ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis MSc Behavioural Economics - Marketing track
Academic year 2019/2020

Reducing egocentric advice discounting: the effect of advice reasoning on revising judgments

Name student: Simeon Georgiev
Student ID: 545586sg

Supervisor: AC Peker
Second assessor: Prof. dr. A Baillon

Date final version: 10.08.2020

Table of Contents

Acknowledgements

I would like to show gratitude to my thesis tutor, Cem Peker, for his continuous support, patience and guidance. His assistance proved of great value during this study. I also appreciate my thesis co-reader, Aurelien Baillon, for his efforts over the academic year.

Deserving of thanks are my parents, Rosinka and Galentin, who have supported and encouraged me to work hard. They inspire me every day and I would not have been able to achieve what I have without them. Furthermore, I am grateful for my partner, Viktoria, and her boundless support and belief in my capabilities.

Last but not least, I appreciate the Erasmus School of Economics and the contribution of every individual who took part in my experiment, without whose assistance I would not have been able to conduct my research.

Executive Summary

When faced with a difficult task people turn to the opinion of others to enhance the accuracy of their decision. Seeking out advice and combining it with one's own intuition has been shown to improve judgements, which is why businesses are often reliant on collaboration between professionals. Despite the benefits of combining advice with one's own opinion, individuals tend to put more faith in their own initial judgment than the advice. This thesis focuses on differential information as the primary cause of this behaviour. This is the notion that individuals only have access to their own reasons for reaching their judgment and hence put more weight on their own opinion. What is novel for this study is the use of advice reasoning as a tool to solve this information asymmetry. An experiment was conducted across a convenience sample and its goal was to detect whether subjects who receive both an advice and advice reasoning overweight their own opinion to a lesser extent compared to subjects who were provided with an advice only. The results confirmed that individuals discount the incoming advice egocentrically. Nonetheless, the experiment was not able to find sufficient evidence for the effectiveness of advice reasoning as a tool to minimize the effect of differential information. To conclude, the study provides further support for the prevalence of advice discounting, which has practical implications as organizations and individuals increasingly rely on advice and use it suboptimally. It remains inconclusive whether the provision of advice reasoning is an appropriate technique for reducing egocentric advice discounting.

Chapter 1 - Introduction

Nowadays individuals face an increasing number of choices of various complexity in most aspects of everyday life and making a satisfactory decision can often be difficult. In light of this, social psychology maintains that when individuals face uncertainty, they often consult others with the intention to obtain their opinion to make better-informed choices (Festinger, 1954). Hofmann, Lei and Grant (2009) define this act of seeking advice as an inquiry of information and support from others. It is people's increasingly interconnected way of life that predisposes them to seek out the opinions of others. This practice is used universally in formal or informal settings to overcome a variety of technical, business, legal, medical, psychological and relationship issues. More precisely, the advice-taking paradigm categorizes two actors - advisors who give advice and judges who take the advice and make a decision (Hütter & Fiedler, 2019). This system explains that once advice is incorporated into the decision-making process, individuals evaluate the extra piece of information and update their initial judgment (Yaniv & Kleinberger, 2000).

From a practical standpoint, modern organizations increasingly rely on open offices and team collaboration to accomplish their objectives due to the benefits of collective knowledge and the fact that groups are wiser than single individuals (Mannes, 2009). Thus, modern businesses structure their staff in teams to benefit from more effective collective decision making (Sloman & Garratt, 2010). Such assemblies are capable of collectively discussing a given problem, identifying multiple potential solutions and reaching an improved decision. Indeed, Harvey and Fischer (1997) establish that combining judgments can positively affect decision accuracy and thus this technique is widely accepted in many professional fields. In practice, seeking and effectively consuming advice efforts appear to be integral for the success of business organizations, public institutions and other collectives. Hence, gaining a better understanding of this widely utilized social behavior can lead to identifying actionable improvements in business communication and effectiveness.

