MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Aggregation of individual judgment"

verfasst von / submitted by

Clara Imani, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066915
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Betriebswirtschaft
Betreut von / Supervisor:	Ass. – Prof. Steffen Keck, PhD

Abstract

The concept of ‘wisdom of crowds’ is well reputed in the field of judgment and decision making. Several works consistently show that pooled estimates are better than estimates of an individual and at times it can be the best estimate too. Research also shows that estimations get better when participants were asked again and thus simulating ‘Wisdom of Crowd’. This is known as ‘wisdom of many in one’. With the wisdom of many in one mind we take a second estimate of the same person and average it, which leads to better assessments. The cognitive processes behind the estimation and decisioning is of interest, to reason why averaged quantitative estimates are performing better. Existing literature also shows that employing varying cognitive methods to obtain multiple estimates and averaging them improves the accuracy of the results. In this work we assess the cognitive processes that underly human judgments and investigate whether combining ‘wisdom of many in one’ with the ‘cognitive diversity’ enhances the outcome. We evaluate this hypothesis through an empirical study conducted via online survey. The survey was designed to involve two different groups consisting of randomly chosen individual participants. We formulated percentage-based knowledge questions and collected responses from each of the group, by providing separate sets of instructions. One with general instructions and the other one with special instructions simulating two cognitive processes, intuitive guessing and analytical guessing. Responses from both the groups were compared through statistical methods and the results are presented. Looking at the data evaluation we can say that our results are partly supported.

Kurzfassung

Das Konzept der „Weisheit der Massen“ ist im Bereich der Entscheidungsfindung bekannt. Mehrere wissenschaftliche Arbeiten zeigen übereinstimmend, dass zusammengefasste Schätzungen besser sind als Schätzungen einer Person und manchmal auch die beste Schätzung darstellen können. Weitere Literaturrecherche zeigt auch, dass Schätzungen besser werden, wenn die Teilnehmer erneut befragt werden und somit „Weisheit der Massen“ simulieren. Dies ist als „Weisheit vieler in einem Kopf“ bekannt. Mit der „Weisheit vieler in einem Kopf“ nehmen wir eine zweite Schätzung derselben Person und bilden den Durchschnitt, was zu besseren Einschätzungen führt. Die kognitiven Prozesse hinter der Schätzung und Entscheidung sind von Interesse, um zu begründen, warum gemittelte quantitative Schätzungen besser abschneiden. Die vorhandene Literatur zeigt auch, dass die Verwendung unterschiedlicher kognitiver Methoden, um mehrere Schätzungen zu erhalten und den Durchschnitt zu bilden, die Genauigkeit der Ergebnisse verbessert. In dieser Arbeit werden die kognitiven Prozesse bewertet, die der Urteilsfähigkeit zugrunde liegen, und untersucht, ob die Kombination von „Weisheit vieler Köpfe in einem“ mit „Methoden kognitiver Vielfalt“ das Ergebnis verbessert. Wir prüfen diese Hypothese durch eine empirische Studie, die durch eine Online-Umfrage durchgeführt wurde. Die Umfrage wurde an zwei verschiedene Gruppen gesandt, die aus zufällig ausgewählten einzelnen TeilnehmerInnen bestehen. Die Befragten wurden mit Wissensfragen konfrontiert deren Antworten Prozentangaben waren und die Antworten wurden von jeder Gruppe mit unterschiedlichen Anweisungen erhoben: von der einen Gruppe mit allgemeinen Anweisungen und von der anderen mit speziellen Anweisungen, wobei die zweitgenannten kognitive Prozesse simulieren, nämlich intuitives Raten und analytisches Raten. Die Antworten beider Gruppen wurden mit statistischen Methoden ausgewertet und die Ergebnisse präsentiert. Die Datenauswertung hat ergeben, dass die Ergebnisse teilweise von der Theorie unterstützt werden.

Acknowledgement

Foremost, I would like to thank my supervisor Steffen Keck, PhD for his continuous support, constructive comments and advice on my master thesis.

Besides my supervisor, especially I would like to express my deepest appreciation to my boyfriend Udayavanan Mathivanan for the ongoing emotional support, his unwavering enthusiasm and the encouragement to finish the thesis.

Also, I would like to thank my family for their constant source of positivity.

Table of Contents

1 Introduction ... 1
2 Literature Review ... 2
2.1 Decision Making ... 2
2.1.1 Findings on ‘Thinking Process’ ... 2
2.1.2 Alternative Approaches ... 3
2.2 Heuristics ... 5
2.2.1 The recognition heuristic ... 6
2.2.2 The Fluency heuristic ... 8
2.3 Anchoring ... 8
2.4 Bias ... 12
2.5 Averaging ... 13
2.6 Intuition ... 14
2.7 Wisdom of Crowd ... 15
2.7.1 Other Studies on the Wisdom of Many in one mind ... 18
2.7.2 Discussions of the existing results ... 19
3 Problem Definition ... 22
4 Empirical Study – Design & Methodology ... 24
4.1 Identification of Groups ... 25
4.2 Survey setup with “soscisurvey” ... 26
4.3 Survey design ... 26
4.3.1 Questions ... 27
4.3.2 Restrictions ... 28
4.3.3 Pretesting ... 29
4.3.4 Contacting Participants ... 29
5 Results and Discussions ... 30
5.1 Dataset Analysis ... 30
5.1.1 Modification of the variables ... 31
5.2 Demographics ... 31
5.3 Result Analysis ... 32
5.3.1 Error Analysis ... 33
5.4 Statistical Tests ... 35
5.5 Feedback from the participants ... 41

6 Conclusion ... 42
6.1 Limitations and Future research ... 42
6.2 Final note ... 43
References ... 44
List of Figures ... 47
List of Tables ... 48
Appendix ... 49
Questionnaire - Control group ... 49
Questionnaire – Target Group (Intuitive / Analytical Group) ... 55
Responses – Control Group ... 62
Responses – Target Group (Intuitive / Analytical Group) ... 64

1 Introduction

In the last few decades research in decision-making theory was done from many different perspectives: In the discipline of neuroscience the processes of the brain was investigated, in the psychological disciplines they looked more closely to the cognitive thinking processes behind the decision making process itself. Plenty of research is available on processes of memorizing as part of judgments-recognition processes, as well as retrieval. Another point of view was provided by Advocates of the rationality approach but rejected by advocates of the use of heuristics and the fact that biases exist and might perturb decision making. Different influence factors that might alter or change a person's beliefs All in all, depending on the field of research, there are many different facets of the human decision-making process. As statistical advocates found methods to increase the results of judgments, an incorporation of those statistical findings into the field of testing decision-making empirically shows to be successful. Numerous options have been tried out to optimize the estimation process itself, whereas some requirements need to be present and tested.

In the master thesis at hand, we will first dive into theory of decision-making, then review the process behind it and alternative ways of 'how to make a decision'. Moving on we will describe influence factors and effects that can appear while making a numerical guess. In chapter 2.3. we describe Anchoring itself and its underlying process. Biases plays a crucial role in the decision-making process. This is described in chapter 2.4. Improving the guesses themselves in order to achieve better answers is subject to the next section. The concept of averaging will be introduced. As the basis of the wisdom of many in one mind research object is the 'wisdom of crowd', we will also review the existing literature on this phenomenon.

After establishing the theoretical background, we propose a hypothesis by combining the 'wisdom of many in one' effect and two distinct ways of making decisions. To test whether the theoretical findings can be supported empirically we setup an experiment and shared the description of the study, its methodology and design in detailed steps in section 4. Finally, we devote the last two sections for data analysis and discussion of the results, with final conclusions.

2 Literature Review

This chapter covers the scientific research that had been done on the topics of Decision Making, Wisdom of Crowd and other relevant areas which will set the necessary background and platform for my work.

2.1 Decision Making

In this section, we will focus on highlighting different approaches on how to make a decision, factors like mood and circumstances that influence decision as well as insights of the decision-making process itself. Considering judgment, we can look at several scenarios are possible: One possible approach to a decision could be that either the person recognizes and automatically knows the answer or detects uncertainty of his or her knowledge with no immediate explication. And, if the participant has no resemblance at all, cannot provide with a solution, or an algorithm is not been followed due to the fact that it is too hard to solve, they have to find out a plan or consider an heuristic. Heuristics are also known as "rules of thumb" (Cooke 1991, p.63 & Furnham & Boo 2011, p.35) or put in other words "shortcut strategies" (Plessner, Betsch et al., 2009, p.8), which we will explain later.

2.1.1 Findings on 'Thinking Process'

Searching for a solution at hand we investigate thoughts. A possible thinking processes underlying the similarity assessment must be mentioned, whereas they search and meanwhile comparison of the present issue with past problems happens and a prove if this can be applied to the present issue runs in the mind.

Further examination leads to circumstances, that the decision maker faces within making up his or her mind. Possibly there could be restrictions, which could influence the decision. Confronted with this constraint, we look at preferences in a sense that they decide which resource is of more importance and convenience than others. (Kirchler & Hoelzl, 2012, p.27)

Also, from importance is the thought whether it depends if it is a certain or uncertain event. The term risk is defined by the probability (Hens & Rieger, Financial economics, p. 80 and Kirchler, 2012, p.38) As Kirchler, 2012 pointed out, humans prefer certain decisions compared to uncertain ones and even avoid ambiguity. (Kirchler, Wirtschaftspsychologie, p.)). Parnell, Bresnick et al names a few possible influence factors that could hinder arriving at helpful options to choose from. By a “block” they mean that the flow of solution mechanisms and consideration of problems is interrupted. Put into three categories of “blocks” they differ between “perceptual blocks, emotional blocks, cultural and environmental blocks and intellectual and expressive blocks.” (Parnell, Bresnick et al, p. 154)

In the category of emotional blocks, they “struggle to deal with ambiguity and uncertainty”- similarly as Kirchler pointed out that the preferred version is dealing with certainty.

2.1.2 Alternative Approaches

Let us evaluate another point of view, which can influence the decision maker as well. We think of the following factor. The mood of a person, who is choosing is also relevant. In other words, the emotional state plays an important role while forming an individual judgment. Being in either a good or a bad state of mind, people apply it as an information source and decide, depending on how they feel. (Englich and Soder, 2009, p.41). Another paper also points out that the mood before making a judgment has an impact. In the case of evaluating the respondents "subjective well-being" they point out that also the current mood of the time they have been asked is important. It even turns out to lead to a biased outcome as they withdraw information from their own emotion to an overall estimation of their "subjective well-being". The current mood determines the direction of the statement they made. It appears that there will be exaggeration through overestimation coming from a good mood. (Wortman, Scherer et al., 2017). To take up on Englich and Soder’s second observation: The authors say it matters in which emotional state you find yourself in- happy, sad or neutral. This determines the quality of your guess. Moods are registered as controlling or affecting the procedure of understanding facts in two ways.: They seek advice from "moods" and towards which direction it points, knowledge is controlled and differed in choosing with the given knowledge. (Englich and Soder, 2009, p.41)

If we look at other fields where decisions are made, let us take for example finance decision models, variables have been considered illustrating „economic and cultural factors" (Hens & Rieger, Financial economics, p.62). So broadly speaking each decision no matter the type, do have various variables that influence the decision itself and its outcome.

Let us now assume, that we can divide deciding into stages. Before reaching an agreement, a person must pass through those stages to reach the final guess. Looking at the structural approach, those efforts are directed to ideas and a constructed concept. While Process theory focuses on the process how someone comes to a certain result through the concluding: In other words, this theory relates to forming an opinion or reaching a guess. (Svenson, 2016, p.885)

To come to conclusions about a proclaimed opinion, a tool called CTA is used by “natural decision making” to have a closer look at procedures and interpretations about opinion-forming. (Klein & Kahneman, 2009) The early shoes of CTA, used to form the foundation of the evaluation of Analytical strategies and decisions. CTA stands for cognitive task analysis. CTA is defined as “semi-structured interview techniques that elicit the cues and contextual considerations influencing judgments and decisions.” (Klein & Kahneman, 2009, p.417) As an example where CTA works, they mentioned a case of nurses skilled to track down infections.(Klein & Kahneman, 2009, p.417) The appearance of supplementary methods of CTA, has been used for several other purposes. Many of the used concepts of CTA have the following approach: First an introductory information assembling, which is followed by illustration of information. The next step would be implementation of assessment with a subsequently check of the methods, furthermore the examination of data. Lastly, they finish with "format results for the intended application." (Klein & Kahneman, 2009, p.417) In the context of the second step, which is the recognition of the knowledge representation, we differentiate between two types of knowledge: procedural and declarative knowledge. The early shoes of CTA used to form the foundation of the evaluation of the Analytical strategies and decisions.

The differentiation about the level of information in knowledge, which is that the decision-making person sustained the likelihood, to which extent they will be apparent (Toplak & Weller, 2017, p.12).

In the context of describing different approaches and circumstances of judgment and decision making and individuals decision-making processes, we now have a look at Hammond's findings with respect to this subject area. The accuracy of collecting information before coming to a decision depends on the following factors:

In Hammond's lens model, which was established in the year 1952, we look at the processes of forming a decision in a different light. The lens model has later been developed into the so-called social judgement theory, in short "SJT". It also enhances the assessment scheme before deciding.

In their "social judgment theory"- in short SJT, they make the distinction between different types of "judgement situations". Investigated is the process, before a person comes to a decision and it revises the given exercises. (Hammond, 20, p.13) Different scenarios are possible with two or more participants, but in the following illustrations, the focus is on findings concerning one individual. (Dhami & Mumpower, 2018, p.) The SJT not only includes the assessment schemes before coming to a decision, but also highlights the way to seek an enhancement concerning "cognitive performance". (Dhami & Mumpower, 2018, p.2).

