert. You think you would like to see both these events, although you expect to like the rock concert more. You work for both professors (****50*** *hours for literature and 5 hours for music) and get paid with the two tickets (theatre and rock concert respectively). As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are non-refundable, nor can they be exchanged. You can use only one of the tickets and not the other.*

Which ticket will you use?

-Theatre performance ticket
-Rock concert ticket

Original Navarro and Fantino

Imagine that you are the leader of a copper-mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek." So far, your group has been digging at Shady Creek for the past 60 days.
[The ground so far has been soft, so the time spent digging has been easy. This morning, you and your group found a vein of copper! / The ground so far has been hard. The time spent diging has been very difficult and has involved a lot of effort from you and your group.]
However, there is a thick wall of quartz, which is easy to dig through, covering most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small-about 10 pounds. Typical mines contain upwards of 500 pounds.

You have a choice to make:

A. dig the 10 more days to collect the 10 pounds of copper.
B. abandon this "Shady Creek" mine and go home.

Altered Navarro and Fantino

Version 1: Imagine that you are the leader of a copper-mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek." So far, your group has been digging at Shady Creek for the past 60 days. ***The ground so far has been soft, so the time spent digging has been effortless.*** *This morning, you and your group found a vein of copper! However, there is a thick wall of quartz, which is easy to dig through, covering most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small-about 10 pounds. Typical mines contain upwards of 500 pounds.*

You have a choice to make:

A. dig the 10 more days to collect the 10 pounds of copper.
B. abandon this "Shady Creek" mine and go home.

Version 2: Imagine that you are the leader of a copper-mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek." So far, your group has been digging at Shady Creek for the past 60 days. ***The ground so far has been very hard, so the time spent digging has been complicated and cost you and your group a lot of trouble and effort.*** *This morning, you and your group found a vein of copper! However, there is a thick wall of quartz, which is easy to dig through, covering most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small-about 10 pounds. Typical mines contain upwards of 500 pounds.*

You have a choice to make:

A. dig the 10 more days to collect the 10 pounds of copper.
B. abandon this "Shady Creek" mine and go home.

Original Arkes and Blumer

On your way home you buy a TV dinner on sale for $3 at the local grocery store. A few hours later you decide it is time for dinner, so you get ready to put the TV dinner in the oven. Then you get an idea. You call up your friend to ask if he would like to come over for a quick TV dinner and then watch a good movie on TV. Your friend says "Sure". So you go out to buy a second TV dinner. However, all the on-sale TV dinners are gone. You therefore have to spend $5 (the regular price) for the TV dinner identical to the one you just bought for $3. You go home and put both dinners in the oven. When the two dinners are fully cooked, you get a phone call. Your friend is ill and cannot come. You are not hungry enough to eat both dinners. You cannot freeze one. You must eat one and discard the other.

Which one do you eat?

-$3 dinner
-$5 dinner
-No preference

Altered Arkes and Blumer

On your way home you buy a ***ready-made meal for 20 kr****. at the local grocery store. A few hours later you decide it is time for dinner, so you get ready to put the* ***ready-made meal*** *in the oven. Then you get an idea. You call up your friend to ask if he would like to come over for a quick ready-made meal and then watch a good movie on TV. Your friend says "Sure." So you go out to buy a second* ***ready-made meal. When you get to your car, it won't start. Instead you decide to go by bike. On the way to the local grocery store your wheel hits a rock and your tire punctures, why you need to walk your bike the rest of the way. At the store you are buying the second ready-made meal for 20 kr. Returning to your bike you discover that you have forgotten your bike key, which means that you have to carry your bike all the way home. Finally at home you*** *put both dinners in the oven. When the two dinners are fully cooked, you get a phone call. Your friend is ill and cannot come. You are not hungry enough to eat both dinners. You cannot freeze one. You must eat one and discard the other.*

Which one do you eat?

-The meal bought on the way home
-The meal bought for your friend
-No preference

ATTACHMENT **Q**

Questionnaire STB

Based on experiences in Experiment STA, small aspects of each of the questions were decided to be changed. In the following the changes made between Experiment STA and STB are highlighted.