On the other hand, further exploring the subject of advice taking is also of academic relevance. Researchers have explored the topic from various angles and have focused on for instance how paid advice benefits from significantly reduced opinion discounting (Gino, 2008), how the more a group grows, the more influential its advice becomes (Mannes, 2009), or how averaging of opinions can lead to improved revised judgments (Soll & Larrick, 2009). Yet, a considerable gap in existing literature becomes evident as little attention is given to tackling one of the causes of advice discounting - differential information. Yaniv and Choshen-Hillel

(2012) explain it as a phenomenon occurring when an individual has full access to their own knowledge, but lacks that same amount of access to the internal thinking processes of another person and vice versa. In other words, a decision making judge lacks an understanding of the cognitive process and subjective experiences that formed the advisor's opinion.

In this situation of information asymmetry, a problem arises in the combination process of one's own judgment with the opinions provided by advisors. That is to say that individuals generally tend to discount advice in an egocentric manner, which means that the own judgments receive more consideration than the advisor's input so that in the end the two lines of thought are simply not objectively weighted (Yaniv & Kleinberger, 2000). Even though seeking advice has been shown to boost accuracy, it is this discrepancy of opinion weighting that leads to a substandard improvement in decision quality (Yaniv & Choshen-Hillel, 2012). As a result, businesses model their organizations in ways for staff to benefit from sharing information, nevertheless, individuals routinely underweight their colleague's input and thus make suboptimal revised judgments.

This study aims to contribute to the academic literature by exploring the little-observed issue of differential information by testing a method for decreasing its negative effect on revising judgment. Therefore, my objective is to answer the following research question:

What is the effect of providing advice reasoning on the advice discounting process?

An experiment will attempt to reveal whether reducing differential information between an advisor and a judge can reduce the decision maker's egocentric discounting. It is important to test the research question as advice discounting appears to be an involuntary and unintentional action (Lim & O'Connor, 1995). Additionally, Lim and O'Connor (1995) suggest that people entrust their own initial forecasts more than independent statistical advice even when they have been told it poses greater accuracy. Therefore, gaining a better understanding of advice discounting's root cause and testing a method to diminish its negative effect would shed a light on a little attended academic corner. Additionally, exploring this topic further may provide practical implications for policy makers, commercial organizations and consulting businesses providing professional services to clients.

This study relies on an experimental methodology to answer the aforementioned research question. An experiment was conducted via an online survey that was distributed across a convenience sample of 118 subjects. The survey consisted of ten quiz-like questions that required participants to provide an exact yearly estimate for historical events. The

experiment assesses how individuals react to receiving various forms of advice and how they revise their judgment as a result. The participants were randomly assigned in roughly even in size control and treatment groups, where the former exposes the subject to advice in the form of a yearly estimate by an anonymous fellow. Meanwhile, the treatment group has access to the same advisor's estimate along with a brief explanation of the reasoning behind their opinion, which acts as a treatment condition in the experiment. The research question is answered by testing whether there is a statistically significant difference between the average advice discounting in each group.

I find that advice is indeed not weighted equally against the subject's initial opinion thus supporting my first hypothesis about egocentric discounting. This finding confirms prior academic literature's notions about advice discounting, which state that people tend to overweight their own initial opinion and thus underweight the incoming advice. In this regard, my results provide additional support for existing theories on egocentric advice discounting and revision process that leads to suboptimal judgments. Furthermore, I did not find sufficient evidence to support my second hypothesis regarding differential information. My results suggest that the administered treatment in the form of advice reasoning does not lessen the amount of advice discounting and thus does not solve the issue posed by differential information.

My thesis continues with an overview of relevant literature on the topic of advice discounting and revising judgment. Then the two key hypotheses will be developed along with the methodology approach that is used to test them. The experiment's findings will be presented and analyzed. They are followed by a discussion on study limitations, avenues for future research and how my results fit in the contemporary academic consensus.

Chapter 2 - Literature review and hypothesis development

This chapter explores the literature on the topic of advice and outlines key concepts that are relevant for my research. Context is provided on advice taking, advice discounting with its effect and causes. Moreover, this section introduces the theoretical framework, the two hypotheses and their predictions.