2.2 Heuristics

A possible way to approach decisions is the use of heuristics. Heuristics are view "strategies that people use deliberately in order to simplify judgmental tasks that would otherwise be too difficult for the typical human mind to solve." (Gilovich, Griffin et al., 2009, p.4)When ANOVA and multiple regression among other instruments are seen as "optimal or rational strategies", heuristics can be seen as representing "discrepancies between these rational strategies and actual human thought processes" (Goldstein and Gigerenzer, 2002, p.75) In contrast to seeing heuristics as discrepancies, Gilovich, Griffin et al., (2011) see heuristics in a more positive perspective : in their point of view "strategies that people use deliberately in order to simplify judgmental tasks that would otherwise be too difficult for the typical human mind to solve." (Gilovich, Griffin et al., 2011, p.4)

As Tversky and Kahneman (1974) pointed out, the usage of heuristics has a positive effect. In their article "Judgment under uncertainty: heuristics and biases" they suggested that heuristics "reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (Kahneman &Tversky in 1974, p.1124) Considering decisions under uncertainty, heuristics facilitate the technique to reach at the right guess.

One possible outlook for the underlying process of the decision-making with the use of heuristics can be described: Humans select the first possible option that comes to mind. Once a suitable option is picked, they proceed further with a mental search process, which they call as "adjustment and anchoring". Processing information can result in biases. Findings about memory models show "judgments and decisions are produced by storing and subsequently retrieving objective evidence in memory and that biases are the result of distortions in this mental process" (Hilbert, 2012, p.6).

Another definition is made by Hilbert (2012): Heuristics are "short-cuts in our information processing that aim at reducing cognitive effort" (Hilbert, 2012, p.5).

Though heuristics can be covering a wide range of concepts, we focus on two specific types which are relevant to the context of information process and decision making. The Recognition and Fluency Heuristic Hertwig, Herzog et al. (2009).

2.2.1 The recognition heuristic

To look where it originates from, we start with a notion called the "cognitive effort". The "cognitive effort" implies an information search, which opens into its incorporation. Two requirements need to be present for the recognition heuristic to work: one option out of two is chosen due to its higher value. As we speak about a quantitative criterion named as "objects". That the recognition heuristic reaches to a good performance works if and only if just one object is recognized. The "recognition heuristic" is described as a process, where perceived (findings) "memories" are proven to be ranked as superior than the others (motive).

The "recognition heuristic" applies under the condition that only one out of two targets are being identified by the person trying to remember. Oppenheimer (2003) describes it in other words as "no other information aside from recognition is taken into account in the judgment." (Oppenheimer, 2003, B2) So only few cognitive processing is demanded for this heuristic. But it does not suit into the textbook scheme for heuristic processing. Also, the correlation between either the criterion, the recognition, and the retrieval fluency should be present. In comparison, Goldstein and Gigerenzer, (2002) also point out different rules for the recognition heuristic to work: Firstly, a strong correlation between recognition and criterion has to exist, and the heuristic operates with success when "lack of recognition" is present. They found out that the "lack of recognition" has a systematically distribution, which is the case, when the needed strong correlation with the criterion is in place. (Goldstein and Gigerenzer, 2002, p.76) Goldstein and Gigerenzer, (2002) illustrate two examples where the recognition heuristic arises. People have the assignment called "paired comparison or two-alternative forced choice". (Goldstein and Gigerenzer, 2002, p.76) Two groups were asked to answer a question. Their answer proved to be made by recognition or no recognition. They let the participants guess, whether the population of San Diego or San Antonio is greater. As both cities are in the united states, it was surprising that only two thirds of the Americans made the right guess, whereas hundred percent of Germans made the right guess that San Diego is larger. Nevertheless, most Germans simply did not have knowledge about these cities, all of them made the right guess. As a second attempt, they verified this theory with guesses of a soccer game. (Goldstein and Gigerenzer, 2002, p.76)

So, what can be observed is, that the process of retrieving a memory record, is that the recognition knowledge reaches the mental stage and is prepared to surpass the deductive process, whereas another knowledge does not surpass. The ability to surpass proves that it aims to succeed. (Hertwig, Herzog et al., 2009, p.1192).

The use of recognition heuristic, its importance and applicability are widely spread across different judgments like for example "judgment of demographic, geographic, and biological quantities". (Hertwig, Herzog et al., 2009, p.1192).

2.2.2 The Fluency heuristic

Fluency heuristic serves as an example that within “retrieval from memory” retrieval fluency is the concomitant. The “fluency heuristic” needs to fulfill the requirement of fluency, which is necessary to access permanently saved information and need. The “fluency heuristic” needs to fulfill the requirement of fluency, which is necessary to access permanently saved information and needs. When both objects are recognized, the fluency heuristic is suitable. In contrast to the recognition heuristic, the fluency heuristic needs both “objects” to be recognized, whereas the “recognition heuristic” applies under the condition that only one out of two targets are being identified by the person trying to remember. So, both heuristics converge to a good output, less of cognitive effort will surprisingly lead to an achievement of good performance. (Hertwig, Herzog et al., 2009)

2.3 Anchoring

During the process of decision-making, humans use several cognitive heuristics as tools leading them to their decisions. While trying to find out a set of possible and accurate options to choose from what the brain suggests, some “cognitive biases” perturb this process. One such cognitive bias can be “anchoring” (Parnell, Bresnick et al., 2013, p.152).

Tversky and Kahneman (1974) refers ‘Anchoring effect’ as the disparate impact on decision makers in forming judgements which are inclined towards an ‘initially presented value’. To look at it, not as a perturbance of the choosing process of several options, but as a mean to approximate a value, anchoring can be viewed as an aid to form a decision given a fictitious number representing the supposed "anchor".

Anchoring is also seen as an ‘adjustment heuristic’ when participants involving in a judgement look at a decision, that is uncertain. A value is adjusted to reach the final answer. This guess under uncertainty arises because "anchoring and adjustment heuristic is assumed to underlie many intuitive judgments" and consequently might lead to judgmental biases (Epley& Gilovich, 2006, p.311).

Trying to find out the parameters and backgrounds forming the anchoring process, a three-step procedure is introduced: In step one the anchor itself is examined. The second step is named as “confirmatory search”, where the extraction of information from memory is detected, that supports the use of the anchor to form a judgment. Integration & adjustment is seen as the third step. It can be executed by using various techniques such as “averaging” (Lopes, 1985), “insufficient adjustment” (Tversky & Kahneman, 1974) and “adjust to range of plausible values” (Quattrone et al., 1984). The last technique not only includes the anchor but also an adjustment considering a “target value”. The direction of the anchor surprisingly points into the direction of the anchor-meaning towards the anchor- to form the final answer. (Wilson, Houston et al., 1996, p.388)

Similar to the third technique a modification would be to include upper and lower values into the estimation. We consider the following scenario: A person is asked if a they would conform that a value is below or above an indicated number, which they call as "anchor". Many studies show that this boundary, a slight modification of just having one value to compare your guess with, the "anchor", also proves to influence the answer.

Mussweiler & Englich (2005) looks at what and how the adapting procedure of guessing a value happens with the help of the anchor: the anchor value in the decision forming process seems to manipulate the participant's answer unconsciously. Not simply providing a direction or a hint helping the participant to reach at a preferably good estimate, but the anchor itself could even be anticipated as the right guess.

The anchor itself is a given number, trying to influence the participant's response. Strack & Mussweiler (1997) defines it as “a biased estimate toward an arbitrary value considered by judges before making a numerical estimate”. (Strack and Musssweiler, 1997, p.437) Examining how anchor values are treated during the judgement, instead of choosing the anchor itself as the solution, it is also possible to take the proposed value, the anchor, and adapt it to an interval of possible predictors we have in mind. This is possible in the case that the anchor presents a value that is way off the mark. (Strack and Mussweiler, 1997, p.438)

Having stated different characteristics of anchoring and the underlying concepts in ways the anchoring effect could be used for guessing values, a review of the literature at hand shows the impact of anchoring. The result is surprising: not many requirements need to be fulfilled that anchoring is working: The anchoring effect is present even if the provided number is not connected to the question asked, meaning that they picked a random value to be the anchor. (Wilson, Houston, et al., 1996, p.387)

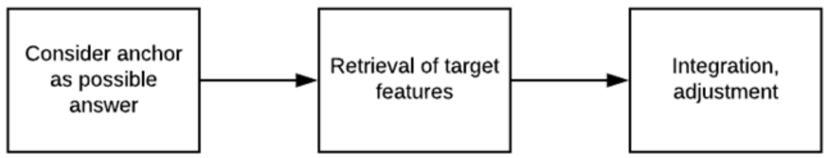

Figure 2.1: Three steps of Anchoring (Lopes, 1985)

In this context the findings of a well-known "anchoring study" should be noted, which Tversky and Kahneman (1974) conducted. According to the findings, anchoring has an influence on a person's guessing. The study was executed in the following way: What they did was posing questions where they indicated a value, the anchor, which was between the range of zero reaching up to hundred. For example, one of the questions was, if participants can guess "whether the percentage of African nations in the United Nations was higher or lower than that number" With that number, the anchor, they mean an anticipated and simultaneously artificial percentage. Once participants were asked to answer with their guess, they formed an estimate according to the anchor. Participants who were provided with a high percentage, gave high estimates as answers whereas giving a low percentage as an anchor, led them to decide for answering with a low value. (Strack & Mussweiler, 1997, p.437 & Wilson, Houston, et al., 1996 p.387).

So, the anchor achieved determining the direction of participants final guess. Bottom line, the anchor served as a guideline in which direction the estimate should go.

In 1995 Jacowitz and Kahneman altered the finding of this study, which was executed in 1974, choosing a different approach: This time they wanted to "determine whether a target value is higher or lower than that of a given anchor". This anchor, which is off the mark of reasonable estimates, will lead to approximations where they only settle for a value, when the participant thinks they reached the close by endpoint of a fictitious estimate range. But as previously mentioned before this so-called “adjustment process” was criticized: It leads to insufficiency but still the anchor proves to work. (Strack & Mussweiler, 1997, p. 438)

Searching for further modifications or scenarios that possibly can result in different guesses, telling participants beforehand about the influence of the anchor might have an influence on their guess. According to Epley& Gilovich (2006) aiming a good estimation originating of the use of an anchor “forewarnings and financial incentives” cannot be counted as being reasons for influencing the guess in a positive way: They detected just a little or no effect at all, which is not sufficient enough to be counted as influencing the guess (Epley& Gilovich, 2006, p.311). If participants were given the instruction to ignore the given value, the effect of the anchor still works. (Wilson, Houston, et al., 1996, p. 387)

Other additional findings about conditions of anchoring have been established again by Mussweiler et al. Several conditions have been listed where the anchoring effect works: Surprisingly, there is no need for expert knowledge, neither the presence of motivation is necessary, as well as anchors which do not contribute for the participant to find the correct answer or values, which are way off the true value. During the discussion of these circumstances two studies were introduced. According to those, the anchoring effect except for using a "corrective strategy" like "consider the opposite", which leads to weak anchored results, proves to work.

With all the above described characteristics it is important to know that the scope of anchoring is valid and applicable for several domains “general knowledge questions (Russo & Schoemaker, 1989), utility assessment (Hershey & Schoemaker, 1985), causal attribution (Quattrone, 1982), the detection of human deception (Zuckerman, Kioestner, Colella, & Alton 1984) predictions of future performance (Switzer & Sniezek, 1991) , predictions of the likelihood of future events (Plous, 1989) and task persistence (Cervone & Peake, 1986).”(Wilson, Houston, et al., 1996, p. 387)

In the context of highlighting conditions and underlying processes of anchoring, participants might search for the help and use of a heuristic, which called the anchoring and adjustment heuristic. Put in other words they look at a decision, that is uncertain. A value is adjusted to reach the final answer. This guess under uncertainty arises because "anchoring and adjustment heuristic is assumed to underlie many intuitive judgments" and as a consequence might lead to judgmental biases (Epley& Gilovich p.311). As the purpose of this chapter is about anchoring, for further details and definitions about intuitive judgement I refer to Chapter 2.1. intuitive judgements and Chapter 2.3 providing insights of biases.

2.4 Bias

As people form decisions, they go through a process of evaluating the preferable option. This search is perturbed by the occurrence of so called “cognitive biases”. (Parnell, Bresnick et al. 2013, p.152) Possible biases are for example “anchoring, availability, comfort zone bias and motivational bias.” Anchoring bias means that a value is influenced by the choice and it describes the struggle to see other possible options when the mind already provides a suggestion. The availability bias is that easily remembered thoughts are emphasized more compared to the ones which are hard to recollect. Thinking outside of the suggested options is not likely in this case. The comfort zone bias explains the fact that possible ways which are not well known may not be taken into consideration because of the lack of feeling prepared or properly ready towards this alternative option. They might not even consider an alternative that would lead to a useful option. Another bias is the so-called motivational bias. This bias intervenes with the number of options that are taken into account by reducing it down to their favorite option, which is not guaranteed to be the most desirable one. On the contrary Cooke (1991) had a different approach in mind, which defined the motivational bias as being a delusional opinion derived from intentional false information. (Cooke, 1991, p. 63)

Besides the biases outlined above Dutta, Mandal et al. (2012) mentioned other biases in their book ‘Bias in Human Behavior. The perspective is the following: We extract these findings of information processing, as they are important for analyzing possible distortions, which can appear through this stage. Because it has an influence on the decision-making process which subsequently is the stage that follows, after the brain handles the information with which it is confronted. (Dutta, Mandal et al., 2012) The focus is put on biases, which they call "Attentional bias, Reading bias, visual processing and auditory processing."(Dutta, Mandal et al., 2012, p.5f) As the attentional bias is associated as being a "cognitive variable" and its occurrence is more probable when there is an asymmetric stimulation to the left and the right side of the brain.