Altered Arkes and Blumer

Part 1: *As the president of a computer company you have devoted 5 years of the research department´s time in a research project. The purpose was to develop a new kind of operating system that offers a special new technology to the user. When the project has been running for 4 years and 6 months, another firm begins marketing an operating system using special new technology. Also, it is apparent that their system is much faster and far more economical* ***and all together much better*** *than the system your company is building. The question is: should you devote the last 6 months of the research department's time on finishing the operating system?*

Part 2: *As the president of a computer company you are considering to devote 6 months of the research department´s time in a research project. The purpose with the research is to develop a new kind of operating system that offers a special new technology to the user. However, another firm has just begun marketing an operating system that uses this special new technology. Also, it is apparent that their system is much faster, far more economical* ***and all together much better*** *than the system your company is building. The question is: should you devote the 6 months of the research department´s time on developing the operating system?*

Altered Soman

Imagine that you recently saw an advertisement on the bulletin board. A literature professor was looking for a research assistant to work for about 50 hours. The payment was in the form of a ticket to ***an event (event 1)****. On the same bulletin board, a music professor was also looking for a research assistant to work for about 5 hours and this assistant would be paid with a ticket for* ***another event (event 2)****.*

You think you will like to attend both events, although you expect to like ***event 2*** *more. You work for both professors (50 hours for literature and 5 hours for music) and get paid with the two tickets. As you are putting the tickets away in your wallet, you notice that both events are scheduled for the same evening and are both at good locations on campus. The tickets are non-transferable, nor can they be exchanged. You can use only one of the tickets and not the other.*

Which ticket will you use?

-***Ticket for event1***
-***Ticket for event2***

Altered Navarro and Fantino

Version 1: Imagine that you are the leader of a copper-mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek". So far, your group has been digging at Shady Creek for the past 60 days. The ground so far has been soft, so the time spent digging has been effortless. This morning, you and your group found a vein of copper! However, there is a thick wall of quartz, which is easy to dig through, covering most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small-about 10 pounds. Typical mines contain upwards of 500 pounds.

You have a choice to make:

A. dig the 10 more days to collect the 10 pounds of copper.
B. abandon the "Shady Creek" mine to give up the dig at "Shady Creek" and search for another place to dig

Version 2: Imagine that you are the leader of a copper-mining group. Your group's job is to dig the ground in search of copper. Currently, your group is digging at a local spot known as "Shady Creek". So far, your group has been digging at Shady Creek for the past 60 days. The ground so far has been very hard, so the time spent digging has been complicated and cost you and your group a lot of trouble and effort. This morning, you and your group found a vein of copper! However, there is a thick wall of quartz, which is easy to dig through, covering most of the copper. Some special equipment you have (which is 100% accurate) indicates that it will take 10 more days to dig through the quartz and collect all the copper. Unfortunately, your equipment also indicates that the amount of copper is small-about 10 pounds. Typical mines contain upwards of 500 pounds. You have a choice to make:

A. dig the 10 more days to collect the 10 pounds of copper.
B. abandon the "Shady Creek" mine to give up the dig at "Shady Creek" and search for another place to dig

Altered Arkes and Blumer

On your way home you buy a ready-made meal for 20 kr. at the local grocery store. A few hours later you decide it is time for dinner, so you get ready to put the ready-made meal in the oven. Then you get an idea. You call up your friend to ask if he would like to come over for a quick ready-made meal and then watch a good movie on TV. Your friend says "Sure". So you go out to buy a second ready-made meal.
However, they are sold out so you have to go to a second grocery store about 2 km away. Having arrived at the store you are buying the second ready-made meal for 20 kr. *At home again you put both dinners in the oven. When the two dinners are fully cooked, you get a phone call. Your friend is ill and cannot come. You are not hungry enough to eat both dinners. You cannot freeze one. You must eat one and discard the other. Which one do you eat?*

STA and STB Comment Box Replies

In this section the comments test subjects have made in the "explain your choice" box in Experiment STA and STB are presented. Not all comments are present, but only a selection of the most interesting.

Experiment STA

The comments are organised in such a manner that the comments for the computer company are presented first, followed by the rock/theatre ticket, the dig at Shady Creek and last the ready made meal.

Computer Company

- There is much more to a operating system and speed and economy. It would be very unlikely that our OS could not differentiate itself in some way and hence compete. But if 90% of the time wasn't passed it would be another situation.
- Ellers er tiden spildt.
- It's a had choice but both solutions can make sense - I think!
- Maybe my system will have a better security and is more easy to use, it could be a better program.
- Dumt at spilde 4 år + 6 mdr. Der må være en målgruppe til jeres operativ system også.

Rock or Theatre Ticket

- I'd prefer the rock concert... I'm just more of a music person.
- Jeg vil hellere til rock koncert så er det lige meget med de 50 timer jeg har spildt.
- I have worked hardest for this ticket.
- I like rock better.
- I like rock better, even though I have worked hard for the theater ticket.
- because it took me 50 hours to earn this ticket, and because I expect to like the rock concert more, I would be more likely to go to another rock concert later. this could be an eye-opener.
- I worked the hardest for the theatre tickets - I'd give the other tickets to someone else.
- If the rock concert is the event that I want to see the most, I would defenitely go to the rock concert, and ask somebody else if they would have theatre ticket.
- Though it would be exstremely annoying to have worked 50 hours for nothing, I still find that it makes more sense to work 55 hours to see something you expect you will enjoy the most, than work 55 hours to see something you might not like that much.

