Do we make better decisions on the Internet?

An experimental analysis of online versus offline decision-making

Linda Smolka

Erasmus Universiteit Rotterdam, The Netherlands

10-08-2016

Supervisor: Paul van Bruggen

Student number: 431333ls

E-mail: linda.smolka@student.eur.nl

E-Mail private: lk.smolka@gmail.com

Abstract

This paper focuses on whether there are differences in decision performance between decisions made online on a laptop versus offline on paper. Decision performance was measured by decision speed and decision quality. The latter was defined as rationality measured by first-order stochastic dominance (FOSD). The fewer violations of FOSD, the more rational the decision and the higher the decision quality. I conducted an experiment with two treatment groups, online and offline, consisting of students from the Goethe University Frankfurt am Main, Germany (GUF). In both treatments, subjects filled in a survey including questions on demographics, area of study, choice tasks on FOSD, follow-up questions on attitude towards the medium and previous knowledge of FOSD. The results showed that there are no significant differences between media both with regard to speed and rationality. The relations between rationality and area of study and between rationality and previous knowledge of FOSD are significant. Moreover, the relations between speed and gender and between speed and medium familiarity when writing were significant. To summarize, there is not enough evidence to conclude that the medium itself has an impact on rationality or speed. Therefore, researchers and practitioners can rather focus on confounding factors like self-selection bias when moving services and surveys online.

Keywords: medium, the Internet, online, offline decision-making, rationality, speed, first-order stochastic dominance

Table of Contents

List of Tables

List of Figures

1. Introduction

There are many different areas in which we now make decisions online instead of offline. One example is the financial sector as can be seen by the increase of financial technology, so called "FinTech", startups. These businesses offer online financial services, such as e-wallets and online banking. Other examples are businesses which collect information to make inferences about consumer behavior or polls conducted in politics. Recently, new tools such as SurveyMonkey and Google Forms made it easy to conduct surveys online which provide the ability to target a specific audience, to have a generally large and global reach and to save time and money (Evans & Mathur, 2005).

Literature in the field of behavioral economics demonstrates that the decision-making process can be easily influenced, e.g. by varying the wording or procedure of a choice task. According to papers in survey research methodology, even the medium itself can influence decision performance. Barber and Odean (2002) find that online investors trade more actively while they also trade more speculatively and less profitably as compared to offline trading. There is not much research on the reasons for the difference in online and offline trading. Both practitioners and researchers should know whether switching from offline to online service provision or survey implementation affects how rationally a customer or participant makes a decision. This leads to the following research question:

Does decision performance as measured by rationality and speed differ between decisions made online (on a laptop) versus offline (on paper)?

In the second section of this study, I provide literature in the area of behavioral economics to show that decision performance can be easily influenced through several factors in general and through the Internet in particular. Since I define decision performance as speed and rationality, I show how to measure rationality. Furthermore, I analyze the way the medium can affect decision performance as part of survey research methodology to derive my hypotheses. In the third section, I develop an experiment to test for differences in decision performance between online and offline decisions. The fourth section is a presentation of the results, which I discuss in the fifth section. In section six, I draw my conclusions from this analysis.

2. Theoretical Background

2.1. Decision-Making Influences

There are behavioral economists who hold the view that preferences are constructed during the decision-making process (Kahneman & Tversky, 1979; Payne, Bettman, Schkade, Schwarz, & Gregory, 1999; Slovic, 1995; Tversky & Thaler, 1990). During this cognitive process, decision-makers are easily influenced by *heuristics* and *biases*. The former are "rules of thumb, or mental shortcuts, the human brain uses to quickly solve complex problems" (Fuller, 1998, p.10). The latter are thinking patterns that lead to erroneous judgments (Tversky & Kahneman, 1975). This reveals decision-makers *bounded rational* behavior and that they can be easily influenced. Methods to alter someone's decision can be *framing* the decision problem in different ways (Tversky & Kahneman, 1981) and varying the procedure (Tversky & Thaler, 1990) or description (Tversky & Kahneman, 1986) of a decision task. Moreover, one can adjust the number of options, change the social setting (Johnson et al., 2012), change the default and give feedback (Thaler, Sunstein, & Balz, 2014). The decision-maker does not have to be aware of these influences in order for them to be effective (Johnson & Goldstein, 2003; Wansink, 2012). Furthermore, the "factors that lead us to make a mindless suboptimal or unhealthy choice can often be reversed to help us make a mindless better choice" (Johnson et al., 2012, p.500). To adjust the context in which a decision is made – the so called *choice architecture* – in a way that it encourages a certain decision is called *nudging*. Meanwhile, it does not deprive the decision-maker of his/her freedom to deviate from the encouraged decision (Thaler & Sunstein, 2008).

When analyzing the influences of decision-making online versus offline in particular, one can observe that decisions are easily influenced. They can be influenced by website trust for instance, which in turn can be easily shaken by errors in information provided on a website (Bart, Shankar, Sultan, & Urban, 2005). Moreover, decisions can be influenced by trust in a medium. Reeves and Nass (1996) found that people treat new technology like real people and as such they treat technology as objects of trust. There are studies which analyzed the trust in a website rather than general trust in the Internet as a medium (e.g. Corritore, Kracher, & Wiedenbeck, 2003). This study, however, focused on the Internet as a medium whose influence on decision performance cannot

be altered (except if one refrained from using the Internet) rather than factors that can be changed by the online service provider or online survey conductor.

In order to detect which factors could influence decision performance (see section 2.3. Online Versus Offline Decisions), I first define decision performance. Decision performance is high when subjects arrive at a good decision in a timely manner. Decision performance is therefore measured by decision speed and decision quality. The definition and assessment of decision quality is described in the following section.

2.2. Measuring Rationality

When I mention decision quality in this study, it means that the more rational the decision-maker, the higher the decision quality. There are two reasons why a decision which is rational is considered to be a "good" decision. Firstly, expected utility theory (EU) is generally accepted as a normative model of decision theory. Secondly, rational decisions are considered a *social norm*, meaning that it is socially desirable to behave rationally:

> "Man has rationality as a norm, as a second-order disposition of the following kind: once one becomes aware that one has fallen into irrationality, one will tend to adjust one's belief, attitudes, and actions such as to make them more rational. [...] Man is a rational animal in the sense that man has rationality as a norm." (Føllesdal, 1982, p.316).

Irrational decision-making can lead to poor choices. If someone is presented a lottery, for instance, he/she may decide to reject a profitable investment opportunity due to *loss aversion* which is the overweighing of negative outcomes (Kahneman, Knetsch, & Thaler, 1991). Another example of irrational behavior is when someone is subject to *status quo bias* meaning that the person chooses not to change a situation or previous decision even when the cost of doing so is small and the importance of the decision is large (Kahneman, Knetsch, & Thaler, 1991). Since I used rationality as indicator of decision quality, I needed a way to assess the level of rationality. First-order stochastic dominance (FOSD) tasks are used as an objective and widely accepted measure of rationality. Since EU satisfies *stochastic dominance,* EU predicts that none of the

subjects violate FOSD. However, when faced with a choice task of two lotteries, people often tended to violate FOSD (Birnbaum, 2005a), revealing *bounded rational* behavior. An example of such a choice task described in the following.

A person decides from which urn he/she wants to draw a marble at random, if the color of the marble determined the amount of money the subject receives.

Table 1. Example of a first-order stochastic dominance task (Birnbaum, 2005a).

Urn A	Urn B
90 red marbles to win €96	85 green marbles to win €96
05 blue marbles to win €14	05 black marbles to win €90
05 white marbles to win €12	10 yellow marbles to win €12

In general, some gamble A stochastically dominates a gamble B under two conditions: 1) the probability of gaining an outcome x or higher in gamble A is greater than or equal to the probability to gain the same outcome x or higher in gamble B for all outcomes x, and 2) this probability is strictly greater for at least one outcome x. In the example above, the probability to win €96 or more is 0.9 in gamble A and 0.85 in gamble B. The probability to win €90 or more is 0.9 in both gamble A and gamble B. The probability to win €14 or more is 0.95 in gamble A and 0.9 in gamble B. The probability to win €12 or more is the same for both gambles. Therefore, both conditions hold and gamble A stochastically dominates gamble B (Birnbaum, 2005a, p.263f.).

In the subsequent section, I review survey research methodology to identify factors which could influence decision speed and rationality in different ways for online and offline decisions. I thereby distinguish between confounding factors that are controlled for and the factors of interest which are the ones inherent to the medium.

2.3. Online Versus Offline Decisions

2.3.1. Confounding Factors

Because there was a lot of progress in survey techniques and technologies, an increasing number of surveys are conducted online instead of offline. Some of the major

strengths of online surveys are flexibility, speed and timeliness, low administration cost, control of the answer order and a required completion of answers. One downside may be that instructions have to be as understandable as possible since online surveys are self-administered (Evans & Mathur, 2005) when they are not conducted in a lab. This is not an inherent characteristic of the medium since surveys can be self-administered or non-self-administered in both online and offline surveys.

Another downside of online surveys is that there may be a self-selection bias since people using the Internet are not representative of the general population. Barber and Odean (2002) for instance explained the difference between online and offline traders with a self-selection bias since overconfident investors are more likely to switch to online trading and also are more likely to trade more speculatively. That certain people tend to use online or offline surveys is inherent to those people but not an inherent characteristic to the medium.

There may also be *test mode effects*, meaning that the flexibility to skip, review and change answers on a laptop can change decision performance (Spray, Ackerman, Reckase, & Carlson, 1989). These are characteristics which can be more easily controlled for in an online survey. Technically, however, the ability to skip, review and change answers can be implemented both online and offline. Therefore, *test mode effects* as well as self-administration and self-selection bias should all be controlled for in an experiment which analyzes differences in decision speed and rationality caused by medium.