Arriving at a point of defining the various types of bias, it is important to state the connection between bias and error. A suggestion for that is pointed out by Cooke (1991). We can talk about the error as being a bias when it originates from heuristics, which lead to a decision. (Cooke, 1991, p. 63) The use of heuristics can result in „predictable errors." The term "error" is, put in other words, a bias.

2.5 Averaging

The origin of the averaging principle was the fact that combining two mindsets instead of one will target a more convenient result. Research showed, that when more than two minds have to judge, there is an effect which they call "Creative Plus". This arises and is based on a statistical fact, which deals with the following key point: "Aggregation of imperfect estimates reduces error." (Larrick, Soll, 2006, p.2) The aim is that the usage of the averaging principle leads to good estimates. We speak about estimates that will come from the same person. (p.283) Larrick and Soll (2006) further describe the averaging effect with comparing two different measurements called the MAD, which stands for mean absolute deviation. This is also a kind of average.

Another factor that leads to reducing the error while using averaging, was pointed out by several authors: Independency of the estimates. This being present will lead to good results using the averaging method. (Vul &Pashler (2008); Herzog &Herwig (2009); Surowiecki (2004)) Furthermore, Rauhut & Lorenz (2011) pointed out that the independency of the guess is based on an “internal distribution”. (Rauhut & Lorenz, 2011, p.191)

To highlight one survey conducted by Müller-Trede in 2011 participants were asked to make a quantitative judgment. Examined are one-person estimates but with a different approach. The outcome of the questionnaire is also influenced by two factors: in the best case, the two estimates should be contrary, which means independently. This is described in Herzog and Hertwig's paper "The wisdom of many in one mind" as two estimates being at the other side of the "true value", but they considered two estimates. The second factor is that the person questioned should approve the use of averaging the two estimates., which turned out not to be the case. (Müller-Trede, 2011, p.289) In an experiment they found out that 4 out of 6 questions provide the predicted gain that originate from averaging-Participants were asked to answer five questions.

Two findings have been established so far regarding the use of averaging. First, research showed that you can differentiate between three approaches forming a third answer. People who have been asked to guess a third time, form either a completely new answer apart from the first two answer or they stick with the first or the second guess. Out of these answers only a fraction of the participants averaged their statements. A possible explanation for this low rate could be that people are not aware of the improvement of combining their first two guesses.

Müller-Trede (2011) as well as Larrick & Soll (2006) use the following for possible explanations of participants not averaging their guesses but instead deviating or replicate their original answers. The occurrence of the respondent's shortcomings evolves from the lack of information about averaging benefits. That is why the provided answers failed to lead to better estimations.

2.6 Intuition

As part of our survey will be to ask participants to make intuitive guesses, we will quickly define what we mean by intuition. From the viewpoint of Kahneman (2003) intuition is described as" thoughts and preferences that come to mind quickly and without much reflection "(Kahneman, 2003, p.697)

Sauter (1991) presumed intuition to be "compressed expertise, a way of rapidly accessing chunks and patterns of knowledge formed from previous experience" (Sauter, 1999, p.110)

Those two terms are general definitions of the term intuition but for our purposes the most convenient definition is "the recognition of patterns stored in memory" (Kahneman & Klein, 2009, p 516). So, when the participant makes an intuitive guess in the survey, the person will try to recollect from memory while answering, which we call as intuition. A possible downside of just using intuitive judgments alone was pointed out by Sauter (2011). They also point out that within making a decision intuitively a wrong idea might be taken into consideration and might lead to incorrect conclusions. (Sauter, 2011, p.110). Using intuition "They may reach conclusions too quickly, ignore relevant facts, or follow an inspiration when it is clearly bad " (Sauter, 2011, p. 111)

2.7 Wisdom of Crowd

As the concept „wisdom of the crowd" suggests, that asking either experts or participants with no prior experience to estimate the outcome of future events and averaging their guesses will lead to surprisingly good and precise results. Herzog and Hertwig (2009) found a method that makes it possible to compare opinions of different people with two opinions from the same person. Conducted was a survey with hundred and one participants, which were rewarded either monetarily with ten swiss francs, to win an iPad out of two or with points for their universal studies. The setting of the questionnaire was to put the participants into two different groups. They had to answer knowledge questions. The first group, which they name as "immediate condition", was asked to do the questionnaire a second time, right after completing it. The second group answered the same questionnaire three weeks later, which they called the "delayed condition". Both groups were not forewarned that they would be asked to answer the questionnaire a second time. Results were the following: a comparison of the "mean squared error" of the first, the second and the average estimations of the "immediate condition" show that "the mean squared error" of the average is lower than those of the first and second estimations. Even postponing the second implemented questionnaire, which is three weeks later, the second condition, again leads to a "lower mean squared error" of the average as to the first and second answers. So, in sum, the error of the average is always smaller compared to the first or second estimate no matter if you look at the immediate or delayed condition. As the second guesses' "mean squared error" was higher than the first guess, we can conclude that the answers of the second guesses were even worse than the first estimates meaning further away from the right answer.

Since the second estimates turned out to be like that, a boost in information during those three weeks can be highly doubted.

If we consider this case it is important, that the second estimate, which they call “dialectical estimate”, is based on different premises and insights than the first one, which we see as they deviate. They call these two guesses as “conflicting opinions”. (Herzog and Hertwig, 2009, p.231)

Looking further at the results of the survey we can say that the consequence of delaying the second time of executing the questionnaire for three weeks, has a positive effect. Also, it can be stated that the independence criteria are fulfilled in the delayed condition due to the fact that an information gain can be doubted. As a consequence of independence being present, the gain from averaging is greater in the “immediate” than in the “delayed condition”. We define the gain as the difference in “mean squared error” between the first guess and the average. This difference in error is bigger and there is a reduction of noise in the immediate condition rather than in the delayed condition. Those estimates, which will be averaged, need to have different origins leading to different errors.

Looking at averaged values, we can say that two positive effects of averaging are: systematic error is diminished and random error crosses out. It can be better explained through the idea of bracketing by Larrick and Soll (2006) where using estimates that would come from “the same side of the truth” leads to the same outcome as choosing randomly between two estimates, whereas, when the estimates come from the opposite sides of the truth, averaging shows improved outcome. In order to explain the term dialectical bootstrapping, the following condition is evaluated: When is the average of the estimates better than the first estimate? Assuming there exists a true value, which is the right answer to the initial posed question, two estimates from the same person, the first and the second estimate and their averaged value. As stated earlier the two estimates arise from different knowledge to some extent as Vul and Pashler (2008) names it as independent estimates. For the above-mentioned condition to be true, the second estimate or “dialectical estimate” should lie within a range called ‘gain range’. Evaluating the position of the guesses we define the gain range. The limits of the gain range are defined by two boundaries. An upper-boundary and a lower-boundary. The upper-boundary reaches from the first estimation until the true value and the lower-boundary has the extends in the opposite side of the true value equivalent to three times the error of the first estimate. Under the condition that the second estimate must lay in the gain range, the average guess is better than the first estimate. The second estimate is located in the gain range, when the two estimates are partially based on diverse knowledge. Herzog and Hertwig (2009) call it as “nonredundant knowledge”. Having said that, the errors of those estimates will have different errors.

The paper refers to the insights of two different methods comparing errors aiming to enhance a better outcome: The first is called dialectical bootstrapping and the second evolves from debiasing research. They found out that the approach of “consider the opposite” technique provides a similar outcome as using the method of dialectical bootstrapping. (Herzog and Hertwig, 2009, p.233) Dialectical bootstrapping takes the average of the estimate of solely one participant, which boosts accuracy. The use of dialectical bootstrapping is for 72% of the participants beneficial. Although the method of “consider the opposite” collects opinions of two persons the outcomes both seem close to the true value as the estimates differ from each other, because they are based on different knowledge. In order to receive an average that boosts accuracy, the second estimate has the freedom to lie within a range that allows it to be three times more off than the first estimate.

Another advocate of averaging is Vul and Pashler (2008). As Herzog and Hertwig (2009) starts his argument considering the wisdom of the crowd’s average proves to be better than considering the estimates of the individual.

But the following concern is expressed: Due to the fact, that the first estimate could even be closer to the right answer, asking the same person for a second guess might consequently result in a bad average value. This is the case when the second estimate is worse than the first as it was the case for the survey conducted by Herzog and Hertwig (2009).

To add to established facts by Herzog and Hertwig (2009), Vul and Pashler (2008) conformed and highlighted another possible viewpoint: He found that the two estimates we are looking at are based on a "internal probability distribution". In this case the average of two estimates will be better than the estimate, when the bias is independent.

To measure the weight, how much a supplementary estimation of group members is worth compared to an individual, the authors introduced λ and what it tells us.

If this factor is 1, using the method of averaging on results, where people were asked twice or taking the result of another person will lead to the same result. Consequently, λ equal to zero, while using the method of averaging (beforehand), means that there is no improvement at all if you compare the following.

This formula demonstrates the "reduction in mean squared error" if an individual has been asked a second time: $1/(1+\lambda(N-1))$. In the immediate condition of the evaluation you get a factor of 1.11 where the 0,11 can be interpreted that asking the same person twice leads to an improvement about 1/10th. Looking at the delayed condition, we see a bigger improvement. The postponed answers of asking the same interviewee again, are 1/3 more useful. With these present findings, it is plausible to say that the answer the persons give, right after completing the first questionnaire are possibly predominated by the first answer they gave. This leads to a smaller improvement. At this time, the anchoring effect should be mentioned.

2.7.1 Other Studies on the Wisdom of Many in one mind

Similarly, to Herzog and Hertwig's paper from 2009 "the wisdom of many in one mind", authors like Müller-Trede (2011) among others also consider the one-person context thinking one step further in investigating reasons how to alter the setting and the execution of their survey.

They want to find out whether the type of question asked matters. With conducting a survey with the purpose of trying to find out whether three answers from one participant instead of two will elicit better results. They name the averaging as being beneficial for better answers. Better in the sense that they are closer to the right answer in question.

In his viewpoint he looks at the performance of judgements. It is being compared in different environments as well as how they control their prior judgments, when in total they make three judgements. Possible estimations that could be investigated are the year estimation, percentage share and, general numerical questions. In their terminology they use the terms potential gains and realized gains. A distinction between potential and realized gains has been made, whereas potential gains are defined as the averaged two estimates. Realized gains represent the third guess and their ultimate answer.

The findings of Müller-Trede's (2011) survey, can be summed up as followed: First of all, the findings of year estimations from Herzog and Hertwig (2009) have been affirmed but in for percentage values. Investigating how participants formed their third estimation, it turned out that the number of participants using averaging consistently is 10% which can be considered as being very low. (Müller-Trede, 2011, p.288) The survey consisted of two stages, a control group, with several conditions and sub-conditions present. Comparing different measures, they did not detect an improvement from participants' ultimate answer compared to their first and second answer: Agreeing to Rauhut and Lorenz's theory that more than two estimations from the same individual will lead to a decrease in returns proved to be right, especially for the general numerical inquiries. Also, the difficulty of the inquiries plays a role: If the questions could easily be answered, the accuracy gain rises whereas the returns do not fall at the same proportion. (Müller-Trede, 2011, p.290) So to conclude, an accuracy gain for the responses of general numerical inquiry is not worth to mention. For analyzing and then putting the acquired results in perspective, several equations represent the interpretation and connection of the described variables. For example, deviations between estimates and right answers have been calculated of the question asked, and potential gains were not detected.

2.7.2 Discussions of the existing results

In their article Rauhut and Lorenz (2011) took up the findings of the wisdom in one mind described in articles of Herzog and Hertwig (2009) and Vul and Pashler (2008) regarding this topic. From the viewpoint of Vul and Pashler (2008), the calculation of the number of persons needed compared to individual guesses with asking another person

Concerning T, they detected the following: Being able to calculate an estimator we would need the variance and “the population bias”. So, the estimator cannot be calculated due to a low number of available guesses from a single individual. This might evoke a difficulty from the authors’ point of view. Furthermore, the discussion of possible problems continues with the consideration of other factors and formulas used in the model, which I will leave out here (Rauhut and Lorenz, 2011).

The survey consisted of five knowledge questions, which can be answered with numbers and were rewarded for achieving good estimations close to the right answer.

They capture two statements of Vul and Pashler’s article “Measuring the Crowd Within”. Two statements were taken into consideration: First, asking questions one-person infinite times will on average not lead to any improvements of the answer. In that context the following T-values were shown: The T-values provided by Vul and Pashler’s study from 2008, reached 1.11 for the participants that were asked as second time right after answering the survey the first time and 1.32 subsequently for a delay of three weeks for the second attempt. The two T-values can be interpreted in a way to be a measure of exchanging the guess of the same person with another person’s guess. As Vul and Pashler’s words the exchange percentage can be defined as “representing is the proportion of an additional guess from another person that an additional guess from the same person is worth.” The conclusion of the immediate answer was that the value of 1.11 showed a benefit of approximately one tenth and delayed answers with the value of 1.32 subsequently represented a result originating from two guesses of one person that is 1/3 better compared to asking another person for his or her judgment.