- I like rock better, even though I have worked hard for the theater ticket.
- I like rock better, but I have worked hard for the other ticket, and would give it away.
- I like rock better, but I have worked hard for the other ticket, and will be irritated that I have worked 50 hours.
- Muligt man ikke har arbejde så meget for den, men tænk på hvor meget man har lært, på de 50 timer, som nogen måske mener er spildt, når man ikke bruge fortjenesten.
- If the rock concert is the event that I want to see the most, I would defenitely go to the rock concert, and ask somebody else if they would have theatre ticket.

Dig at Shady Creek

- I have already used 60 days.
- When we've finally found the copper, then why not get it out of the ground - 10 more days out of the already 60 that has passed isn't a lot.
- Because I would not like that someone else came and used my 60 days of work to easily get to the copper, plus I wouldn't know if there was copper in the next mine I would go to.
- Since there is no mention of other places to dig at the moment, I would take what I can get.
- It would be a waste of time and money

Ready made meals

- jeg har brugt mere energi på at få fat i den.
- i would find it easier to get rid of the one that i just bought because it's filled with bad luck :)
- Because they cost the same (and unlike the theatre-ticket, I have "only" spent couple of hours getting the meal, so it doesn't matter which meal I'll chose).
- It would be the same.
- Der er nok ikke forskel i smagen :)

Experiment STB

The comments are organised in such a manner that the comments for the computer company are presented first, followed by the rock/theatre ticket, the dig at Shady Creek and last the ready made meal.

Computer Company

- If it is clear thet my company could not make it better I see no reason for wasting time on it.
- Hvis de andres alligevel er bedre, virker det som spild af tid.
- The problem is already solve more effectivly så the rest of the time would be a waste.
- There is no reason to waste time developing the system if its useless.

- Just do some nice marketing. Both Coca Cola and Apple seem to do well like this.

Rock or Theatre Ticket

- Difficult to answer since I dont know to what degree I prefer event 2. But generally I would always go for the event I liked no matter the "cost" or "waste".
- I would take event 1 because I have spent 50 hours working for it.
- Very tough question. It comes down to how much more I like event2. But since I chose to work 50 hours for ticket1, the difference can´t be too big.
- I think 50 hours give me more things than 5 hours. So the event 1 might be more interesting.
- I like event 2 the most. Men svært at sætte sig ind i når man ikke direktet har følelsen af at have brugt 50 timer på noget man alligevel ikke kan komme til.
- Although having spent 10 times the time on event 1 the better experience of event 2 is worth the loss.
- If i expect to like event 2 the most, then I would choose event2. (and give the other one away).
- I worked more to obtain the ticket for event 1 so I want to compensate my 50 hours of work.
- I want to go to event 2 the most. And I might sell ticket 1, which must be worth more.
- Since I worked 10 times harder to get ticket 1, it must be more worthy, and I don´t know which event is the best yet, even though I preferred the 2nd event.
- Since I expect to like this event more I´ll attend this (and give the other ticket to a friend).
- I want to use my time on what I find meaningful.
- Although I have spent a conciderable amount of time earning the ticket to event 1, I think I would go to the event most enjoyable for me, not worrying about wasted time.
- Preferable event. No choice really.

Dig at Shady Creek

- Should have given up earlier if those 10 days cant be spared. Could be 60 more days before copper is found again and the amount might be even smaller.
- You´ve already spent 60 days on the damn dig why not continue for 10 more? If not you´d have wasted 60 days whereas you´d waste 70 days but gain copper if you decide to keep digging.
- I think it is like a reward for the 60 past days of work and who knows we might find more.
- I am stubborn.
- Since I want to give my workers some kind of success, I´ll let them dig up some copper, so they wont feel we have wasted 60 days. Had it been me alone I would have taken the 2nd option.
- We´ve come so far so lets get what we can.
- Compared to 60 days 10 days is not that much. So digging more will prevent that the 60 are completely wasted.

- Already went through the trouble. so the 60 days wont have been wasted on nothing.
- I think “a little” is better than nothing, WHEN you already have used time on the project.
- But my answer may not be correct as I have no knowledge of the cost/benefit relationship. Would make a SWOT analysis before deciding.

Ready made meals

- If I had to disgard my friend´s meal, it would feel like the trip to get his meal was wasted.
- They are the same.
- Because my own meal is the most fresh one.
- Its the same.
- It is just food. Might be that one is fresher than the other, but I do not really care.

DVD Content

1. Digital report in pdf
2. Readme guide on how to run the Visual Basic programs
3. Visual Basic program utilised in Experiment ST1, FM2, FM5 and DF1
4. Visual Basic program utilised in Experiment ST2 and DF2
5. Visual Basic program utilised in Experiment FM1 and FM3