2.3.2. Factors of Interest

Dillon (1992), Noyes and Garland (2008) provided an overview of experimental comparisons between online and offline surveys. The tasks observed in the experiments of survey research methodology were mostly reading comprehension. In their overview of different studies, Noyes and Garland (2008) differentiated between variables observed before and after 1992. The variables observed in studies before 1992 were predominantly reading speed, accuracy and comprehension. Mayes et al. (2001) for instance found reading on a screen to be significantly slower. With regard to decision quality, Chen, Cheng, Chang, Zheng and Huang (2014) found that for strongly

demanding comprehension tasks most studies did not find significant differences between paper and computer. For less demanding comprehension tasks, however, subjects had significantly higher scores using paper which confirms former studies (Mangen, Walgermo, & Brønnick, 2013; Wästlund, Reinikka, Norlander, & Archer, 2005 as cited by Chen et al., 2014).

Since FOSD choice tasks are rather demanding, this would suggest that there is no difference between media in this study. With the rationality tasks, however, there is one big difference compared to reading comprehension tasks: there is risk involved which is strongly affected by trust (Jarvenpaa et al., 2000 as cited by Shankar, Urban, & Sultan, 2002). Trust in turn is closely related to security (Grandison & Sloman, 2000), I therefore expect the following concerns to weigh more heavily than with reading comprehension tasks. Subjects may differ in their privacy and security (Evans & Mathur, 2005; Buchanan & Hvizdak, 2009) or ethical concerns such as "consent […], anonymity, confidentiality, and autonomy" (Buchanan & Hvizdak, 2009) when comparing different mediums. People might feel that decisions can be easier traced back to them in offline surveys since they are handed-in personally. On the other hand, participants might feel concerned that online data can be illegally acquired by third parties. Hence, there may be differences between online and offline surveys in both decision speed and rationality caused by privacy, security and ethical concerns.

After 1992, researchers additionally started observing more refined and task-specific variables. One such variable is *cognitive workload* (and memory retrieval which is not relevant for this study). *Cognitive workload* is measured by assessment of the participant's mental demand, physical demand, temporal demand, performance, effort and frustration (Hart & Staveland, 1988). It can reveal differences in processing difficult tasks which can explain performance variations between decisions made on screen and on paper (Noyes & Garland, 2008). Although the majority of the more recent studies which use more advanced technologies revealed no significant differences in *cognitive workload* (Mason et al., 2001; Mayes, Sims, & Koonce, 2001; Noyes & Garland, 2003; Bodmann & Robinson 2004; Garland & Noyes, 2004 as cited by Noyes & Garland, 2008), there are very recent studies which did reveal differences in cognition due to device-related navigation and visual fatigue caused by emitting light, and individual differences in cognition (e.g. Chen et al., 2014). Differences in cognition could affect

decision performance as a whole. A factor which could only influence differences in speed could be the ability to scroll on a laptop screen which can lead to spatial instability (Mangen et al., 2013). The ability to scroll is an inherent characteristic of the medium Internet and is therefore maintained in the study.

Another more refined variable which had an effect on decision performance in reading comprehension tasks altogether is *computer familiarity* (Yu, 2010) which can be assessed through a computer familiarity questionnaire (CFQ) (Chen et al., 2014). The factors which influence *computer familiarity* are "access, attitude, experience or use, and experience with related technology" (Eignor, Taylor, Kirsch, & Jamieson, 1998). According to Eignor et al. (1998), only attitude and experience with related technology appeared as distinct factors, while access and experience or use correlated with attitude and experience with related technology. Also Zheng, Cheng, Xu, Chen and Huang (2015) found that attitude correlated with frequency of use and experience. I assumed that all students have experience with related technology. Therefore, I was only interested in differences in attitude towards media. Figure 1 provides an overview of the all the previously mentioned factors which may cause differences in decision performance between different media.

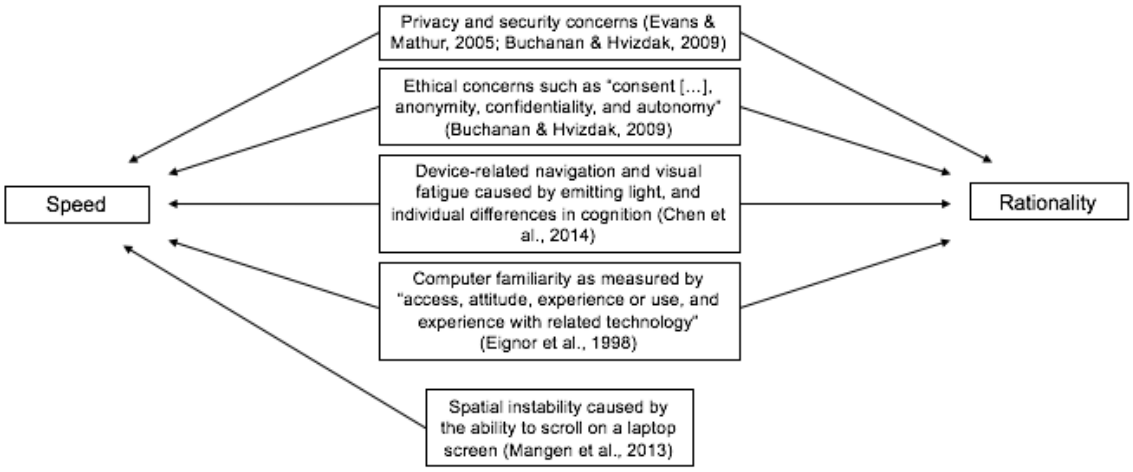


Figure 1. Factors influencing decision performance.

This leads to the following hypotheses:

H_1. *Subjects make faster decisions offline (on paper) than online (on the laptop).*

H_2. *Subjects violate FOSD less often offline (on paper) than online (on the laptop).*

When comparing different media, it is important to assure that the surveys are as similar as possible (Noyes & Garland, 2008). In the study of Walt, Atwood, & Mann (2008) for instance – which tests for differences between computer and paper – the survey changed in the way it is displayed as it is transferred from paper to online survey. As a result, they violated similarity of surveys which could have led to *framing effects* (Bettman & Kakkar, 1977; DeSanctis, 1984; Vessey, 1991 as cited by Speier, Vessey, & Valacich, 2003). When similarity is violated, the difference between media can not be attributed solely to the medium. The method I used to establish equivalence between the survey forms is part of the experimental design in the following section. After the description of the experimental design and participants, I describe how I implemented the experiment.

3. Experiment

3.1. Participants and Design

The experiment consisted of two treatments and was conducted with 88 students from the GUF. Students were equally distributed among treatments (see Appendix A.1.). The treatment was the use of different media: the Internet (experimenter laptop) or paper. In both treatments, the task was to fill in a survey. The content of the survey consisted of questions on gender, age, nationality and area of study; seven choice tasks on FOSD; follow-up questions on which medium the subject prefers when reading and writing; and a question about previous knowledge of the FOSD. I also asked participants to provide contact details in order to be able to conduct a random-lottery to give subjects monetary incentive to participate in the experiment (Wakker, 2007). The monetary incentive was for the subject to participate in a random-lottery. The incentive provided by this lottery was not very high since the probability to win money at all was 1/88 and the probability to win the maximum amount of €9,60 was even lower. However, I wanted to provide an incentive compatible with the task, so I had to connect the lottery prize to the decision outcome. Providing an incentive of €96 was not possible due to the budget constraint. Since I provided monetary incentives, I had to ask participants for their contact details. Therefore, I could not provide anonymity.

The FOSD example and tasks used in the experiment of this study are the same as in study 1 in Birnbaum's experiment (Birnbaum, 2005b). They are specifically designed to trigger violations of FOSD (Birnbaum, 2005a). I chose this instead of tasks that do not trigger violations of FOSD to generate more variation between tasks and treatments. The different FOSD choice tasks are not altered with filler choices as in Birnbaum's (2005a) study. This could lead subjects to answer the questions according to a specific pattern or strategy. This affects both treatments and therefore it should not negatively affect the results comparing the relative differences between media. The graphical representation of the FOSD choice problem can influence the decision (Dertwinkel-Kalt & Köster, 2015). Since the graphical representation is the same in both experimental treatments, and since I am interested in the relative differences between media, this does not affect the results of the experiment. Regarding the factors which could explain potential differences between media mentioned in section 2.3., I only included the factor medium familiarity by asking which medium a subject prefers when reading and writing. The reason for this is that I had limited resources and it was not be feasible to conduct a survey which lasts longer than about ten minutes without being able to pay each subject for their participation.

The online survey was created using Qualtrics (see Appendix A.2.). In order to ensure equivalence between treatments, I edited the print version of the survey in Photoshop (since downloading the survey from Qualtrics completely changed the layout). The FOSD choice tasks were randomized to control for ordering effects. This could be implemented automatically in the online survey form. However, to randomize questions in the print version, I had to edit every questionnaire individually. I used an online randomization tool[1] to generate random orders of the FOSD choice tasks which I then implemented in Photoshop. The surveys for both treatments had the same font type, font color, proportions and layout. The page or screen size (and therefore the font size) differed since the screen size was 13" and the paper had an A4 format. The page or screen size is an inherent characteristic of the medium and was therefore maintained. Since *test mode effects* are not inherent to the medium, I provided subjects in both treatments with the ability to review and change answers in both media and the inability

[1] https://www.randomizer.org/

to skip questions in either medium to exclude these effects as factors causing differences between media.

Optimally, the subjects should be able to both win and lose money in the gambles of the FOSD choice tasks to make the experiment as realistic as possible. This, however, is not practically feasible since I could not ask the students to lose their own money in the experiment, while giving them a budget in the beginning of the experiment could lead to the *house money effect*. The *house money effect* suggests that people value their own money more than money they are given by the experimenter which could lead to riskier decisions (Ackert, Charupat, Church, & Deaves, 2006). In order to obtain a homogenous group of subjects, the experiment was conducted on campus of the GUF. Thereby, I also controlled for speed since people who filled in surveys at home may take breaks from filling in the questionnaire. The downside was that subjects might have felt time pressure as the experimenter was present. Since I was interested in relative differences in speed between media and the experimenter was present in both treatments, it did not affect the validity of the experiment. Moreover, experimenter presence ensured that treatments were equal with regard to self-administration of the survey.