Rauhut and Lorenz (2011) provide a T-value of 1.1 confirming the initial value of 1.11 from Vul and Pashler’s results of their survey. On average asking oneself two times is equal to asking 1.1 other persons. To find a T-value that represents the condition that a person is asked unlimited of times they provided two T-values calculated. One value they got was 1.28 and the other was 1.21. The first was calculated with a structured approach, the other one was calculated with an unordered calculating technique. 1.28 and 1.21

in any which case is lower than 1.32 so we can conclude that asking another individual is not better than asking one person a second time. (Rauhut and Lorenz, 2011, p.195)

Summing up we can say that the benefits of the averaging principle are mostly not recognized and not well understood by the participants asked. The use would use to improvements in their guesses. (Larrick & Soll, 2006 and Müller-Trede, 2011)

In addition, independency is the condition, which target/lead to better and accurate results. (Vul & Pashler (2008); Herzog & Hertwig (2009); Surowiecki (2004))

Furthermore Rauhut & Lorenz 2011 as well as Müller-Trede 2011 found out that the gain from the averaging method does not aim to an accuracy gain, if you apply it to more than 2 estimates of one individual. Müller-Trede (2011) even reinforced that “accuracy gains decreased substantially when averaging more than two estimates from the same judge.” (Müller-Trede, 2011, p.284) Positive accuracy gains can be detected from his survey, where four out of six questions confirmed such a benefit.

3 Problem Definition

From the theoretical foundation laid in the previous chapter, the effect “wisdom of crowds” leads to better results is very well established. Leveraging this fact, Herzog & Hertwig (2009) and Vul & Pashler (2008) found out that a similar benefit can be achieved by simulating a “wisdom of crowd” with a second estimate coming from the same person. So those two concepts improve the accuracy gain. Another approach chosen by Keck & Tang (2019) was to combine estimates generated through discrete cognitive processes in order to boost the estimate’s performance. When the estimates are independent from each other the aggregation of estimates yields better results. One such method to create independent estimates is to combine judgments, which are intuitive and analytical. Intuitive and analytical judgements evolve from different cognitive processes. Similarly, Sinclair & Ashkanasy (2005) also suggest a model “of integrated analytical and intuitive decision making” as well as Sauter (2005) can be counted as an advocate for the combination of intuitive and analytical estimates.

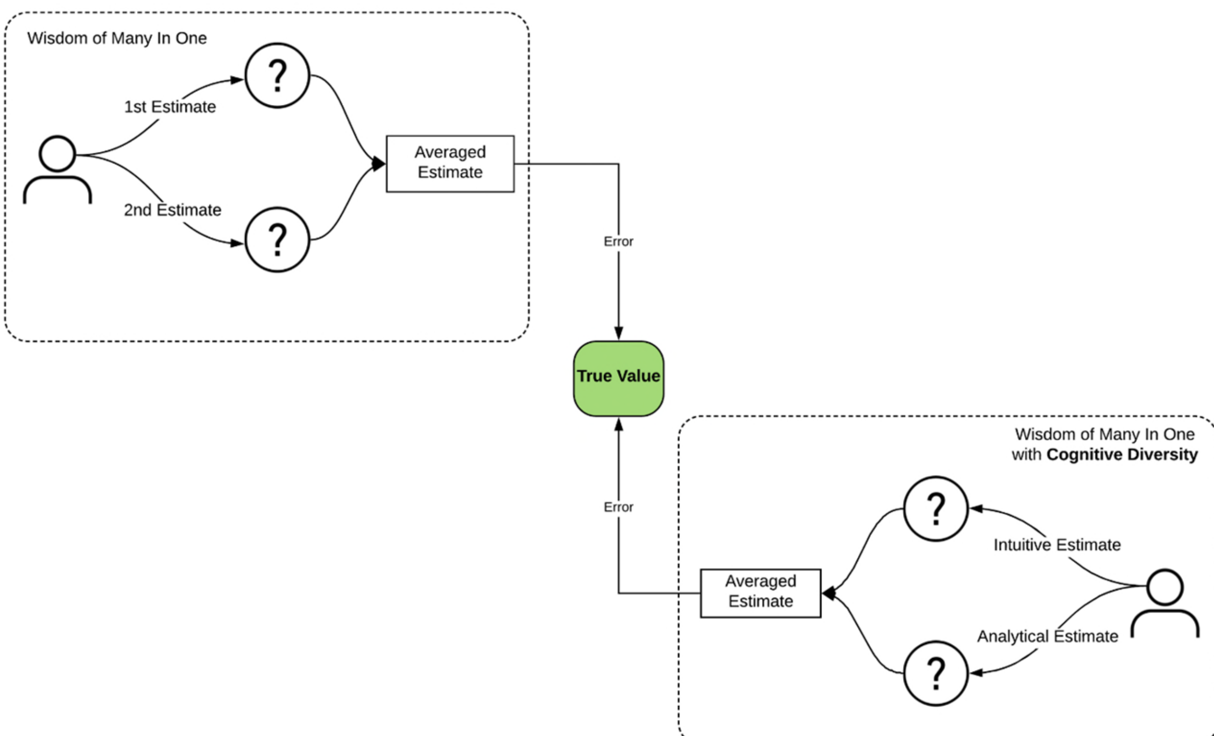

Figure 3.1: Conceptual representation of the Hypothesis

Taking both approaches and combining them together, we want to find out whether *"Aggregated estimates obtained through diverse cognitive process yields more accurate results than aggregated estimates obtained from a less diverse cognitive process."* This is our first hypothesis. Also, what is interesting is to know if there is a connection between the gender and the fact whether the participant made a "most accurate" "accurate" or "deviating" error. To find out whether the estimation was close to the right answer we will take the true value (green box in Figure 3.1), which is the right answer of the question, and subtract it from the average of the two guesses (averaged estimate). Through this we get the error of the estimate. If the error is small, that means that the estimation of the participant was very close to the true value and therefore good.

In order to validate the above hypotheses, we will conduct an empirical study and collect data, to see if there is a statistical relevance to the performance of the results yielded.

4 Empirical Study – Design & Methodology

Empirical studies are helpful tools to find out if certain theories can be proved. It can be a single or group experiments or it can be interviews. Existing literature offers a broad range of options on how to conduct different types of empirical studies. For our study, we used the approach suggested by Malone (2018) and adapted his step by step approach for designing and executing surveys. The following figure and section below give a brief overview of those steps.

Figure 4.1: Survey design steps (Malone, 2018, p.87)

Step 1: Supported by the literature review, we formed the Hypothesis to find out whether the theoretical assumptions are in accordance through the survey.

Step 2: Sampling strategy: The ideal participants would have different characteristics. A description of the participants is in section 3.1 & 4.3.

Step 3: We worked on a general design of questions first and detailed the sub-sections of the survey. More details in the section 3.3 & Appendix.

Step 4: A pilot of the survey was tested by using the 'pretest' feature of the survey tool, where the graphical and contextual parts of the survey was validated like the layout of the boxes and the Conditions of Groups etc.,

Step 5: To get the survey started, we sent out the link via message or e-mail.

Step 6: The collected data was available online which can be accessed via the homepage of soscisurvey.de. Additionally, it was downloaded and store on the local storage of the computer and on an external flash drive in order to prevent any data loss.

Step 7: Soscisurvey.de provides data download in either an SPSS file, CSV(Excel) or text document. The data itself, the variables, and the values come in separate files.

4.1 Identification of Groups

Based on the hypothesis, the experiment will consist of two groups.

Group 1 or ***Control group*** which will be presented with a set of 8 questions twice to get two estimates without any instructions. Let us call this group "no control group" = NCG

Group 2 or ***Target group*** which will be presented with the same set of 8 questions but with the following variations. For the first question set the participants will be asked to make an intuitive guess as well as an analytical guess, when they were presented the second set of questions. Let us call this group as "intuitive/ analytical group" = IAG.

With the NCG Group we want to simulate the wisdom of many in one mind. Two estimates that come from the same person. As shown in the literature, averaging those two will result in a better guess.

The two responses from NCG group will be NCG_1, NCG_2 and the Average denoted as Avg_{NCG}

The two responses from IAG will be IAG_1, IAG_2 and the Average denoted as Avg_{IAG}

4.2 Survey setup with “soscisurvey”

For this study, we chose an ‘online survey’ as a means to conduct the experiment. As there are plenty of technology tools out there for organizing and conducting online surveys, we evaluated quite a few of them including Survey Monkey, Microsoft Forms, Google Forms, Type Form, Soscisurvey etc., The comparison and evaluation criterions of these tools are not mentioned in this work as this is out of scope of this study. We finalized soscisurvey as the tool of choice due to various advantages that fits this study compared to other tools.

For our purpose of doing data analysis, a diverse set of data collection is important. With soscisurvey we can download three files: values, variables and data. All three were downloaded as csv files. In the file variables, we can find each question as a variable and the socio demographic values age, education and gender. They also offer other variables like the starting the time spent per question. In the values file the variables are put into categories. For example, you take gender, where 1 represents ‘Male’ participants and 2 represents female participants. The data file contains all the information from the files variables and values (categorized), as well as the guesses themselves.

4.3 Survey design

As an important element of this study is the cognitive diversity, the emphasis is to simulate the necessary cognitive processes by prompting the participants with the right information and instruction. Keeping this in mind, we started drafting a general framework of how the questions and instructions can be formulated. Though the questions themselves are one and the same for both groups, depending on the group, the participants will encounter different sequence of messages and directions to guide them through the survey.

Figure 4.2 shows the framework we designed. Group 1 setup: It consists of general instructions, followed by knowledge questions one to eight, then participants were asked about their socio-demographics information such as age, gender and education. Again, they were reminded of the general instructions, what they will be asked to answer, then they will answer the same set of knowledge questions again to record the second estimate.

Regarding Group 2, the IAG, before the start of the first estimate, the participants will be asked to judge quickly/intuitively by hinting them with the following instruction "Please make a quick guess not taking longer than 5 seconds." The second time, the instruction was: "Think of reasons why your first guess might be right or wrong. Think thoroughly before making the guess." This hint was repeatedly placed for all the questions.

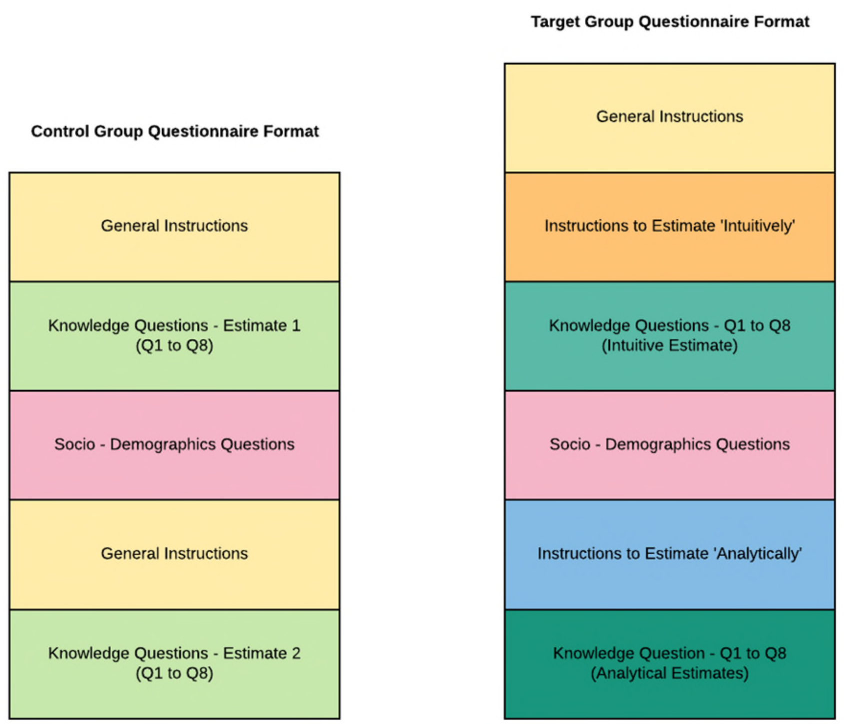

Figure 4.2: Framework for survey design

4.3.1 Questions

The core part of the survey were Knowledge questions. Using Herzog and Hertwig survey as a pointer for the study of this master thesis at hand, we chose eight questions from the factbook of Herzog and Hertwig's paper. The question set consist of eight questions, (highlighted with varying color shades of green boxes in the above figure)

Q1: What percent of the world's population lives in either China, India, or the European Union?

Q2: What percent of the world's airports are in the United States?

Q3: What percent of the world's population is Christian?

Q4: What percent of the worldwide income does the richest 10% of households earn?

Q5: What percentage of the world's countries have a higher life expectancy than the United States?

Q6: What percent of the earth's surface is covered by water?

Q7: The area of the USA is what percent of the area of the Pacific Ocean?

Q8: What percent of the world's telephone lines are in China, USA, or the European Union?

Participants were asked to type in a number into a box which was provided.

4.3.2 Restrictions

As all eight questions were percentage questions, the only possible range to type in the answer is from zero until hundred. It facilitates and saves us to cross out the participants.

With the socio-demographic questions we chose the following format for the survey:

AGE: A fill out box was provided to type in their age.

GENDER: They had to click on two suggested bullets-one for female, one for male.

EDUCATION: We suggested the participant to choose between categories from one to four, where 1 is Bachelor's Degree, 2 stands for Master's Degree, 3 is for PhD and 4 was Dr. If none of the above mentioned was applicable for their education, they could type in their completed degree themselves.

All in all, important to achieve a valuable dataset, are the following points:

- All questions should be answered, so that there are no missing data
- Participants should only be able to type in numbers reaching from zero to hundred
- To avoid missing data, we also cross out that participants could type in letters

4.3.3 Pretesting

Two pretests were created to see, what the actual survey would look like. Conducted were the eight questions with fill out boxes, as well as three sociodemographic questions. The eight questions were knowledge questions which were repeated a second time. After creating the link, filling out the questions, the survey offers the features to write comments in feedback boxes, which allowed us to improve the questions based on our notes. A first outline of the survey itself was created, in order to check whether all the requirements we need to work with the results and further analyze them are met.