2.1 Advice taking

Having given some insight into the relevance of the topic of advice taking in the previous chapter, it is valuable to provide a more detailed overview of when is advice requested. Yates, Price, Lee and Ramirez (1996) conclude that individuals are opposed to solving problems in a vacuum, hence they are open to incorporating sought after advice into the decision making process. Additionally, people interpret forecasts in a precise sequence, where initially a worthy opinion is obtained, then it is reviewed and finally combined with one's own outlook (Yates et al., 1996). This process serves the purpose of filling in gaps of knowledge, helping to determine the appropriateness of different alternatives or providing a new perspective on the problem at hand. It follows that advice would be sought after in situations that fit the aforementioned purposes well. When the topic of advice is discussed, it is essential to cover when, why and how it is best used.

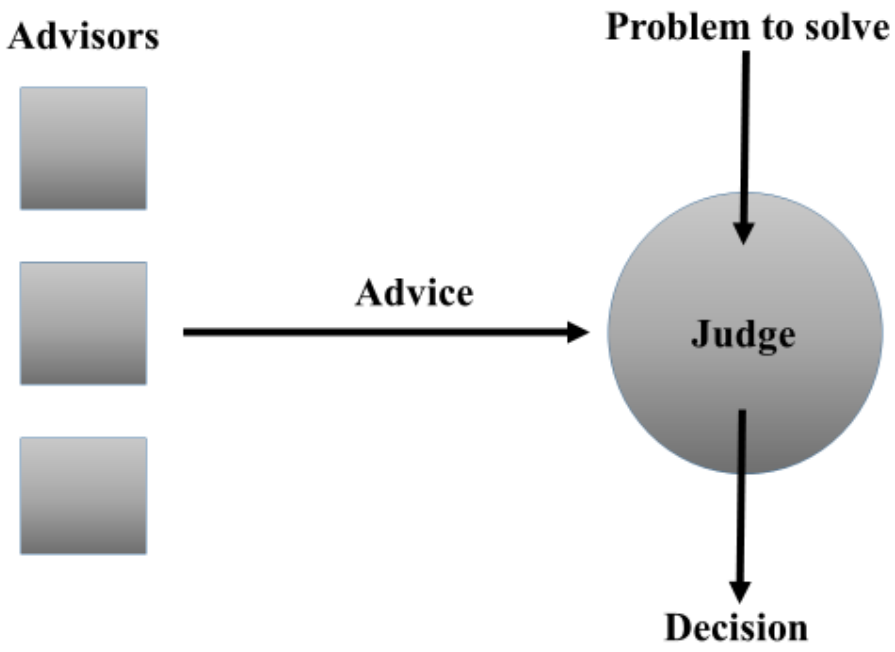
Specific situational conditions have been linked with the increased use of advice seeking. To illustrate, an additional opinion is often requested in instances that induce anxiety (Gino, Brooks & Schweitzer, 2012). It is not unreasonable to assume that a feeling of uneasiness is common when an uncertain problem is presented and a solution is required. This is directly interrelated with task difficulty as the likelihood of seeking and using advice increases as the complexity of the task grows (Schrah, Dalal & Sniezek, 2006). As a result, in the face of anxiety-inducing sophisticated problems individuals experience uncertainty in their capabilities and judgements (Festinger, 1954). Given the above, exposure to such conditions serves to explain the human desire to consult with the judgement of others.

As elaborated above, advice seeking is utilised in a wide variety of situations and it is worthwhile to explore why it benefits decision making. A vocal majority of researchers (Larrick & Soll, 2006; Harvey & Fischer, 1997; Yaniv, Choshen-Hillel & Milyavsky, 2009) argue that using advice is a great tool that enhances judgement precision. To be more precise, it is