3.2. Procedure

First, I randomly approached potential subjects for the experiment on campus to obtain a randomized sample. I asked them to participate in an experiment as part of my Master thesis at Erasmus University Rotterdam. I told them that while they participated in the experiment, they also had the chance to participate in a lottery through which they had the chance to win up to €9,60. If they agreed to participate, I randomized them into one of the two treatment groups to control for self-selection bias. I let each subject draw one folded piece of paper from a bag which was filled with equal amounts of papers with the letters “L” (for laptop) and “P” (for paper) at the beginning of the experiment. The subjects did not put the piece of paper back into the bag so that I ended up with 44 subjects per treatment. Participants received the experimenter laptop with the online survey or the paper-based survey accordingly. The subjects filled in the survey while I, the experimenter, inconspicuously kept track of the time the subjects needed to fill in

the survey in the offline treatment. In the online treatment, the time was automatically stopped by Qualtrics. After the subjects completed the survey, I collected the laptop or paper and thanked the subject for their participation. In the offline treatment, I wrote down the time a subject took on the first page of the survey he/she filled in.

After I conducted the experiment, I played out the lottery used as monetary incentive to participate with the help of the online randomization tool I also used to prepare the paper survey. First, I gave every participant an ID composed of a number between one and 77. I had to exclude 11 subjects because they did not want to indicate their names and/ or email addresses for the lottery. Some people verbalized that they were concerned about either privacy or being spammed with advertisement. Others mentioned that they did not want or need an incentive to participate. I generated a random number between one and 77 to identify the winner of the lottery which was the subject with ID number 34. Secondly, I generated a random number between one and seven to identify the gamble that would be played out. This turned out to be FOSD choice task 1 (see Appendix A.2.). The subject with ID number 34 chose Lottery A for this task and a marble was drawn at random which turned out to be red. Therefore, the subject won (€96 / 10 =) €9.60. She was notified via email that she won the lottery and the prize money was transferred to her bank account.

4. Results

There were 88 independent observations at the level of the individual. Of those 88 observations, 44 observations were collected in the online treatment and 44 observations in the offline treatment. I dropped one observation in the offline treatment due to a mistake in the survey form. Thus, I ended up with 87 observations overall for the data analysis.

4.1. Variable Definitions

There were 17 variables in the analysis presented in Table 2 on the next page.

Table 2. Variable Definitions.

	Variable Name	**Variable Type**	**Variable Definition**
1	online	dummy variable	takes value 1 if the survey is filled in online and 0 if it is filled in offline
2	rationality	categorical variable	sum of violations of FOSD throughout all seven choice tasks, ranges from 0-7 with 0 being the most rational and 7 being the least rational
3	speed	continuous variable	speed to fill in the survey in seconds
4	nationality	categorical variable	1=Germany, 2=Rest of Europe, 3=Asia and 4=South America
5	gender	dummy variable	takes value 1 for female and value 0 for male participants
6	age	continuous variable	age of the participant in years
7	area_of_study	categorical variable	1=Economics and Business, 2=Psychology, 3=Social Sciences, 4=Human Sciences, 5=Computer Sciences, 6=Law, 7=Natural sciences, 8=Teaching, 9=Engineering
8	med_fam_read	dummy variable	takes value 1 if participant prefers paper to laptop when reading, and value 0 if he/she prefers laptop to paper when reading
9	med_fam_write	dummy variable	takes value 1 if participant prefers paper to laptop when writing, and value 0 if he/she prefers laptop to paper when writing
10	FOSD_knowledge	dummy variable	indicates whether a subject has previous knowledge of FOSD where the value 1 stands for "yes" and the value 0 for "no"
11-17	FOSD_task_1, ..., FOSD_task_7	dummy variable	takes value 1 if participant violated FOSD, and 0 if he/she did not

In the following section, I describe the data for each variable in the same order they are represented in the table.

4.2. Variable Descriptions

When looking at the variable rationality, one can observe that 33.72% of the participants did not violate FOSD. Thus, they made decisions consistent with fully rational behavior while 66.28% revealed *bounded rational* behavior by violating FOSD up to seven times. The frequencies of violations are displayed in Table 3 below.

Table 3. Percentages of FOSD violations.

0	1	2	3	4	5	6	7
33.72%	10.47%	15.12%	10.47%	9.30%	6.98%	4.65%	9.30%

Table 4 below summarizes in which of the tasks FOSD was most frequently violated. One can see that FOSD was violated most frequently in FOSD task 1, and violated least in FOSD task 7. In fact, the percentage of violations decreased between all FOSD tasks with task 4 and 5 having an equal percentage of violations.

Table 4. Frequency of violation in the different FOSD choice tasks.

task 1	task 2	task 3	task 4	task 5	task 6	task 7
50%	38.37%	32.56%	31.40%	31.40%	27.06%	25.58%

The speed in which participants completed the survey was 424 seconds, or around 7 minutes, on average. For the online treatment, the average speed was 433 seconds. The minimum was 178 seconds and the maximum was 650 seconds. For the offline treatment, the average speed was slightly lower with 414 seconds. The minimum was below the minimum in the offline treatment with 140 seconds and the maximum was above the online treatment with 703 seconds. There were more extreme values for the offline treatment than in the online treatment. The average of the offline treatment, however, was below the average of the online treatment. The descriptive statistics for the variable speed are summarized in Table 5 below.

Table 5. Descriptive statistics of the variable speed.

	N	Min (seconds)	Average (seconds)	Max (seconds)
speed overall	87	140	424	703
speed, online	44	178	433	650
speed, offline	43	140	414	703

With regard to nationality, the majority of participants, namely 76 participants, were German. The exact distribution of participants with regard to nationality is illustrated by the pie chart below.

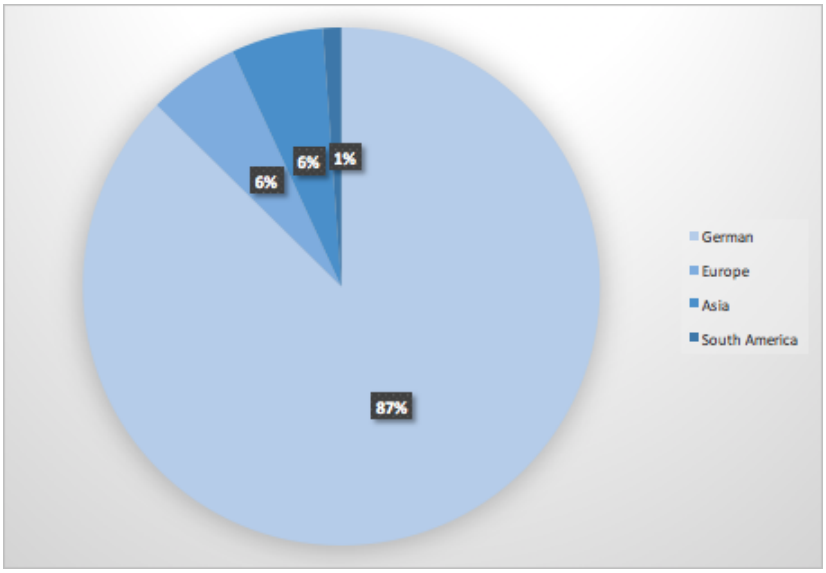


Figure 2. Descriptive statistics of the variable nationality.

In the following, there is another pie chart showing the number of subjects from different areas of study. The largest group, consisting of 29 participants, had a background in Economics and Business. Unlike for the other variables, there were only 86 observations instead of 87 because one subject did not fill in his/her area of study.

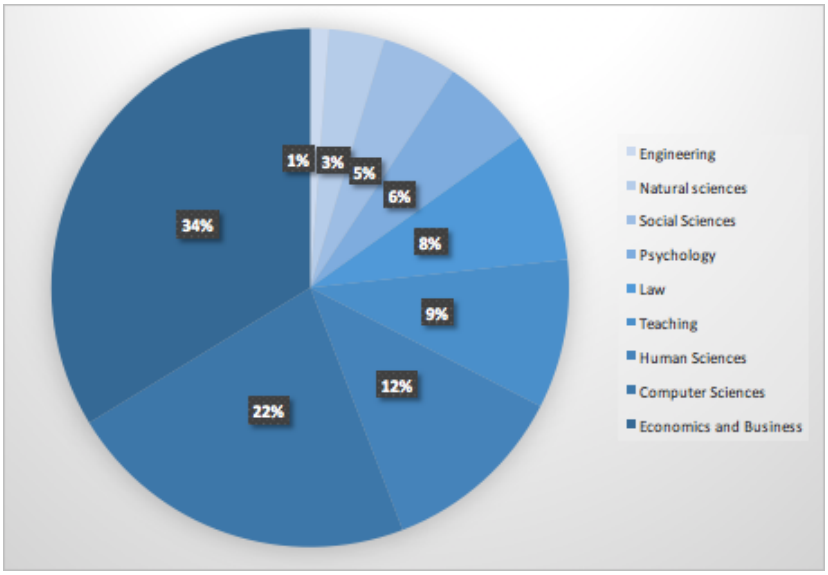


Figure 3. Descriptive statistics of the variable area of study.

The distribution of genders was approximately equal with 53.49% being male and 46.51% being female. Participant's age was 23 years on average with a minimum age of 18 years and a maximum age of 36 years. The different ages are displayed in the histogram below.