4.3.4 Contacting Participants

Participants were contacted through different channels of communication: they were contacted per message or per email, whereas some received the first link and some the second. The email included either the link of Group 1 (*https://www.soscisurvey.de/projectmaster*) or Group 2 (*https://www.soscisurvey.de/IAG*)

Email template

Subject: Master thesis survey

Dear Participants,

The following survey will collect data which will be subject in my master thesis from the chair of strategic management. The topic of the thesis is "Aggregation of individual judgment". We will ask knowledge questions and collect percentage estimations.

Now please click on the link below to start the survey:

<one of the two links mentioned above>

Thank you very much for contributing to my survey!

Yours sincerely,

Clara Imani, BSc

0664/3437603

5 Results and Discussions

The homepage Soscisurvey offers three different formats to download the actual data: As a user it is possible to download the results as an SPSS file, CSV(Excel) or as a text document. We chose to download the data as a csv file, which provides in my opinion the clearest view. With different charts we can sort the information we gathered throughout the survey.

5.1 Dataset Analysis

In total we had 121 participants answering both surveys. 49 participants responded as Group 1 of the survey, whereas we collected initially 75 responses from Group 2. In the chart we see how many participants answered on which day, where the first graph represents the data collected per day for Group 1 and the second graph shows the same for Group 2. The grey portion of the chart represents who started but never completed.

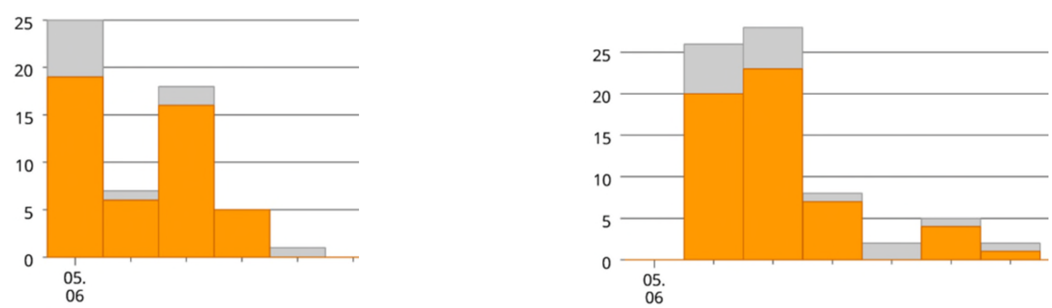

Figure 5.1: Number of Responses collected per day (Image from Soscisurvey.de dashboard)

It is to be mentioned that the downloaded data sheet from soscisurvey contains a categorization, whether the interview has been finished (reached last page) or not. One stands for finished and zero stands for not finished. So, it was easy for us to identify 'completed' datasets. Out of 49 participants of Group 1 we had to cross out fourteen participants, due to data issues, leaving valid data from 35 participants intact. In Group 2, from the 75 responses we randomly picked 35 responses to balance both the groups.

After identifying the data set for analysis, we performed a simple data exploration by examining the results. The raw dataset consists of several components: Sixteen estimations, the age, gender and education of the participant, Information about the answers to the questions, time they spent per question, time when the survey was submitted.

5.1.1 Modification of the variables

Grouping the age into three different categories in SPSS: Going to TRANSFORM -> Recode into different variables. Then we define class one with participants from 0 to 25 years with the value 1, class two are the participants from 25 to 50 years, and the last class is class three, which consists of participants above 50 years. Class three has the value 3.

The education is grouped in four classes. 1 stands for Bachelor's degree, Class 2 stands for Master's Degree, 3 is for PhD and 4 was Dr. If the participant has some other education that was not listed in one of the four classes, they could type in the form of education themselves.

5.2 Demographics

As we see in the pyramid charts below, there were more female participants in either of the two groups. Looking at the control group most participants fall in the age group of 24 to 26 years. Looking at the female section of the IAG Group there were 8 participants in the age group of 24 to 26, 6 participants in group of 27 to 29 and further 6 in their 60s. Regarding male participants, out of the thirty-five 4 belong to the age group 24 to 26 and 4 of them between 27 and 29. Though we collected the education details of the participants we leave those details out due to the little contribution it makes to this work.

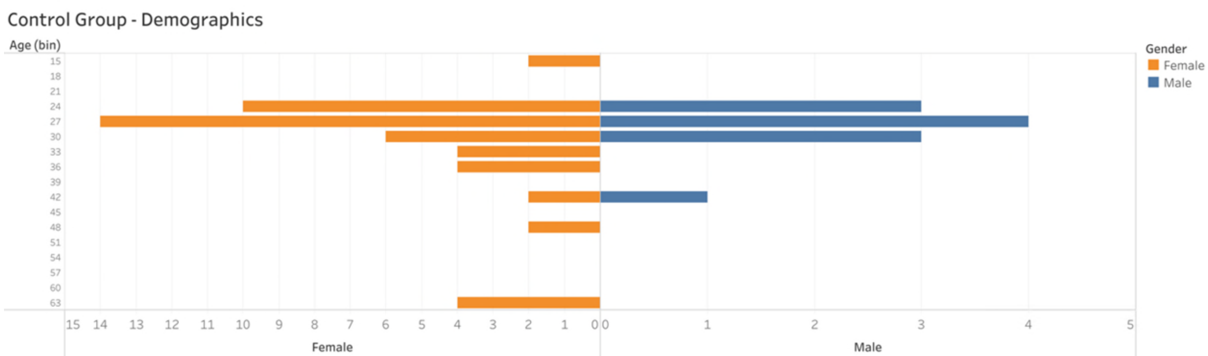

Figure 5.2: Control Group - Demographics pyramid chart

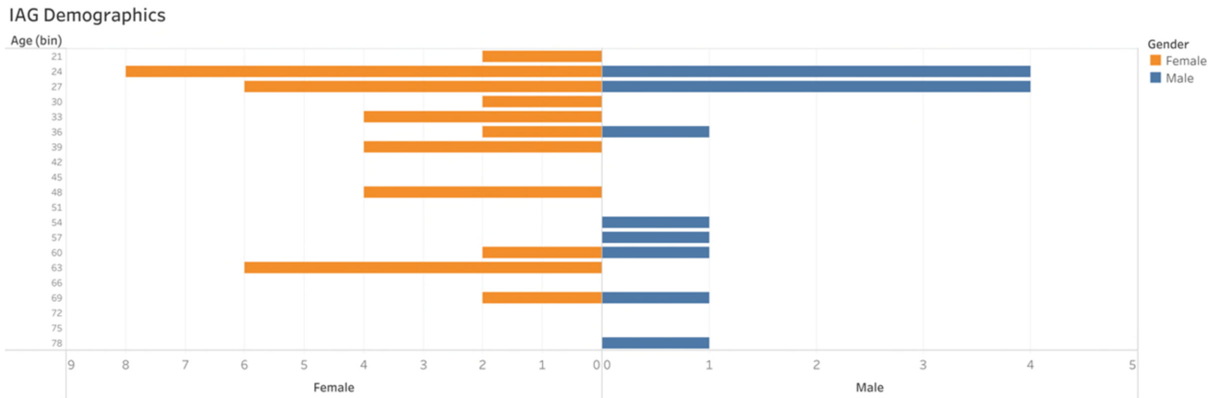

Figure 5.3: Target Group (IAG) - Demographics pyramid chart

5.3 Result Analysis

Now we have a look at the performance of the participants with respect to question one to eight. We calculated the median of two responses for every participant and averaged it across all the participants to see how the group performed. Figure 5.4 shows the results from two groups compared with true value.

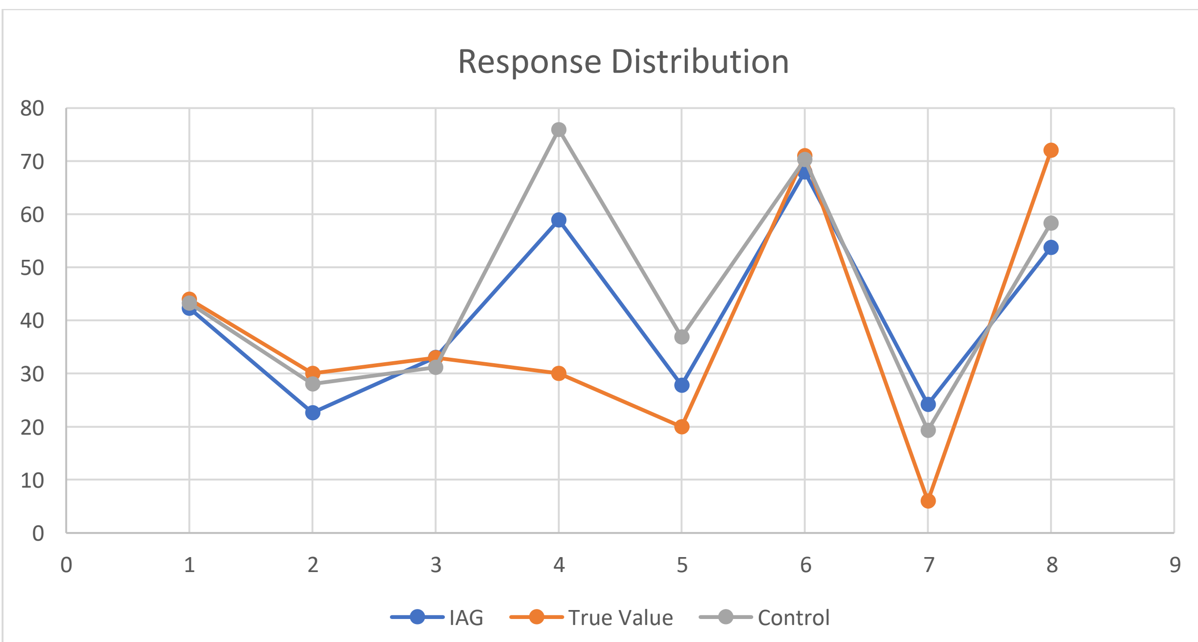

Figure 5.4: Response distribution

The x axis of the graph is the number of knowledge questions, whereas the y axis shows the percentages. The orange line indicates true value, which would be the right answer to the question. As the blue line represents the averaged response for each question by Group 2 (IAG), and the grey line represents the same for the Group 1(NCG). From the analysis, the following are our observations: The first thing that stands out, is Question 4. Question 4 was “What percent of the worldwide income does the richest 10% of households earn?”, we can see that Group 1, as well as Group 2, both deviate a lot from the true value. Three out of the eight questions showed estimations, that were most close to the true value: Question 1, Question 3 as well as Question 6. As the answers to the questions were percentages, the length of each box presents 10%. So, we can see that Question 1, 3 and 6 the difference is very minor. In Question 2 Group 1 performed better than Group 2 since the grey line is closer to the orange one. Question 5 Group 2 is closer to the true value than Group 1. Question 7 again shows results, where group 1, the grey line is closer to the true value.

5.3.1 Error Analysis

We had a closer look at the errors for each question from both the group.

The two responses from NIG group will be NIG_1, NIG_2 and the Average=Avg_{NIG}

The two responses from IAG will be IAG_1, IAG_2 and the Average=Avg_{IAG}

Error = $Avg_{(x)}$ – True Value *where* (x) represents averaged estimate of a question from the respective group.

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Control	Average	43.2714	28.0286	31.1714	75.9286	36.8571	70.2857	19.3000	58.2857
	True Value	44	30	33	30	20	71	6	72
	Error	-0.7286	-1.9714	-1.8286	45.9286	16.8571	-0.7143	13.3000	-13.7143

Table 5.1: Calculation of the Errors of Group 1, NCG

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
IAG	Average	42.3286	22.6143	33.0429	58.9286	27.7857	67.9286	24.1857	53.7286
	True Value	44	30	33	30	20	71	6	72
	Error	-1.6714	-7.3857	0.0429	28.9286	7.7857	-3.0714	18.1857	-18.2714

Table 5.2: Calculation of the Errors of Group 2, IAG

For the calculation of the average we take the data from the survey, which provides the guesses of the participants. Next is forming the average, which consists of the first and the second estimate of a participant divided by the number of estimates. This is done for each question. Respectively to Herzog and Hertwig's approach to find out the most accurate values we need to follow two further steps. First, we take the average values of each of the eight questions. For our estimates that means to sum up the eight averages and divide them by eight. Secondly, we subtract the values from the right answer. We call this value 'error'. Now we follow the same two steps for the second group.

For the first eight questions of Group 1 we see the calculated values in Table 5.1. In the first row, we see the Average answer of the thirty-five participants for question one to eight. The second row represents the right answers to the first eight knowledge questions. Subtracting the average from the right answer, which is the true value, leads us to the error.

In Table 5.2 we calculate the average of the first and the second answer of the survey, then take the right value and subtract it from the right value to reach the error. For this we take the estimates from Group 2, which did have instructions to follow, while answering the eight knowledge questions twice. For further details of the instructions I refer to the subchapter 4.1. Instructions.

Error_Average_Control	-0.7286	-1.9714	-1.8286	45.9286	16.8571	-0.7143	13.3000	-13.7143

Error_Average_IAG	-1.6714	-7.3857	0.0429	28.9286	7.7857	-3.0714	18.1857	-18.2714

Table 5.3: Error comparison

Table 5.3 shows the direct comparison of the errors. Comparing two values with each other we define one estimate to be more accurate, when it is smaller than the other one. As the errors are calculated as the right value minus the estimate itself, we will get some negative values. This is the case when the participants gave estimated that are bigger than the actual answer to the question. In other words, they overestimated the percentages. For this reason, we will only look at absolute values, when it comes to errors.

The first comparison shows that the error of question 1 of Group 1 is smaller than the error of question 1 of Group 2. In terms of value rounded to two decimals we get 0,73% for group 1 and 1,67% for group 2. Question two showed 1,97% for group 1 and the error in group 2 was 7,39%. Given these percentage numbers group 1 performed better. When we look at the errors of questions six, seven and eight Group 1 showed smaller errors than group 2: 0,71% was smaller than 3,07% for question 6, 13,30% was smaller than 18,19% when we look at question 7. Question eight showed 13,71% compared to group 2 with 18,27% of error.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
CONTROL	19.82435	18.30097	11.63915	19.51249	17.95975	12.31109	13.51644	20.45234
IAG	19.14455	14.81508	12.23025	31.91098	17.8611	11.13826	17.69724	22.18152

Table 5.4: Standard deviation of the error

These numbers suggest that group 1 with no instructions performed better than group 2 who were given with specific instructions. On the contrary, it can be stated that group 2 performed better than group 1 in the case of questions three, four and five. The absolute values for those are 0,04% compared to an error of 1,83% of group 1. Question four had an error of 45,92% in group 1 compared to group 2, which deviated with just 28,93%. Participants of question five in group 2 are in error of 7,79% whereas group 1 participants erred by 16,86%.

As an analyzing tool we will now find a suitable statistical test in order to interpret the data set.

5.4 Statistical Tests

The variables described in section 5.1.1, are now put into order to perform different statistical tests. We will run several Chi-squared tests. But first, we introduce a statistical decision tool that was helpful to find out which statistical test to run. On the top of Figure 5.5 set of questions are listed, which represent stages. We will now go through this path, marked as a red line, to conclude which test is applicable in the end. At the end of the decision tree, list of applicable statistical tests can be found. For our available dataset, we find ‘Chi-squared test’ as the most fitting one.

In the following section we will look at Chi-squared Test results. First the error of group 1 was divided into three categories "most accurate" "accurate" and "deviating". "Most accurate" means that the error is between zero and 5 percent, "accurate" means it is between five and twenty. An error that we call as "deviating" means that the error is above 20%.

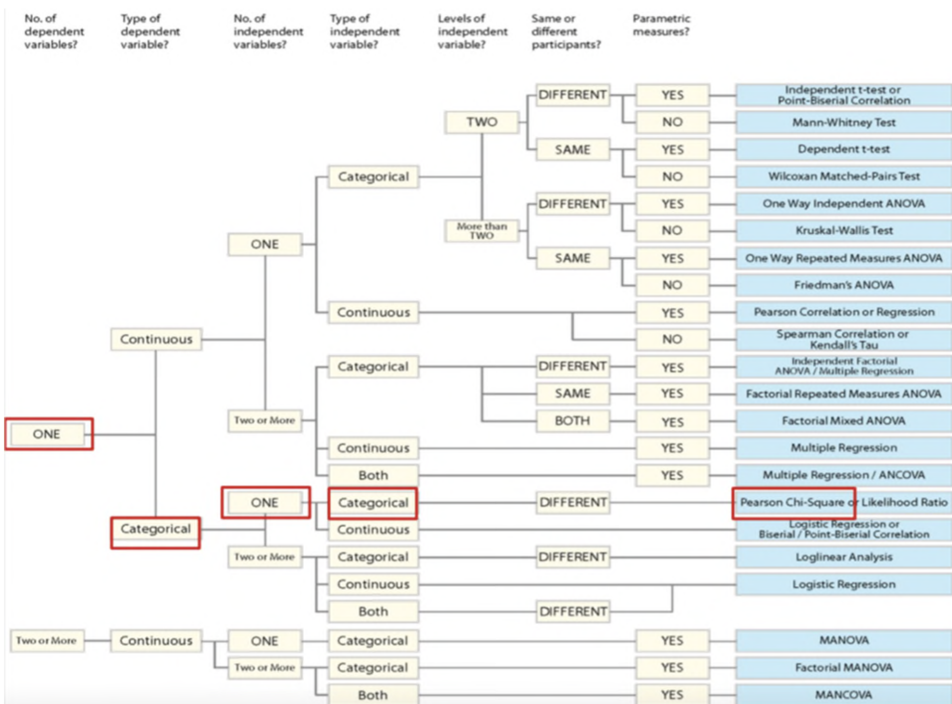

Figure 5.5: Statistical test selection chart (Taken from WU lecture slides)

The gender was categorized in female and male, where male is coded with 1 and female is coded with 2. We define the gender as the dependent and the error of group 1 as the independent variable. Both variables are categorical, and we look at samples with different participants. Our Null hypothesis is that there is no relation between the gender and the fact whether the participant made a "most accurate", "accurate" or "deviating" error. The alternative Hypothesis says that there is a relation between the gender and the fact, whether the participant made a "most accurate" "accurate" or "deviating" guess.

Considering the results from Group 1, which we named as NCG we have a look at Table 5.5 and can say that 34,1% of the male participants had errors in the range of zero to 5. 35,2% of the male participants gave answers with "accurate errors" which means that they only deviated 5 to 20 %. 30,7% of the participants gave answers, which had an error which was over 20%. Under the participants who made a guess that resulted in an error between zero and five percent, were 36,6% male and 63,4% female. Out of participants making either accurate or deviating answers slight under one third were male: 29,2% of the category accurate errors and 29,3% of the category deviating errors. With a non-significant p-value of 0,483 (Exact-two sided) we cannot accept the Null hypothesis but can accept the Alternative Hypothesis.

genderNCG * errorNCG_cath

Crosstab

			errorNCG_cath			
			most accurate	accurate	deviating	Total
genderNCG	Male	Count	30a	31a	27a	88
		Expected Count	25,8	33,3	28,9	88,0
		% within genderNCG	34,1%	35,2%	30,7%	100,0%
		% within errorNCG_cath	36,6%	29,2%	29,3%	31,4%
		% of Total	10,7%	11,1%	9,6%	31,4%
		Standardized Residual	,8	-,4	-,4	
	Female	Count	52a	75a	65a	192
		Expected Count	56,2	72,7	63,1	192,0
		% within genderNCG	27,1%	39,1%	33,9%	100,0%
		% within errorNCG_cath	63,4%	70,8%	70,7%	68,6%
		% of Total	18,6%	26,8%	23,2%	68,6%
		Standardized Residual	-,6	,3	,2	
Total		Count	82	106	92	280
		Expected Count	82,0	106,0	92,0	280,0
		% within genderNCG	29,3%	37,9%	32,9%	100,0%
		% within errorNCG_cath	100,0%	100,0%	100,0%	100,0%
		% of Total	29,3%	37,9%	32,9%	100,0%

Each subscript letter denotes a subset of errorNCG_cath categories whose column proportions do not differ significantly from each other at the ,05 level.

Table 5.5: SPSS results - NCG

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	1,431[a]	2	,489	,483		
Likelihood Ratio	1,410	2	,494	,490		
Fisher's Exact Test	1,422			,483		
Linear-by-Linear Association	1,005[b]	1	,316	,328	,179	,039
N of Valid Cases	280					

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 25,77.

b. The standardized statistic is 1,002.

Table 5.6: Chi-Square results - NCG

Symmetric Measures		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	,071	,489	,483
	Cramer's V	,071	,489	,483
N of Valid Cases		280		

Table 5.7: Symmetricity results - NCG

So, we can conclude that there is no connection between gender and the fact how good the estimates were. We measure this with the three categories: The best estimates are named as “most accurate”, the second best are “accurate” errors and the third best are “deviating”.

We can see in Table 5.7 how strong the connection between error of group 1 and the gender is. For an analysis of a 2x2 Table we would look at the Phi value. As we have three categories, we will use the Cramer’s V. If the value is zero it would mean that there is no connection whereas 1 means that there is a strong effect between those two variables. As Cramer’s V is 0,071 we cannot speak of any or a close connection. It tells us that there is no connection when the value is zero and 1 means that there is a strong connection between the two variables.

The comparison of the error of Group 2, which we named IAG, which stands for Individual and analytical Group, we have a look at the SPSS outputs. The Crosstab shows that 33% of the male participants had errors between zero and five percent. 35,7% of the male participants gave answers with “accurate errors” which means that they only erred between 5 and 20 %. 37,2% of the participants gave answers, which deviated over 20%. Under the participants with “most accurate” errors were 50% female and 50% male. Out of participants giving “accurate” and “deviating” answers around one third were male: 35,7% of the category accurate errors and 37,2% of the category deviating errors. The exact two-sided p-value is 0,122. Since it is greater than 0,05 the Null-hypothesis needs to be rejected and we accept the Alternative-Hypothesis. Looking at the table symmetric measures we can say that the connection between error of group 2 and the gender is non-existent. Again, we will look at the Cramer’s V, since we have three categories. 0,123 is very close to zero, so we cannot speak of a connection. Looking at the above described results, we can conclude that there is no connection between the gender and the fact how good the estimate is. To define how good the estimate is we categorized into “most accurate”, “accurate” and “deviating”. As the category “most accurate” includes errors reaching from zero till 5, these are the best estimates, which are very close to the true value. Second best are the estimates in the category “accurate”. “accurate” includes errors between 5 and 20. We speak of “deviating” errors when the error is not that close to the right answer. In this case we speak about errors reaching over 20% errors.

genderIAG * errorIAG_cath

Crosstab

			errorIAG_cath			
			most accurate	accurate	deviating	Total
genderIAG	Male	Count	37a	40a	35a	112
		Expected Count	29,6	44,8	37,6	112,0
		% within genderIAG	33,0%	35,7%	31,3%	100,0%
		% within errorIAG_cath	50,0%	35,7%	37,2%	40,0%
		% of Total	13,2%	14,3%	12,5%	40,0%
		Standardized Residual	1,4	-,7	-,4	
	Female	Count	37a	72a	59a	168
		Expected Count	44,4	67,2	56,4	168,0
		% within genderIAG	22,0%	42,9%	35,1%	100,0%
		% within errorIAG_cath	50,0%	64,3%	62,8%	60,0%
		% of Total	13,2%	25,7%	21,1%	60,0%
		Standardized Residual	-1,1	,6	,3	
Total		Count	74	112	94	280
		Expected Count	74,0	112,0	94,0	280,0
		% within genderIAG	26,4%	40,0%	33,6%	100,0%
		% within errorIAG_cath	100,0%	100,0%	100,0%	100,0%
		% of Total	26,4%	40,0%	33,6%	100,0%

Each subscript letter denotes a subset of errorIAG_cath categories whose column proportions do not differ significantly from each other at the ,05 level.

Table 5.8: SPSS results – IAG

The comparison of the error of Group 2, which we named IAG, which stands for Individual and analytical Group, we have a look at the SPSS outputs. The Crosstab shows that 33% of the male participants had errors between zero and five percent. 35,7% of the male participants gave answers with “accurate errors” which means that they only erred between 5 and 20 %. 37,2% of the participants gave answers, which deviated over 20%. Under the participants with “most accurate” errors were 50% female and 50% male.

Out of participants giving “accurate” and “deviating” answers around one third were male: 35,7% of the category accurate errors and 37,2% of the category deviating errors. The exact two-sided p-value is 0,122.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	4,240[a]	2	,120	,122		
Likelihood Ratio	4,192	2	,123	,125		
Fisher's Exact Test	4,177			,125		
Linear-by-Linear Association	2,492[b]	1	,114	,133	,067	,018
N of Valid Cases	280					

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 29,60.

b. The standardized statistic is 1,579.

Table 5.9: Chi-Square results - IAG

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	,123	,120	,122
	Cramer's V	,123	,120	,122
N of Valid Cases		280		

Table 5.10: Symmetricity results - IAG

Since it is greater than 0,05 the Null-hypothesis needs to be rejected and we accept the Alternative-Hypothesis. Looking at the table symmetric measures we can say that the connection between error of group 2 and the gender is not really existent. Again, we will look at the Cramer’s V, since we have three categories. 0,123 is very close to zero, so we cannot speak of a connection. Looking at the above described results, we can conclude that there is no connection between the gender and the fact how good the estimate is. To define how good the estimate is we categorized into “most accurate”, “accurate” and “deviating”. As the category “most accurate” includes errors reaching from zero till 5, these are the best estimates, which are very close to the true value. Second best are the estimates in the category “accurate”. “accurate” includes errors between 5 and 20. We speak of “deviating” errors when the error is not that close to the right answer. In this case we speak about errors reaching over 20% errors.

5.5 Feedback from the participants

Five out of seventy participants that answered the survey, commented that they did not understand the first question. Two of them answered all the questions but reached out to me after completing the survey saying that there must be a mistake in the survey because the same questions were asked again. Another three participants were eager to find out the results and how good their estimations were.

6 Conclusion

Through data analysis shared in the previous chapter, we investigated the errors of the guesses, as well as performing a Chi-squared test. We can state the following findings:

Looking at the calculation of the errors of Group 1 and Group 2 five out of eight questions, participants of Group 1, the Control Group performed better. So, we can say that without any instructions the guesses were closer to the true value as the guesses from Group 2 IAG. What we did was comparing two values with each other we define one estimate to be more accurate, when it is smaller than the other one.

So, our first hypothesis that aggregated estimates obtained through diverse cognitive process yields more accurate results than aggregated estimates obtained from a less diverse cognitive process cannot be supported by the data we found.

Secondly from the execution of the Chi-squared test we can say that there is no relation between the gender and the fact whether the participant made a “most accurate”, “accurate” or “deviating” error., since the p-value in both groups was larger than 0,05. (0,483 for Group 1 and 0,122 for Group 2) We performed two Chi-Squared tests, which both led to rejecting the Null-hypothesis, saying that there is no connection between the variables. For Group 1 we got a Cramer’s V value of 0,071 and for Group 2 a value of 0,123, as zero indicates that there is no connection.