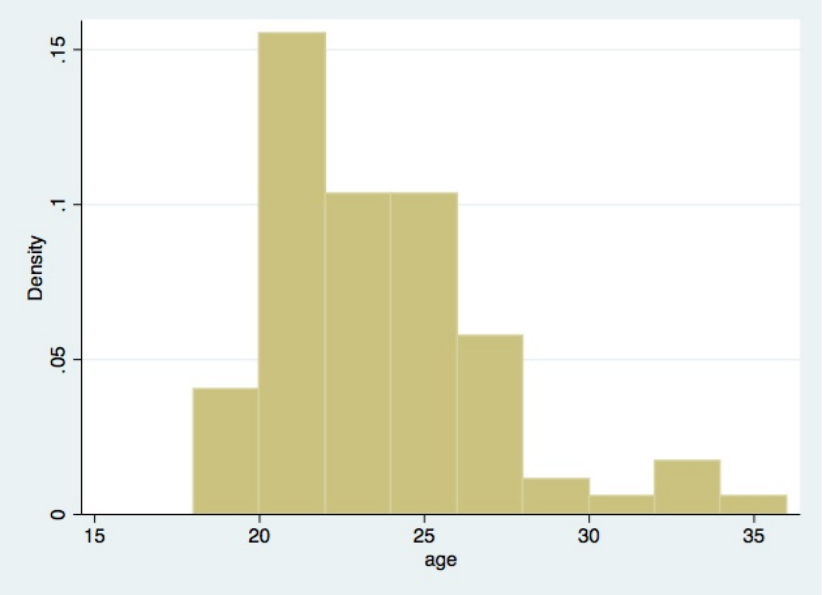


Figure 4. Descriptive statistics of the variable age.

Subjects generally preferred the medium paper to laptop both when reading and writing. There is an overview of the medium preferences in Table 6 below.

Table 6. Medium preferences for laptop and paper.

	laptop preferred	paper preferred
reading	18.60%	81.40%
writing	46.51%	53.49%

The majority of people did not have previous knowledge of FOSD, namely 83.72%, while 16.28% said that they would be able to explain what FOSD is.

4.3. Data Analysis

There were two variables of interest, rationality and speed, for which I conduct a separate data analysis in Stata. The variable rationality was a categorical variable in an ordinal scale which means that I could not use a parametric test for this variable. The variable speed was continuous but did not satisfy several of the conditions to perform a parametric statistical test (see Appendix A.3.). This is why I used a non-parametric statistical test for both variables. In order to be able to conduct a non-parametric test, observations must be independent. This was the case for the whole data set due to the between-subject design.

I chose to use the Mann-Whitney U test to analyze relation between the variable speed and any other binary variable. This test is a rank-sum test used to analyze if two independent samples are from populations with the same distribution. The Mann-Whitney U test indicated that the time needed to make a decision was greater in the offline treatment than in the online treatment ($z=-0.938$, $p=0.3482$). The subjects who preferred laptop when reading were faster than the subjects who preferred paper when reading ($z=0.110$, $p=0.9128$). The subjects who had previous knowledge of FOSD were faster than the subjects who did not ($z=0.231$, $p=0.8173$). However, for the last three results, I do not reject the null hypothesis that there are no significant differences between the distributions of the populations underlying the samples. Moreover, the results showed that female subjects were faster than male subjects ($z=1.781$, $p=0.0748$). The subjects who preferred laptop when writing were slower than the subjects who preferred paper when writing ($z=1.716$, $p=0.0861$). Both of these results were significant at a 10%-level. The results from the Mann-Whitney U test are summarized in

Table 7 below. It shows us the sample size N, the observed ranked sums and the rank sum that would be expected if the null hypothesis (that there is no significant difference) was not rejected.

Table 7. Mann-Whitney U test for speed and other binary variables.

Variable	**Value**	**N**	**Rank Sum**	**Expected**
online	1	43	1781.5	1892
	0	44	2046.5	1936
gender	0	46	2233.5	2024
	1	41	1594.5	1804
med_fam_read	0	16	714	704
	1	71	3114	3124
med_fam_write	0	40	1961.5	1760
	1	47	1866.5	2068
FOSD_knowledge	0	73	3232	3212
	1	14	596	616

To test for significance of the relations between the variable speed and non-binary variables, I chose a Fisher exact test which is a non-parametric two sample test. It is good with small samples and small frequencies of values. To be able to use it, I made a binary variable out of the continuous variable speed. The variable speed_high takes value 1 if speed is above median and value 0 if speed is equal to or below median. The descriptive statistics for the new variable speed_high is in Table 8 below.

Table 8. Descriptive statistics of the variable speed_high.

Speed_high	N	Min (seconds)	Average (seconds)	Max (seconds)
0	44	140	327.05	407
1	43	409	534.88	925

I do not reject the null hypothesis that there is no significant difference in speed (speed_high) for different nationalities (two-sided, p=0.638, obs=87), for different ages

(two-sided, p=0.357, obs=87), as well as for different areas of study (two-sided, p=0.101, obs=86).

For the variable rationality, I could not use the Mann-Whitney U test, since there were many ties among observations. For the same reasons mentioned for the variable speed, I used the Fisher exact test for the variable rationality. Fisher exact can only be used for categorical variables, so I made a binary variable from the continuous variable age. The variable age_high takes value 1 if age is above median and value 0 if age is equal to or below median. The descriptive statistics for the new variable are displayed in Table 9 below.

Table 9. Descriptive statistics of the variable age_high.

Age_high	N	Min (years)	Average (years)	Max (years)
0	52	18	21	23
1	35	24	26	36

I do not reject the null hypothesis that there is no significant difference in rationality for the different media (two-sided, p=0.644, obs=87), for different nationalities (two-sided, p=0.354, obs=87), for different genders (two-sided, p=0.149, obs=87) and for different ages (two-sided, p=0.196, obs=87). Moreover, I do not reject the null hypothesis that there is no significant difference in rationality for different preferences for media when reading (two-sided, p=0.309, obs=87), and writing (two-sided, p=0.945, obs=87). I do reject the null hypothesis that there is no significant difference in rationality for subjects with, and without previous knowledge of FOSD (two-sided, p=0.011, obs=87). The result was significant at a 5%-level. Furthermore, I do reject the null hypothesis that there is no significant difference in rationality for different areas of study (two-sided, p=0.006, obs=86). The result was significant at a 1%-level.

From the data, I conclude that there is not enough evidence to reject the null hypothesis, that there is no significant difference in speed for offline and online decisions. The reason for this could be either that the results are not consistent with H_1 (Subjects make faster decisions offline than online), or because of low test power. Moreover, I do not reject the null hypothesis, that there is no significant difference in

FOSD violations for offline and online decisions. Here as well this could be due to low test power or due to the results not being consistent with H_2 (Subjects violate FOSD less often offline than online).

5. Discussion

5.1. General Discussion

Contrary to the hypotheses of this study, the different media do not have a significant effect on rationality or speed. This study, however, further looks into relations between variables that could explain possible differences between media like area of study, previous knowledge of FOSD, gender and medium preference. The results show a significant relation of area of study and rationality which could be explained by means of *social norms*. Rational decisions are considered to be a *social norm*, however, it may be the case that in economics this *social norm* is more prevalent than in other areas of study. Moreover, I find that subjects with previous knowledge of FOSD decide more rationally. A reason for this as well could be *social norms*. As mention before, a person who knows what rational behavior is, would not want to violate rationality (Føllesdal, 1982). Another reason for this could be that subjects who are familiar with these kinds of choice tasks use this previous knowledge to "solve" these choice tasks because they consider the rational decision to be the "right" one without considering if they would actually prefer the other option. This could also mean that subjects always prefer the more rational option but are unable to identify it without practice. Meanwhile, area of study and previous knowledge of FOSD themselves do not have a significant relation. This suggests that the area of study does not have a significant effect on rationality because people with certain backgrounds tend to be more or less familiar with FOSD.

Furthermore, the results show that female subjects are faster than male subjects. While decision quality or rationality do not significantly differ for the genders, the overall decision performance is still higher for female subjects. This could be caused by many different factors. Female subjects could be more driven and therefore strive to perform better. They could also be more motivated to support the research the best way

they can. However, differences could also be caused by differences in cognition between male and female subjects. These are just a few possible explanations. In order to be certain what causes the differences, however, one would have to do further research. When looking at writing preferences, the subjects who prefer laptop were slower than the subjects who prefer paper. However, there was not enough evidence for a significant difference when looking at reading preferences. It is surprising that the media did not cause significant differences in speed while medium preference did. This could be an objective of further research by designing an experiment to test the influence of inherent preferences on decision-making.

5.2. Limitations

A limitation of the study is that the different FOSD choice tasks are not altered with filler choices which might lead subjects to answer the questions according to a specific choice pattern. This is the reason why the results from all seven choice tasks are not independent and why I combine them into one variable (rationality). This effect might be reinforced by the option to go back and forth and review previous and subsequent questions. Although there is no explicit time constraint, this effect is potentially also reinforced by time pressure posed through the experimenter's presence (Ordonez & Benson, 1997). It probably leads people to either always go for the gamble with 90, 5, 5 marbles or the other option (see Appendix A.2.).

Another limitation is that subjects could have tended to review and change answers more frequently in one of the two treatments. Subjects could initially make less rational decisions in either treatment but then tend to correct their answers to a more rational one by going back and forth thus leading to more similar results of the treatments. If that is the case, however, than that is an inherent characteristic of the medium Internet and is exactly what I was trying to observe. If the effect of going back and forth in one treatment more than in the other is not caused by the medium itself but rather by experimenter presence, however, that is a factor which should have been controlled for. Experimenter presence is established to provide a similar environment for both conditions and to control for both anonymity and self-selection bias. During the experiment, however, I am theoretically only able to observe test mode effects in the

offline condition and not in the online condition due to the nature of a screen being only visible by the subject and paper laying on the table visible for both the subject and the experimenter. This suggests that subjects in the online treatment might initially decide less rationally but have a stronger tendency to change answers when they discover a pattern in the answer options leading to very similar results in both treatments in the end. Another aspect of the experiment that could be improved is that the question on medium familiarity only has two answer options. Using a more differentiated scale, e.g. a five- or seven-point Likert scale, could give a more differentiated insight about subject's attitudes towards different media. This could be implemented in further analyses. More suggestions for further research follow in the subsequent section.