6.1 Limitations and Future research

During the data evaluation a lot of pre-calculations had to be done before converting the data into the SPSS file. As we chose a quantitative data collection tool, an online survey, that beside the advantage of this type of collection-large number of participants can be achieved quickly as well as a large scope can be reached, there are also disadvantages of an online survey: The information about the population itself can be insufficient as well as the anonymity of the participants might lead to dubious answers. Also, the collected data might contain a bias that arises from the self-selection of participants. (Homburg, 2003, p. 271).

In addition, the survey was completely voluntary with no rewards or incentives which will act as an intrinsic motivation. Despite clear instructions given, we could see that participants took longer time than expected to answer the intuitive questions, from the time data which recorded the time spent per page. This provides a possibility of performing the same study in a controlled environment, to see if our findings are aligning.

6.2 Final note

Through the course Strategic Decision-Making, I got an introduction of decision-making processes. This attained my interest to research further on decision making and choice. It is of special interest to me to highlight intuitive and analytical decision making. I find it fascinating what happens consciousness in the process of deciding and what factors may influence the final decision or judgment. Particularly the research about intuition and judgment are my interest. Writing this thesis, I learned a lot about the underlying processes as well as how to improve estimates. Doing the research enchants me to dig deeper into the existing literature about topics with psychological influence factors of judgments and research about its frequency.

References

Betsch, C., Betsch, T., & Plessner, H. (2009). *Intuition in judgment and decision making*. Psychology Press.

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *International Journal Of Forecasting*, *5*(4), 559-583. https://doi.org/10.1016/0169-2070(89)90012-5

Cooke, R., & Shrader-Frechette, K. (1991). *Experts in Uncertainty*. Oxford University Press.

Dhami, M., & Mumpower, J. (2018). Kenneth R. Hammond’s contributions to the study of judgment and decision making. *Judgement And Decision Making*, *13*(1), 1-22.

Dutta, T., Mandal, M., & Kumar, S. (2012). *Bias in human behavior*. Hauppauge, N.Y.: Nova Science Publisher's.

Englich, B., & Soder, K. (2009). Moody experts – how mood and expertise influence judgmental anchoring. *Judgmental And Decision Making*, *4*, 41-50. Retrieved 12 June 2020, from.

Epley, N., & Gilovich, T. (2006). The Anchoring-and-Adjustment Heuristic. *Psychological Science*, *17*(4), 311-318. doi: 10.1111/j.1467-9280.2006.01704.x

Furnham, A., & Boo, H. (2011). A literature review of the anchoring effect. *The Journal Of Socio-Economics*, *40*(1), 35-42. https://doi.org/10.1016/j.socec.2010.10.008

Gilovich, T., Griffin, D., & Kahneman, D. (2009). *Heuristics and biases*. Cambridge: Cambridge Univ. Press.

Goldstein, D., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75-90. doi: 10.1037/0033-295x.109.1.75

Hammond, K., Todd, F., Wilkins, M., & Mitchell, T. (1966). Cognitive conflict between persons: Application of the “lens model” paradigm. *Journal Of Experimental Social Psychology*, *2*(4), 343-360. https://doi.org/10.1016/0022-1031(66)90027-8

Hens, T., & Rieger, M. (2016). *Financial economics*. Springer.

Herzog, S., & Hertwig, R. (2009). The Wisdom of Many in One Mind. *Psychological Science*, *20*(2), 231-237. doi: 10.1111/j.1467-9280.2009.02271.x

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211-237. doi: 10.1037/a0025940

Homburg, C. (2017). Marketingmanagement Strategie - Instrumente - Umsetzung - Unternehmensführung (6th ed.).

Individual Differences in Judgement and Decision-Making. (2016). https://doi.org/10.4324/9781315636535

Jacowitz, K., & Kahneman, D. (1995). Measures of Anchoring in Estimation Tasks. *Personality And Social Psychology Bulletin*, *21*(11), 1161-1166. doi: 10.1177/01461672952111004

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697-720. https://doi.org/10.1037/0003-066x.58.9.697

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515-526. https://doi.org/10.1037/a0016755

Kamleitner, B., Hoelzl, E., & Kirchler, E. (2012). Credit use: Psychological perspectives on a multifaceted phenomenon. *International Journal Of Psychology*, *47*(1), 1-27. https://doi.org/10.1080/00207594.2011.628674

Keck, S., & Tang, W. (2019). Enhancing the Wisdom of the Crowd With Cognitive Process Diversity: The Benefits of Aggregating Intuitive and Analytical Judgments. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3319676

Kirchler, E. (2011). *Wirtschaftspsychologie*. Hogrefe.

Larrick, R., & Soll, J. (2006). Intuitions About Combining Opinions: Misappreciation of the Averaging Principle. *Management Science*, *52*(1), 111-127. https://doi.org/10.1287/mnsc.1050.0459

Lopes, L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin Of The Psychonomic Society*, *23*(6), 509-512. doi: 10.3758/bf03329868

Müller-Trede, J. (2011). Repeated judgment sampling: Boundaries. *Judgment And Decision Making*, *6*(4), 283-294. Retrieved 13 June 2020, from.

Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior And Human Decision Processes*, *98*(2), 133-143. doi: 10.1016/j.obhdp.2004.12.002

Oppenheimer, D. (2003). Not so fast! (and not so frugal!): rethinking the recognition heuristic. *Cognition*, *90*(1), B1-B9. doi: 10.1016/s0010-0277(03)00141-0

Parnell, G., Bresnick, T., Tani, S., & Johnson, E. (2013). *Handbook of decision analysis*. John Wiley & Sons.

Quattrone, G., Lawrence, C., Finkel, S., & Andrus, D. (1984). Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints. *Unpublished Manuscript, Stanford University, Stanford, CA.*

Rauhut, H., & Lorenz, J. (2011). The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions. *Journal Of Mathematical Psychology*, *55*(2), 191-197. https://doi.org/10.1016/j.jmp.2010.10.002

Sauter, V. (1999). *Intuitive decision-making p109-115*. Association for Computing Machinery.

Schnell, R. (2019). Psychologische Grundlagen. In: Survey-Interviews. Studienskripten zur Soziologie. Springer VS.

Sinclair, M., & Ashkanasy, N. (2005). Intuition: Myth or a Decision-Making Tool?. *Management Learning*, *36*(3), 353-370. https://doi.org/10.1177/1350507605055351

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal Of Personality And Social Psychology*, *73*(3), 437-446. doi: 10.1037/0022-3514.73.3.437

Surowiecki, J. (2004). The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. *Choice Reviews Online*, *42*(03), 42-1645-42-1645. https://doi.org/10.5860/choice.42-1645

Svenson, O. (2016). Towards a framework for human judgements of quantitative information: the numerical judgement process, NJP model. *Journal Of Cognitive Psychology*, *28*(7), 884-898. https://doi.org/10.1080/20445911.2016.1188822

Toplak, M., & Weller, J. (2017). *Individual differences in judgement and decision-making*.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124-1131. doi: 10.1126/science.185.4157.1124

Vul, E., & Pashler, H. (2008). Measuring the Crowd Within. *Psychological Science*, *19*(7), 645-647. https://doi.org/10.1111/j.1467-9280.2008.02136.x

Wilson, T., Houston, C., Etling, K., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal Of Experimental Psychology: General*, *125*(4), 387-402. doi: 10.1037/0096-3445.125.4.38

Yap, S., Wortman, J., Anusic, I., Baker, S., Scherer, L., Donnellan, M., & Lucas, R. (2017). The effect of mood on judgments of subjective well-being: Nine tests of the judgment model. *Journal Of Personality And Social Psychology*, *113*(6), 939-961. https://doi.org/10.1037/pspp0000115

List of Figures

Figure 2.1: Three steps of Anchoring (Lopes, 1985) 10
Figure 3.1: Conceptual representation of the Hypothesis 22
Figure 4.1: Survey design steps (Malone, 2018, p.87) 24
Figure 4.2: Framework for survey design 27
Figure 5.1: Number of Responses collected per day (Image from Soscisurvey.de dashboard) 30
Figure 5.2: Control Group - Demographics pyramid chart 31
Figure 5.3: Target Group (IAG) - Demographics pyramid chart 32
Figure 5.4: Response distribution 32
Figure 5.5: Statistical test selection chart (Taken from WU lecture slides) 36

List of Tables

Table 5.1: Calculation of the Errors of Group 1, NCG 33
Table 5.2: Calculation of the Errors of Group 2, IAG 33
Table 5.3: Error comparison 34
Table 5.4: Standard deviation of the error 35
Table 5.5: SPSS results - NCG 37
Table 5.6: Chi-Square results - NCG 37
Table 5.7: Symmetricity results - NCG 38
Table 5.8: SPSS results – IAG 39
Table 5.9: Chi-Square results - IAG 40
Table 5.10: Symmetricity results - IAG 40

Appendix

Questionnaire - Control group

soSci
oFb - der onlineFragebogen

0% completed

Thank you for taking the time to participate in this survey.

- You will now be asked to answer 8 knowledge questions
- Since all the questions are percentage estimations, please enter your answers in the range of 0 to 100%
- It will take you approximately 5 to 7 minutes to complete the survey

Thank you very much!

Next

B.Sc. Clara Imani – 2020

5% completed

1. What percent of the world's population lives in either China, India, or the European Union?

GUESS

Next

10% completed

2. What percent of the world's airports are in the United States?

GUESS

Next

14% completed

3. What percent of the world's population is Christian?

GUESS

Next

19% completed

4. What percent of the worldwide income does the richest 10% of households earn?

GUESS

Next

24% completed

5. What percentage of the world's countries have a higher life expectancy than the United States?

GUESS

Next

29% completed

6. What percent of the earth's surface is covered by water?

GUESS

Next

33% completed

7. The area of the USA is what percent of the area of the Pacific Ocean?

GUESS

Next

38% completed

8. What percent of the world's telephone lines are in China, USA, or the European Union?

GUESS

Next

43% completed

9. How old are you?

Next

48% completed

10. Gender:

Gender:

[Please choose]

[Please choose]
Male
Female

Next

52% completed

11. What is your highest achieved education?

Next

57% completed

Thank you for finishing the first part of the survey.

The following applies:

- You will now be asked to answer 8 knowledge questions
- Since all the questions are percentage estimations, please enter your answers in the range of 0 to 100%
- It will take you approximately 5 to 7 minutes to complete the survey

Thank you very much! Clara Imani BSc

Next

62% completed

12. What percent of the world's population lives in either China, India, or the European Union?

GUESS

Next

67% completed

13. What percent of the world's airports are in the United States?

GUESS

Next

71% completed

14. What percent of the world's population is Christian?

GUESS

Next

76% completed

15. What percent of the worldwide income does the richest 10% of households earn?

GUESS

Next

81% completed

16. What percentage of the world's countries have a higher life expectancy than the United States?

GUESS

Next

86% completed

17. What percent of the earth's surface is covered by water?

GUESS

Next

90% completed

18. The area of the USA is what percent of the area of the Pacific Ocean?

GUESS

Next

95% completed

19. What percent of the world's telephone lines are in China, USA, or the European Union?

GUESS

Next

soSci
oFb - der onlineFragebogen

Thank you for completing this questionnaire!

We would like to thank you very much for helping us.

Your answers were transmitted, you may close the browser window or tab now.

Questionnaire – Target Group (Intuitive / Analytical Group)

soSci
oFb - der onlineFragebogen

0% completed

Thank you for taking the time to participate in this survey.

- You will now be asked to answer 8 knowledge questions
- Since all the questions are percentage estimations, please enter your answers in the range of 0 to 100%
- It will take you approximately 5 to 7 minutes to complete the survey

Thank you very much! Clara Imani BSc

Next

B.Sc. Clara Imani, Universität Wien – 2020

4% completed

Please make a quick guess not taking longer than 5 seconds per question.

Answer the question intuitively.

Next

9% completed

1. What percent of the world's population lives in either China, India, or the European Union?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

13% completed

2. What percent of the world's airports are in the United States?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

17% completed

3. What percent of the world's population is Christian?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

22% completed

4. What percent of the worldwide income does the richest 10% of households earn?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

26% completed

5. What percentage of the world's countries have a higher life expectancy than the United States?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

30% completed

6. What percent of the earth's surface is covered by water?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

35% completed

7. The area of the USA is what percent of the area of the Pacific Ocean?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

39% completed

8. What percent of the world's telephone lines are in China, the USA, or the European Union?

Please make a quick guess taking no longer than 5 seconds

GUESS

Next

43% completed

9. How old are you?

Next

48% completed

10. Gender:

[Please choose]
[Please choose]
Male
Female

Next

52% completed

11. What is your highest achieved education?

1
PhD
Dr.
Master's degree
Bachelor's degree

57% completed

Thank you for finishing the first part of the survey.

The following applies to the second part of the survey.

- You will now be asked to answer 8 knowledge questions
- Since all the questions are percentage estimations, please enter your answers in the range of 0 to 100%
- It will take you approximately 5 to 7 minutes to complete the survey

Thank you very much! Clara Imani BSc

Next

61% completed

Please think of reasons why your first guess might be right or wrong.

Think thoroughly before making the guess.

Next

65% completed

12. What percent of the world's population lives in either China, India, or the European Union?

Please think thoroughly before answering

GUESS

Next

70% completed

13. What percent of the world's airports are in the United States?

Please think thoroughly before answering

GUESS

Next

74% completed

14. What percent of the world's population is Christian?

Please think thoroughly before answering

GUESS

Next

78% completed

15. What percent of the worldwide income does the richest 10% of households earn?

Please think thoroughly before answering

GUESS

Next

83% completed

16. What percentage of the world's countries have a higher life expectancy than the United States?

Please think thoroughly before answering

GUESS

Next

87% completed

17. What percent of the earth's surface is covered by water?

Please think thoroughly before answering

GUESS

Next

91% completed

18. The area of the USA is what percent of the area of the Pacific Ocean?

Please think thoroughly before answering

GUESS

Next

96% completed

19. What percent of the world's telephone lines are in China, the USA, or the European Union?

Please think thoroughly before answering

GUESS

Next

Thank you for completing this questionnaire!