5.3. Future Research

With regard to methodology, I make the following suggestions for further research. I would strongly suggest to replicate this study with a bigger sample to see whether the relation between medium and decision performance is really insignificant or whether there is a lack of test power. One can also conduct an experiment using other samples like less educated people since not only educated people make use of online services or participate in online surveys. Moreover, one could test for differences between generations. The research questions may have different answers for so-called "Digital Natives" who are familiar with new technologies and other generations.

It is important to note that there is a tradeoff between providing incentives and providing anonymity. Providing incentives is important to control for self-selection bias because without monetary incentives only intrinsically motivated participants would agree to take part in the experiment. Providing anonymity would be important for two reasons. One is to control for the *Hawthorne effect*. This means that subjects' decisions are affected by their own awareness of participating in an experiment (Jones, 1992), meaning that they could try to guess what the purpose of the experiment is and decide accordingly. Another reason is that subjects could be influenced by normative mechanisms like social desirability which means that subjects feel the need to respond in ways that are socially acceptable (Carini et al., 2003, p.2f). I suggest to conduct this experiment while granting subjects anonymity. This solves the problem of subjects

feeling monitored while reviewing and changing answers in the offline treatment as described in the limitations section. As new technologies emerge, online surveys can also be accessed through smartphones or tablets if not conducted in a lab. Furthermore, online services are increasingly accessed through mobile devices. Therefore, further research should explore whether it makes a difference if online decisions are made through different devices. Moreover, one can test whether there is a difference in using the laptop of the experimenter and the laptop of the subject.

6. Conclusions

In general, one could expect there to be differences in decision performance between online and offline decisions caused by privacy, security, and ethical concerns, differences in cognition when making decisions on screen or paper and attitude towards different media. According to the results of this study, medium does not affect decision performance as measured by decision speed and rationality. Since there is not enough evidence that the medium affects decision performance, one can benefit from the many advantages of online surveys and online services like flexibility, speed and timeliness and low administration cost. Service providers who move their services online should be aware of the factors influencing purchasing decisions. Since the medium itself does not seem to influence them, they can focus on other factors, e.g. self-selection bias and try to counteract this through targeting customers. Researchers as well as practitioners use online surveys to collect data. They base important conclusions and insights on these surveys which in turn can have a huge impact on science, business, politics and other areas like medicine. Thus, it is important for them to know that decision speed and rationality are not influenced simply by their choice of medium.

References

Ackert, L. F., Charupat, N., Church, B. K., & Deaves, R. (2006). An experimental examination of the house money effect in a multi-period setting. *Experimental Economics*, *9*(1), 5-16.

Barber, B. M., & Odean, T. (2002). Online investors: do the slow die first?. *Review of Financial Studies*, *15*(2), 455-488.

Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. *Journal of marketing*, *69*(4), 133-152.

Birnbaum, M. H. (2005a). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, *31*(3), 263-287.

Birnbaum, M. H. (2005b). Decision-Making Experiment: Choices between Gambles. URL: http://psych.fullerton.edu/mbirnbaum/decisions/exp2_urns_SD.htm, Access Date: 11.07.2016

Buchanan, E. A., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of human research ethics committees. *Journal of Empirical Research on Human Research Ethics*, *4*(2), 37-48.

Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: Does mode matter?. *Research in Higher Education*, *44*(1), 1-19.

Chen, G., Cheng, W., Chang, T. W., Zheng, X., & Huang, R. (2014). A comparison of reading comprehension across paper, computer screens, and tablets: Does tablet familiarity matter?. *Journal of Computers in Education*, *1*(2-3), 213-225.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International journal of human-computer studies*, *58*(6), 737-758.

Dertwinkel-Kalt, M., & Köster, M. (2015). Violations of first-order stochastic dominance as salience effects. *Journal of Behavioral and Experimental Economics*, *59*, 42-46.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, *35*(10), 1297-1326.

Eignor, D., Taylor, C., Kirsch, I., & Jamieson, J. (1998). Development of a scale for assessing the level of computer familiarity of TOEFL examinees. *ETS Research Report Series*, *1998*(1), i-32.

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet research*, *15*(2), 195-219.

Føllesdal, D. (1982). The status of rationality assumptions in interpretation and in the explanation of action. *Dialectica*, *36*(4), 301-316.

Fuller, R. J. (1998). Behavioral finance and the sources of alpha. *Journal of Pension Plan Investing*, *2*(3), 291-293.

Grandison, T., & Sloman, M. (2000). A survey of trust in internet applications.*IEEE Communications Surveys & Tutorials*, *3*(4), 2-16.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives?. *Science*, *302*(5649), 1338-1339.

Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., ... & Wansink, B. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, *23*(2), 487-504.

Jones, S. R. (1992). Was there a Hawthorne effect?. *American Journal of sociology*, 451-468.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives*, *5*(1), 193-206.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*, 61-68.

Mayes, D. K., Sims, V. K., & Koonce, J. M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, *28*(6), 367-378.

Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: Are they equivalent?. *Ergonomics*, *51*(9), 1352-1375.

Ordonez, L., & Benson, L. (1997). Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decision Processes*, *71*(2), 121-140.

Payne, J. W., Bettman, J. R., Schkade, D. A., Schwarz, N., & Gregory, R. (1999). Measuring constructed preferences: Towards a building code. In *Elicitation of preferences* (pp. 243-275). Springer Netherlands.

Reeves, B., & Nass, C. (1996). How people treat computers, television, and new media like real people and places (pp. 19-36). *Cambridge, UK: CSLI Publications and Cambridge university press*.

Shankar, V., Urban, G. L., & Sultan, F. (2002). Online trust: a stakeholder perspective, concepts, implications, and future directions. *The Journal of strategic information systems*, *11*(3), 325-344.

Slovic, P. (1995). The construction of preference. *American psychologist*, *50*(5), 364.

Speier, C., Vessey, I., & Valacich, J. S. (2003). The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sciences*, *34*(4), 771-797.

Spray, J. A., Ackerman, T. A., Reckase, M. D., & Carlson, J. E. (1989). Effect of the medium of item presentation on examinee performance and item characteristics. *Journal of Educational measurement*, *26*(3), 261-271.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving Decisions about Health, Wealth, and Happiness. *Yale University Press*.

Thaler, R. H., Sunstein, C. R., & Balz, J. P. (2014). Choice architecture. *The behavioral foundations of public policy*.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions (pp. 81-126). *Springer Berlin Heidelberg*.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141-162). Springer Netherlands.

Tversky, A., & Thaler, R. H. (1990). Anomalies: preference reversals. *The Journal of Economic Perspectives*, *4*(2), 201-211.

Wakker, P. P. (2007). Message to referees who want to embark on yet another discussion of the random-lottery incentive system for individual choice. URL: http://people.few.eur.nl/wakker/miscella/debates/randomlinc.htm, Access Date: 27.07.2016

Walt, N., Atwood, K., & Mann, A. (2008). Does survey medium affect responses? An exploration of electronic and paper surveying in British Columbia Schools. *The Journal of Technology, Learning and Assessment*, *6*(7).

Wansink, B. (2012). Package size, portion size, serving size... market size: The unconventional case for half-size servings. *Marketing Science*, *31*, 1-54.

Yu, G. (2010). Effects of presentation mode and computer familiarity on summarization of extended texts. *Language Assessment Quarterly*, *7*(2), 119-136.

Zheng, X., Cheng, W., Xu, B., Chen, G., & Huang, R. (2015). An initial development and validation of tablet computer familiarity questionnaire. In *Emerging Issues in Smart Learning* (pp. 69-76). Springer Berlin Heidelberg.

Appendix

A.1. Power Calculation

In order to calculate the optimal number of observations required for each treatment in this study, I take the conventional values of significance with alpha = 0.05 and beta = 0.2 implying a power of 80% and assume that variance remains constant across treatments. Assuming a standard deviation of 0.5 and a 15% difference, I get a minimum desired effect size of δ = (0.5-0.35)/0.5 = 0.3. Now, the number of observations per group can be calculated using the formula

(1) $N = 2 (t \alpha/2 + t \beta)^2 * (\sigma / \delta)^2$

(2) $N = 2 (1.96 + 0.84)^2 * (0.5 / 0.3)^2$

(3) $N = 44$

There shall be 44 observations per treatment, which means 88 subjects overall. It is important to note that this number is a mere approximation since it has been derived using the rule of thumb that assumes that a student t-test is being used which as discussed in the following section, is not the case.

A.2. Survey Questionnaire (Example)

The FOSD choice tasks in the example survey below appear in ascending order.

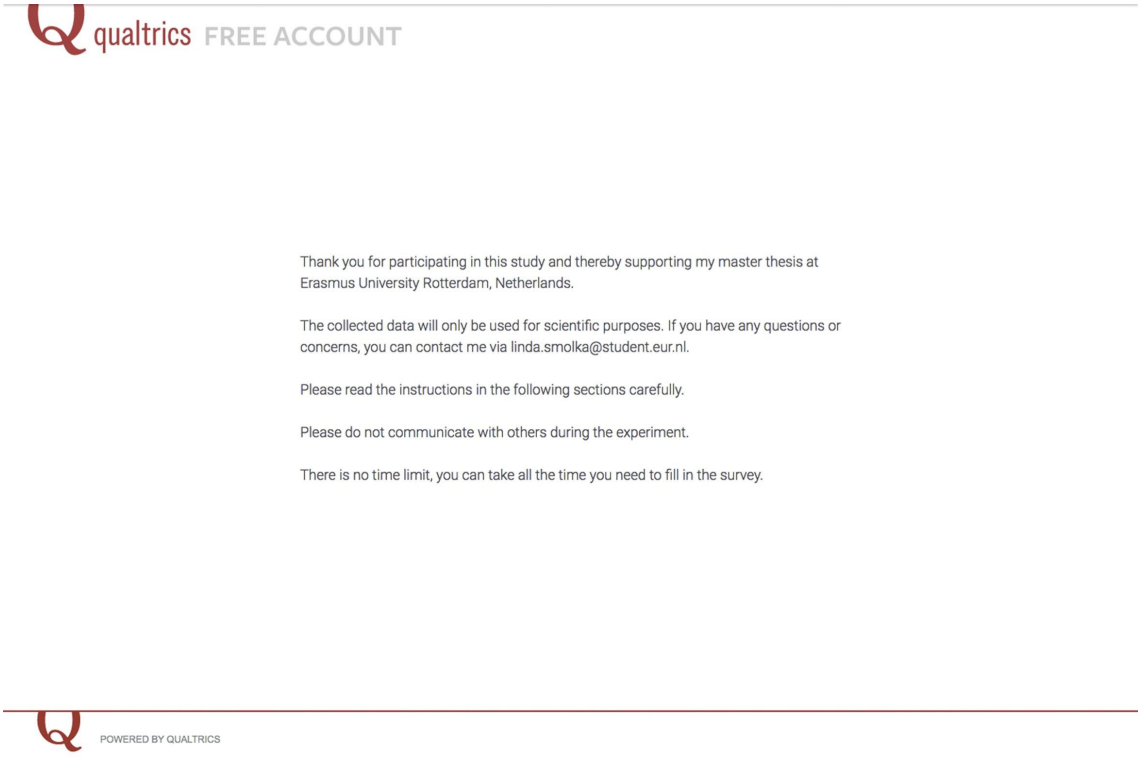


Figure 5. Example of a survey form.

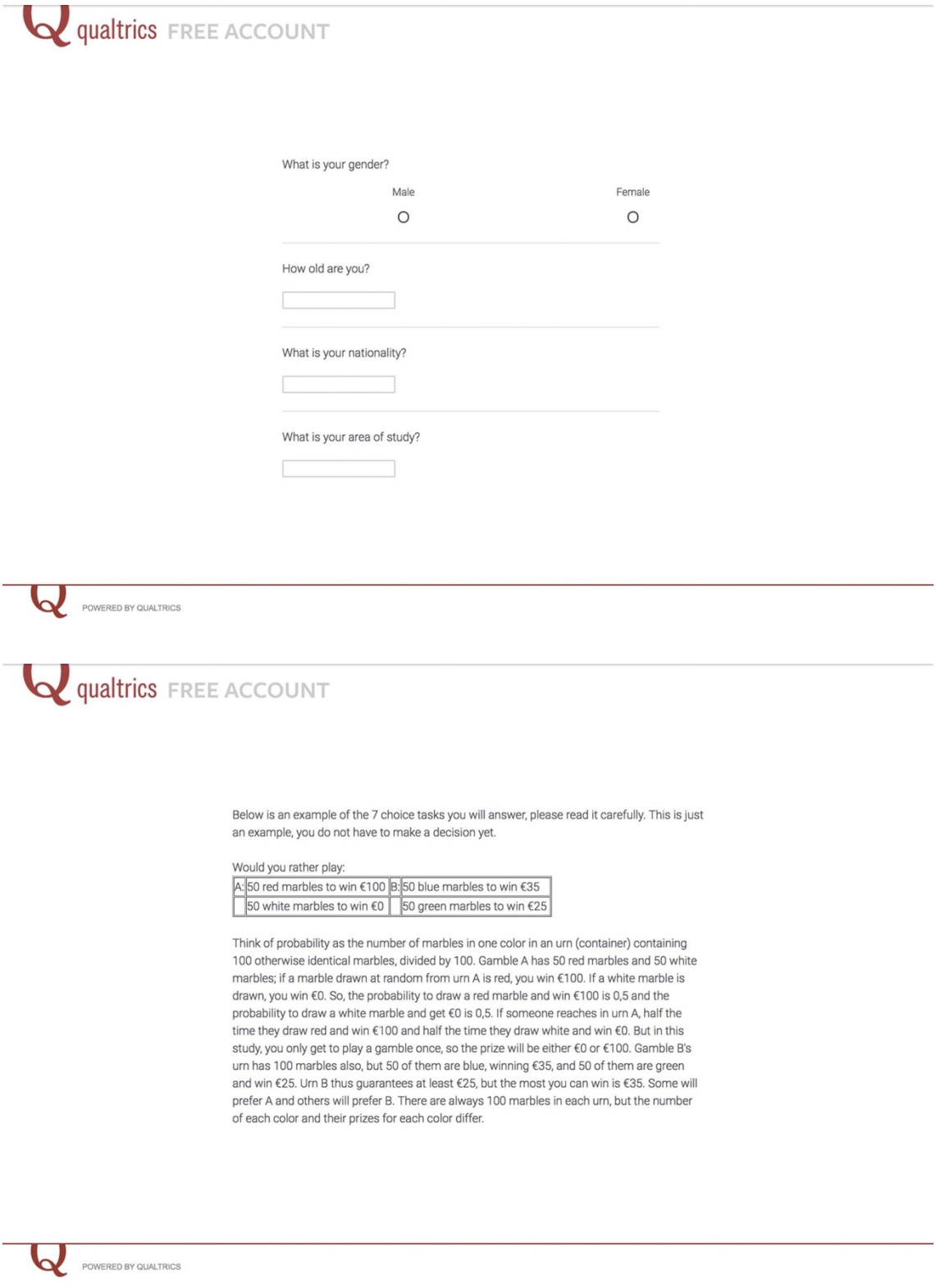


Figure 5. Example of a survey form <continued>

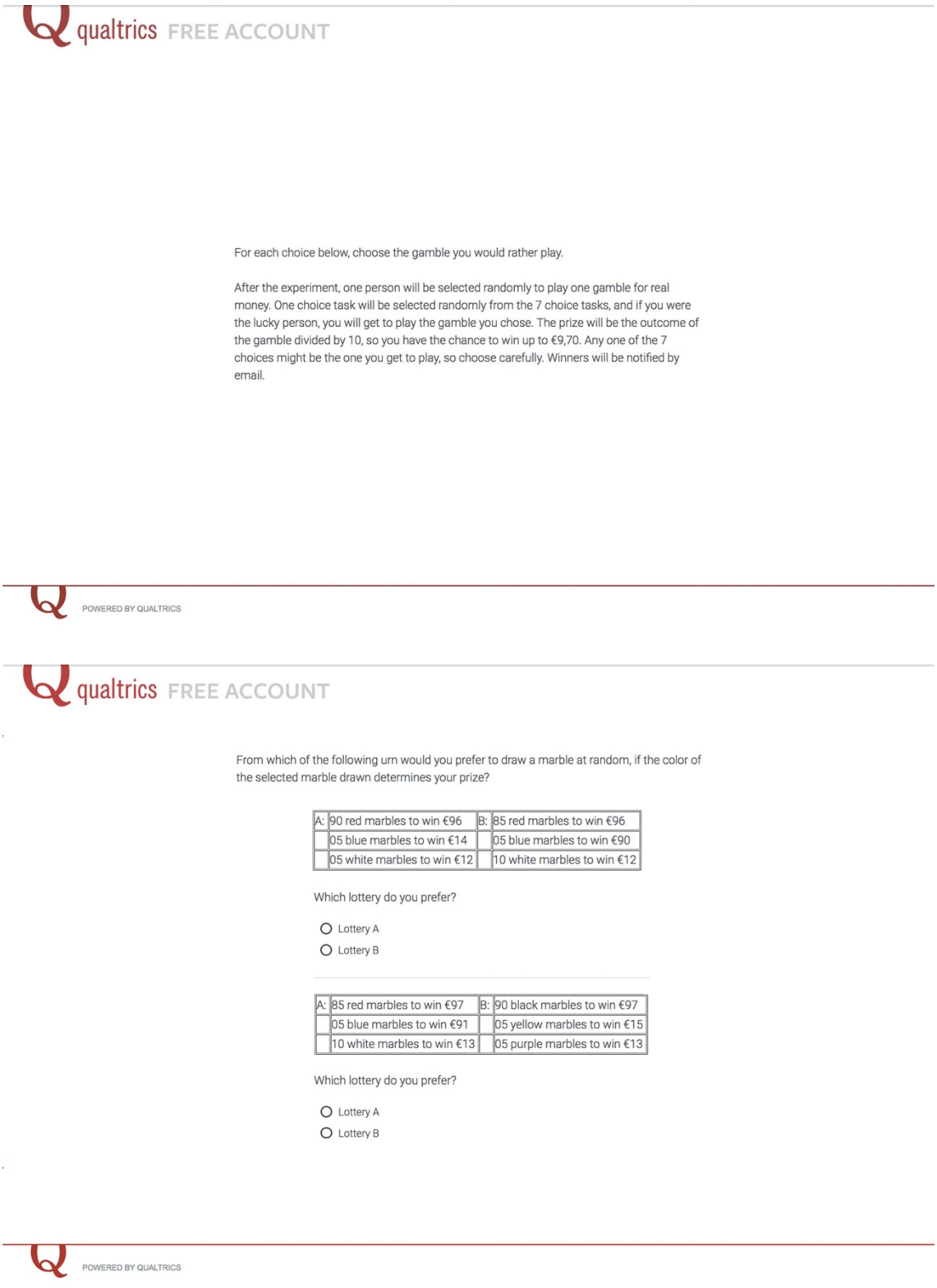
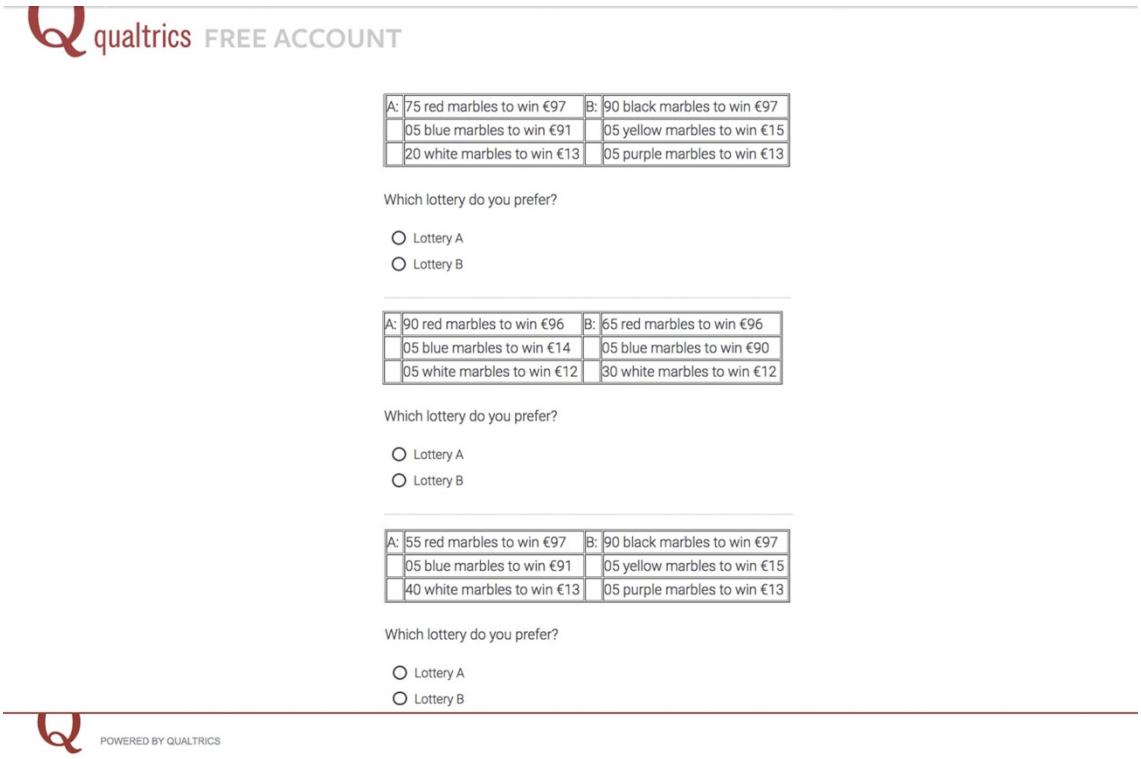