We would like to thank you very much for helping us.

Your answers were transmitted, you may close the browser window or tab now.

Responses – Control Group

CASE	STARTED	A109_01	A110	A111	A111s	A101_01	A102_01	A103_01	A104_01	A105_01	A106_01	A107_01	A108_01
30	6/5/2020 17:36	24	2	1		25	20	50	99	50	67	33	80
33	6/5/2020 17:40	31	2	-2	Magister	60	75	40	60	20	80	30	90
39	6/5/2020 17:55	29	2	2		30	14	20	85	65	60	10	25
43	6/5/2020 18:04	25	1	1		30	2	15	90	40	70	20	60
44	6/5/2020 18:05	26	2	2		30	15	40	90	50	95	10	65
47	6/5/2020 18:13	64	2	4		8	10	50	95	40	70	10	10
48	6/5/2020 18:25	26	2	1		15	40	30	80	30	80	30	20
49	6/5/2020 18:34	64	2	4		10	20	50	95	40	70	10	30
59	6/5/2020 20:05	35	2	2		20	1	20	90	30	90	30	70
67	6/6/2020 11:04	26	2	1		10	10	20	40	40	60	25	60
73	6/6/2020 14:52	29	1	1		60	20	25	95	20	70	10	70
75	6/6/2020 16:00	25	1	2		50	25	33	60	35	70	5	80
76	6/6/2020 16:16	28	1	1		70	30	25	80	30	75	20	50
79	6/6/2020 23:51	28	2	-2	Dr. med	33	30	35	90	15	80	3	95
80	6/7/2020 9:22	37	2	1		55	65	45	80	75	70	8	70
84	6/7/2020 10:19	31	2	2		45	30	30	90	20	70	7	50
96	6/7/2020 13:11	35	2	2		40	30	30	50	60	70	60	80
97	6/7/2020 13:14	32	1	-2	Ingenieur	40	8	30	85	30	70	10	38
98	6/7/2020 13:14	28	1	1		50	5	1	95	20	71	10	81
99	6/7/2020 13:25	32	1	2		70	12	39	77	35	71	5	65
106	6/7/2020 16:41	32	2	1		30	25	12	60	60	70	10	70
112	6/7/2020 18:30	42	1	-2	Matura	30	40	30	80	10	65	25	50
113	6/7/2020 18:41	29	2	-2	Master	60	50	40	20	10	20	10	60
114	6/7/2020 18:42	44	2	-2	primary school	60	50	30	50	20	70	30	40
115	6/7/2020 18:46	15	2	-2	8th grade	27	40	55	73	75	75	42	42
116	6/7/2020 18:52	36	2	2		20	20	40	80	40	70	10	60
122	6/7/2020 21:43	29	2	1		50	20	33	50	25	70	25	70
124	6/7/2020 22:43	29	2	1		20	20	35	90	15	75	60	75
125	6/7/2020 23:21	27	1	4		25	7	30	97	30	66	4	20
126	6/7/2020 23:53	25	1	1		60	40	30	95	20	75	5	70
127	6/8/2020 8:30	26	2	2		80	60	20	60	50	80	30	70
128	6/8/2020 8:39	32	1	1		45	10	35	60	30	60	30	40
129	6/8/2020 9:12	50	2	2		70	60	30	90	20	85	20	85
131	6/8/2020 13:06	28	2	-2	Diploma	40	17	27	70	55	67	8	60
132	6/8/2020 16:19	28	2	-2	Diploma	75	22	15	80	60	80	20	75

CASE	LASTDATA	A109_01	A110	A111	A111s	A115_01	A116_01	A117_01	A118_01	A119_01	A120_01	A121_01	A122_01
30	6/5/2020 17:40	24	2	1		20	20	50	99	70	67	33	80
33	6/5/2020 17:45	31	2	-2	Magister	60	70	60	60	30	80	40	80
39	6/5/2020 17:58	29	2	2		30	15	20	80	60	60	10	25
43	6/5/2020 18:07	25	1	1		30	2	15	90	40	70	30	60
44	6/5/2020 18:08	26	2	2		30	15	40	90	40	95	10	15
47	6/5/2020 18:21	64	2	4		10	30	50	95	40	70	10	20
48	6/5/2020 18:32	26	2	1		15	30	30	80	30	80	30	40
49	6/5/2020 18:38	64	2	4		10	20	50	95	40	70	10	30
59	6/5/2020 20:09	35	2	2		30	20	30	90	30	90	30	80
67	6/6/2020 11:08	26	2	1		10	10	20	30	20	60	30	60
73	6/6/2020 14:57	29	1	1		60	20	25	95	20	70	10	70
75	6/6/2020 16:05	25	1	2		50	25	33	65	35	70	5	80
76	6/6/2020 16:19	28	1	1		70	30	25	80	30	75	20	50
79	6/6/2020 23:57	28	2	-2	Dr. med	30	30	45	90	15	80	2	94
80	6/7/2020 9:30	37	2	1		40	60	35	80	80	65	8	70
84	6/7/2020 10:40	31	2	2		45	30	30	90	30	70	7	70
96	6/7/2020 13:18	35	2	2		80	30	30	50	80	70	70	80
97	6/7/2020 13:37	32	1	-2	Ingenieur	40	8	30	85	35	70	10	40
98	6/7/2020 13:19	28	1	1		50	5	1	95	20	72	10	85
99	6/7/2020 13:31	32	1	2		70	12	39	70	35	71	5	15
106	6/7/2020 16:50	32	2	1		30	25	10	60	50	70	10	70
112	6/7/2020 18:41	42	1	-2	Matura	70	50	25	80	10	65	40	50
113	6/7/2020 18:44	29	2	-2	Master	60	60	50	20	10	20	20	50
114	6/7/2020 18:48	44	2	-2	primary school	60	50	40	50	20	70	30	40
115	6/7/2020 18:59	15	2	-2	8th grade	36	42	29	45	75	75	23	25
116	6/7/2020 19:10	36	2	2		20	20	40	80	40	70	10	70
122	6/7/2020 21:45	29	2	1		50	20	33	50	20	70	25	70
124	6/7/2020 23:02	29	2	1		65	20	35	90	40	60	10	75
125	6/7/2020 23:31	27	1	4		25	7	30	95	30	66	5	20
126	6/7/2020 23:55	25	1	1		60	40	30	95	30	75	10	60
127	6/8/2020 8:35	26	2	2		80	70	20	70	50	80	30	70
128	6/8/2020 8:44	32	1	1		45	30	30	45	20	30	40	60
129	6/8/2020 9:19	50	2	2		75	60	25	95	25	80	15	70
131	6/8/2020 13:19	28	2	-2	Diploma	55	18	27	70	55	67	8	60
132	6/8/2020 16:23	28	2	-2	Diploma	75	25	10	80	60	80	20	70

Responses – Target Group (Intuitive / Analytical Group)

CASE	STARTED	DS01_01	DS02	DS03	DS03s	IE01_01	IE02_01	IE03_01	IE04_01	IE05_01	IE06_01	IE07_01	IE08_01
25	6/6/2020 15:55	26	2	1		40	5	20	2	2	80	10	80
27	6/6/2020 16:00	22	2	-2	Matura	20	80	50	40	50	70	10	90
30	6/6/2020 16:14	26	2	1		30	15	30	20	20	90	5	20
34	6/6/2020 16:48	38	2	1		70	20	40	80	50	60	30	60
44	6/6/2020 18:21	64	2	3		20	30	25	90	40	75	25	20
46	6/6/2020 18:30	33	2	2		35	25	27	85	27	42	20	75
52	6/6/2020 19:06	29	1	-2	Bsc	35	30	35	95	60	65	15	65
54	6/6/2020 19:08	50	2	-2	Akademie	30	30	20	20	30	60	40	50
55	6/6/2020 19:15	26	2	-2	Bachelor	20	20	21	5	40	80	3	15
56	6/6/2020 19:26	39	2	1		40	25	50	40	70	66	20	70
59	6/6/2020 19:50	54	1	-2	University degree	50	15	30	80	25	70	20	75
63	6/6/2020 21:03	41	2	2		30	30	40	20	20	60	40	80
75	6/6/2020 21:41	28	1	2		20	18	25	95	8	75	13	20
77	6/6/2020 21:51	38	1	2		75	10	25	90	3	75	5	50
79	6/6/2020 22:11	28	2	1		15	1	45	15	10	45	5	15
85	6/6/2020 22:41	26	1	1		60	15	40	80	10	70	50	90
87	6/6/2020 23:31	70	1	4		20	10	25	0	25	70	5	30
91	6/7/2020 8:13	29	2	2		25	6	29	65	10	66	15	45
97	6/7/2020 11:04	61	2	2		10	40	20	50	20	60	60	10
98	6/7/2020 11:17	35	2	1		70	1	60	1	3	70	60	80
103	6/7/2020 12:06	63	2	1		30	5	30	80	60	70	20	80
105	6/7/2020 12:14	26	2	1		54	18	65	15	30	45	35	40
107	6/7/2020 13:12	29	2	2		45	20	35	70	50	65	70	70
111	6/7/2020 13:57	25	1	-2	Matura	60	10	30	98	30	80	30	60
112	6/7/2020 14:49	26	1	-2	Higher technical Collage	60	30	30	20	60	70	40	60
113	6/7/2020 14:50	63	2	3		40	15	30	10	20	90	10	80
114	6/7/2020 14:54	26	1	1		50	5	18	90	15	66	15	40
118	6/7/2020 15:39	78	1	2		40	25	50	85	50	75	15	75
119	6/7/2020 15:51	50	2	2		5	30	20	60	20	60	20	40
120	6/7/2020 15:55	70	2	-2	general qualification for university entrance	80	56	12	80	15	80	40	75
123	6/7/2020 16:32	58	1	-2	Matura	80	35	30	80	15	75	5	90
125	6/7/2020 17:11	29	1	2		30	20	40	80	30	70	10	50
128	6/7/2020 18:47	31	2	1		30	7	30	2	10	70	30	40
140	6/7/2020 20:14	29	1	2		65	40	35	80	30	67	5	20
156	6/8/2020 15:04	62	1	1		25	20	30	90	30	66	5	80

CASE	LASTDATA	DS01_01	DS02	DS03	DS03s	AE01_01	AE02_01	AE03_01	AE04_01	AE05_01	AE06_01	AE07_01	AE08_01
25	6/6/2020 16:01	26	2	1		40	20	30	2	3	80	5	30
27	6/6/2020 16:06	22	2	-2	Matura	40	80	60	90	60	70	10	80
30	6/6/2020 16:18	26	2	1		30	5	40	20	10	90	5	20
34	6/6/2020 16:52	38	2	1		60	20	40	70	50	70	40	80
44	6/6/2020 18:45	64	2	3		45	40	30	99	50	70	10	60
46	6/6/2020 18:37	33	2	2		33	27	25	85	27	42	25	56
52	6/6/2020 19:08	29	1	-2	Bsc	40	30	40	95	60	70	15	40
54	6/6/2020 19:17	50	2	-2	Akademie	40	30	20	30	40	60	40	50
55	6/6/2020 19:18	26	2	-2	Bachelor	20	20	21	5	15	80	15	15
56	6/6/2020 19:35	39	2	1		50	10	15	80	65	66	7	70
59	6/6/2020 19:54	54	1	-2	University degree	50	10	30	80	30	70	20	70
63	6/6/2020 21:07	41	2	2		50	30	40	20	40	50	40	80
75	6/6/2020 22:03	28	1	2		20	15	25	95	13	75	13	20
77	6/7/2020 2:41	38	1	2		80	50	12	93	5	85	3	80
79	6/6/2020 22:14	28	2	1		15	10	45	15	10	45	5	15
85	6/6/2020 22:43	26	1	1		70	30	60	80	10	70	30	80
87	6/6/2020 23:39	70	1	4		25	10	35	75	35	70	7	30
91	6/7/2020 8:23	29	2	2		14	7	20	65	8	67	5	40
97	6/7/2020 11:09	61	2	2		15	30	16	50	20	60	50	60
98	6/7/2020 11:21	35	2	1		80	1	60	1	3	70	50	60
103	6/7/2020 12:12	63	2	1		20	10	30	80	40	70	30	80
105	6/7/2020 12:19	26	2	1		45	20	50	50	30	45	40	55
107	6/7/2020 13:15	29	2	2		65	50	50	80	50	65	70	70
111	6/7/2020 14:00	25	1	-2	Matura	80	5	35	98	40	82	25	30
112	6/7/2020 14:55	26	1	-2	Higher technical Collage	60	30	30	20	30	70	40	60
113	6/7/2020 14:57	63	2	3		60	20	40	10	2	90	15	20
114	6/7/2020 14:57	26	1	1		42	12	17	99	18	71	14	25
118	6/7/2020 15:51	78	1	2		50	45	65	85	60	80	85	65
119	6/7/2020 15:58	50	2	2		20	20	20	40	30	40	20	30
120	6/7/2020 16:07	70	2	-2	general qualification for university entrance	80	30	12	85	10	40	38	90
123	6/7/2020 16:59	58	1	-2	Matura	25	15	25	70	20	67	90	85
125	6/7/2020 17:15	29	1	2		60	20	40	80	20	70	10	20
128	6/7/2020 18:55	31	2	1		40	9	28	95	3	70	5	45
140	6/7/2020 20:17	29	1	2		65	40	35	80	30	67	10	30
156	6/8/2020 15:13	62	1	1		25	20	30	90	20	70	5	80