Figure 5. Example of a survey form <continued>



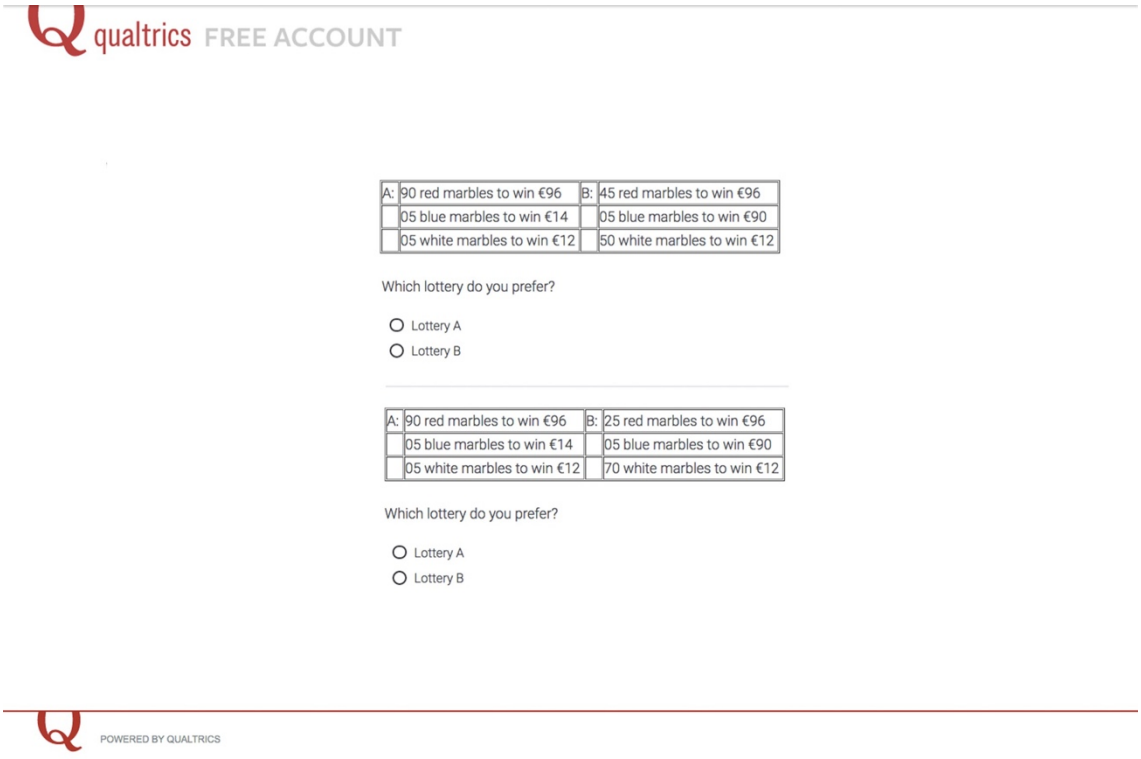


Figure 5. Example of a survey form <continued>

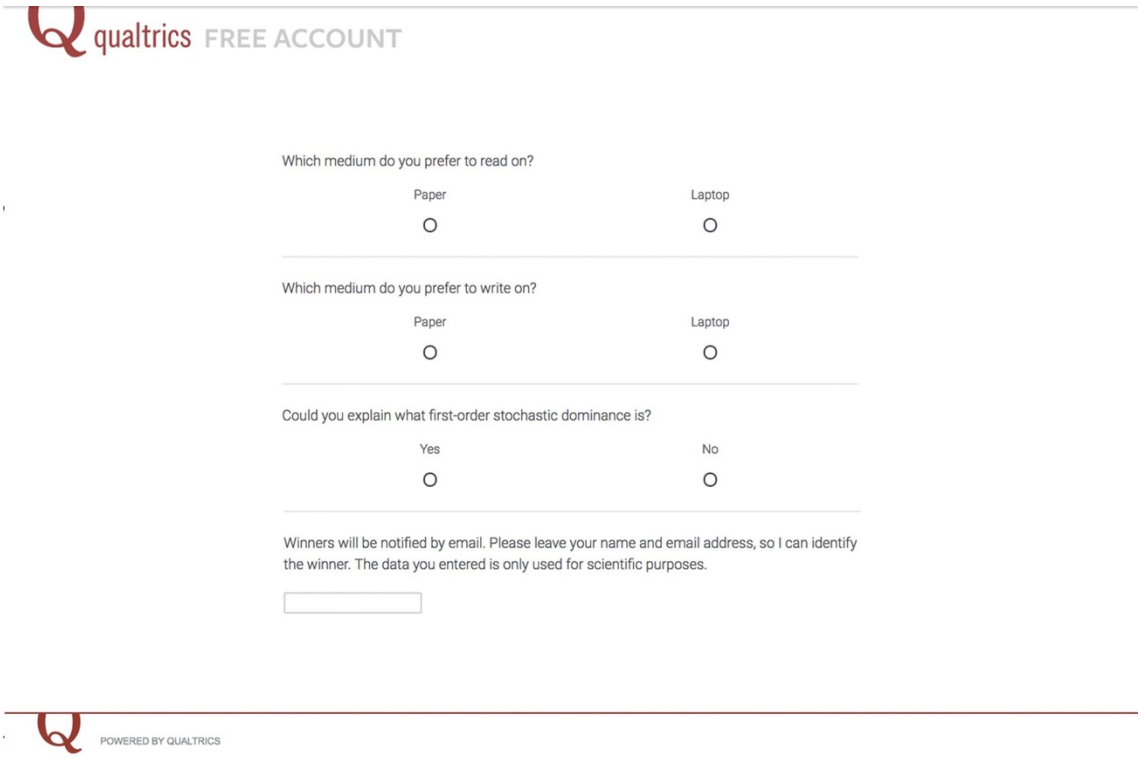


Figure 5. Example of a survey form <continued>

A.3. Further Data Analysis

The variable rationality is a categorical variable indicating the number of violations throughout the seven choice tasks on FOSD. Although it has a ranked order (0 being the most rational and 7 being the least rational), the differences between numbers do not have meaning. Thus, the data on the variable rationality is only in an ordinal scale and does not satisfy the condition of being at least in an interval scale to be able to perform a parametric statistical test. The data on the variable speed is not only in an interval scale, but even in a ratio scale since the differences between numbers have meaning and there is a true zero point at its origin, i.e. two minutes is twice as fast as one minute, and 0 seconds means no time has passed. However, in order to use a parametric statistical test, the following three conditions have to be met as well:

1. The observations are independent.
2. The observations must be drawn from a normally distributed population.
3. The two groups analyzed must have the same variance.

Condition 1 is satisfied as the between-subject design of the experiment assures independent observations. I conducted a skewness and kurtosis test for normality for the variable speed (p=0.0112) and found that the null hypothesis of normality is rejected. Therefore, condition 2 is not satisfied which is illustrated by Figure 6 below.

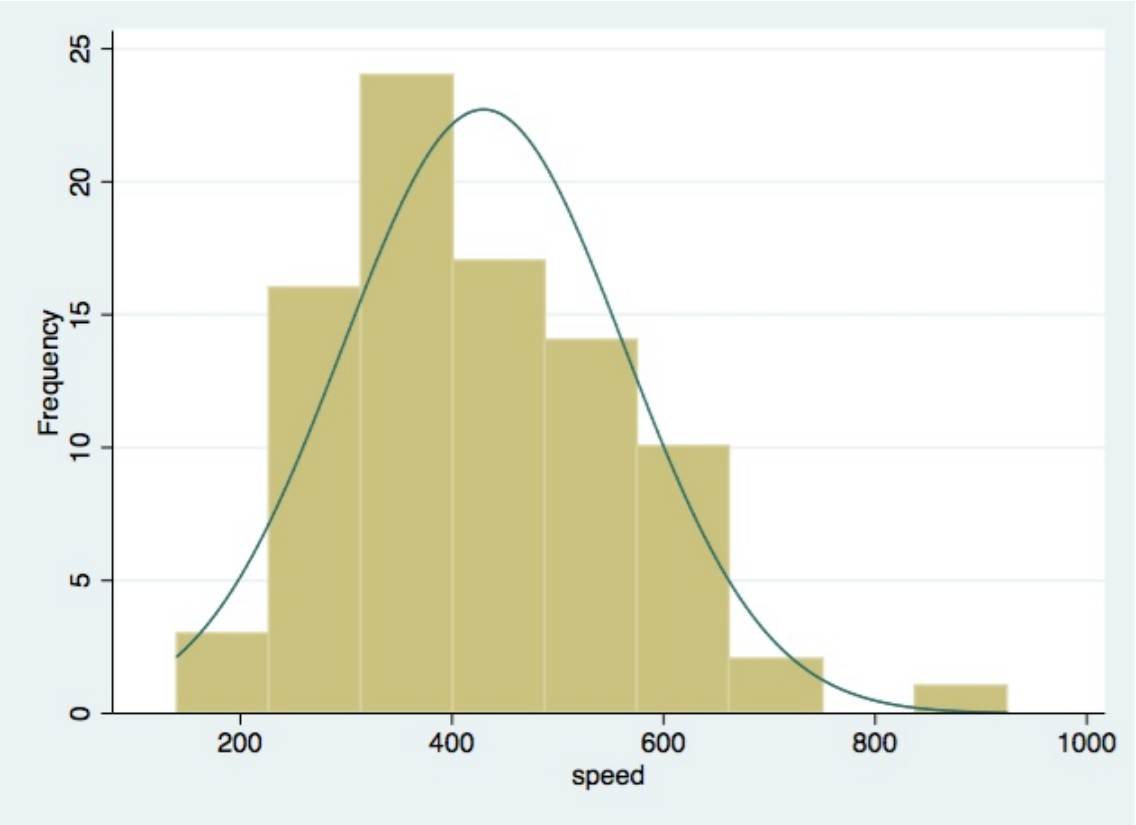


Figure 6. Distribution of the variable speed.

A normal distribution of speed can not be established, neither by taking the logarithm nor the square of the speed variable. I conducted another skewness and kurtosis test for both the logarithm of speed (p=0.0706) and the square of speed (p=0.0000), again the null hypothesis of normality is rejected. This can also be seen in Figures 7 and 8 below.

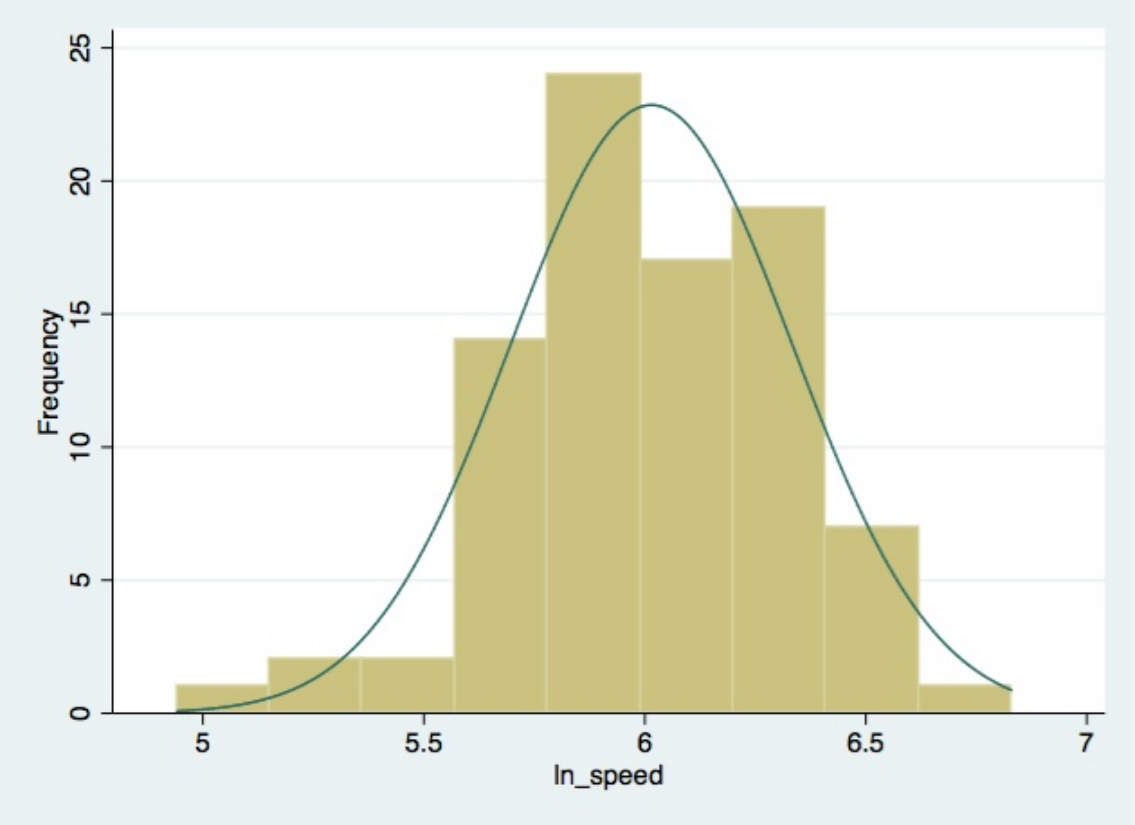


Figure 7. Distribution of the logarithm of the variable speed.

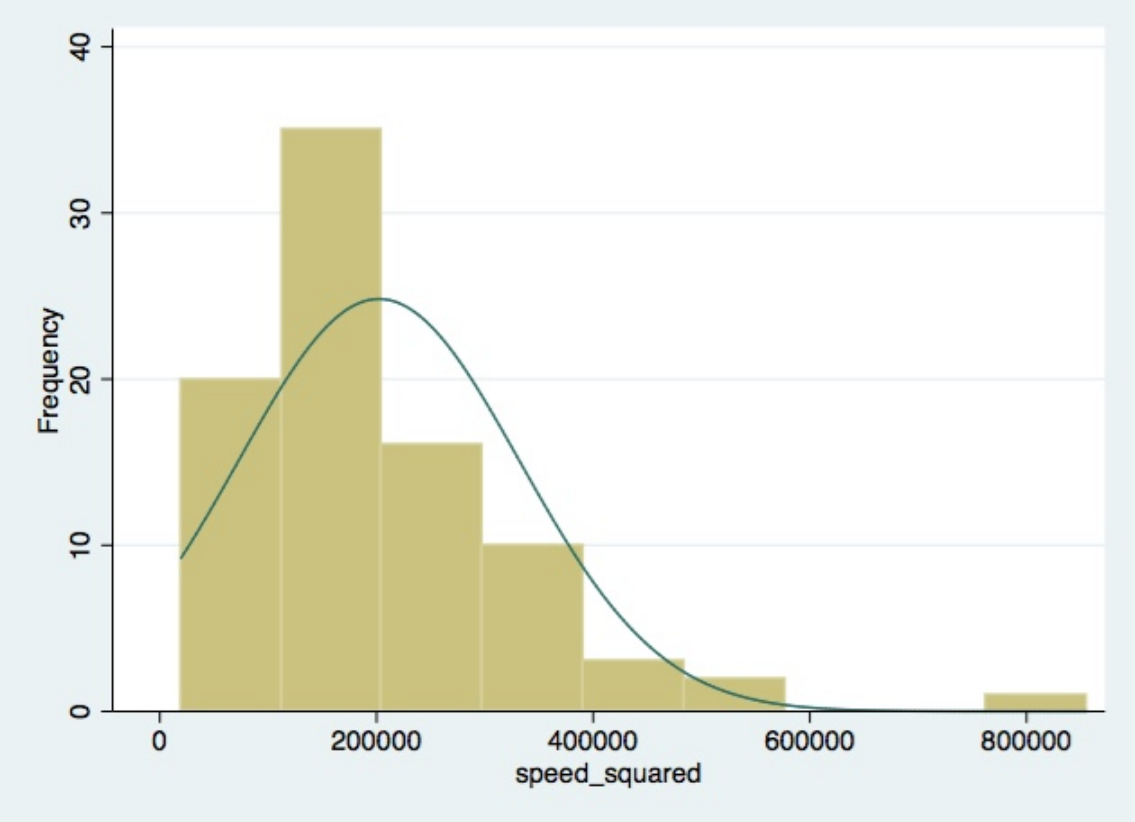


Figure 8. Distribution of the square root of the variable speed.

To test for condition 3, I conducted Levene's test for homogeneity of variances. The null hypothesis is that the group samples are drawn from populations with the same variance. We cannot reject the null hypothesis that the variances for the variable speed ($F(1, 85)=0.7073$, $p=0.4027$) are equal between the online and offline treatment. Therefore, condition 3 is satisfied. The variable speed satisfies all conditions but observations being drawn from a normally distributed population.