suggested that considering advice directly moves one's judgment 20 percent closer to a correct numerical answer (Yaniv, 2004). In line with this, Harvey and Fisher (1997) confirm that advice universally increases accuracy even with various rates of learning in the task completion. Despite the benefit of improved precision, an additional opinion is also sought after for social reasons since by taking advice subjects can share the responsibility of a potential undesirable outcome when the risk of an error is perceived to be great (Harvey & Fisher, 1997). Moreover, asking for an extra opinion in a work setting can be advantageous as it has been shown to increase perceptions of competence in the advice seeker (Brooks, Gino & Schweitzer, 2015). This effect whilst notable is only observed in specific circumstances such as the problem being complex, the opinion being inquired in a personal manner and requested from an advisor with perceived expertise. Furthermore, professionals in fields such as audit seek out advice for impression management reasons (Kennedy, Kleinmuntz & Peecher, 1997). In an experiment, Kennedy et al. (1997) found that requesting advice resulted in a judgement that was perceived as justifiable even if the advice was not followed.

Having covered why advice is valuable to use and seek out, it is imperative to elaborate on how it is incorporated in the decision making process. Most research in the field of advice taking and judgement revising relies on the Judge Advisor System (JAS) which comprises of two acting bodies – a judge and an advisor (Sniezek & Buckley, 1995). The former has decision making power and can act upon the opinion provided by the latter in any way they want. It is necessary to assume that both parties share an interest in finding a solution to the problem and in the traditional JAS the advisor's judgement is provided for free and by default (Sniezek & Buckley, 1995). Additionally, an important distinction is to be made as the provision of advice does not automatically mean willingness to follow the additional piece of information (Brooks et al., 2015). According to JAS, individuals follow a sequence of actions that lead to a final decision. At the outset judges form their own opinion about a potential solution to the problem, then the advisor's opinion is assessed, it is combined with the judge's initial solution, and finally a decision is made (Sniezek & Buckley, 1995). A visual representation of JAS is provided in Figure 2.1 below.

Figure 2.1 Judge Advisor System (JAS)



The aforementioned combination process brings forth this thesis's main problem. To illustrate, the judge subjectively weights their own opinion alongside the one coming from the advisor, which is where a conflict between the estimates can occur (Yaniv, 2004). In this case, the judge has the option to either fully ignore, adjust toward, or fully adopt the incoming advice (Yaniv, 2000). An optimal manner to solve the aforementioned conflict is to weight the two opinions equally by performing simple averaging (Harvey & Fisher, 1997). However, in reality individuals rarely implement this strategy and often behave irrationally, which is further discussed in the next section.

2.2 Advice discounting

Researchers agree that when the decision maker is faced with a discrepancy between own opinion and given advice, he or she tends to discount the incoming recommendation and to place an excessive amount of trust in one's own judgment (Yaniv & Kleinberger, 2000; Lim & O'Connor, 1995; Mannes, 2009; Soll & Larrick, 2009). The experiments of notable researchers in the field take a closer look specifically at judgments defined as estimation of various types of numerical values e.g. historical dates (Yaniv et al., 2000), the caloric value of food items (Yaniv et al., 2009), temperature (Mannes, 2009), and annual salary (Soll & Larrick, 2009).

Accordingly, Krueger (2003) defines this egocentric tendency to discount opinions as the belief that one's own judgments are superior to those of others. The aforementioned researchers share a similar experimental design as they ask subjects about their best estimate on a quantifiable problem and then provide additional numerical estimates that act as advice (Yaniv & Kleinberger, 2000; Lim & O'Connor, 1995; Mannes, 2009; Soll & Larrick, 2009). In line with the JAS, the observed effect of advice discounting occurs at the final stage of the experiments where subjects are prompted to give their final answer, which is hopefully revised as their initial opinion is combined with the incoming advice. The rest of this thesis will specifically refer to a judgment and advice as point estimates on a continuous scale.

Advice discounting poses a problem for decision makers as the purpose of using advice is to increase one's own accuracy when completing a task. Yaniv and Kleinberger (2000) suggest that the uneven weighting of opinions also known as egocentric discounting occurs during the combination of advice and own opinion and leads to subpar judgements. Likewise, Mannes (2009) confirms this notion by stating that individuals generally consider the advice, however their efforts are counterproductive as there is a tendency to inadequately weight the available options against their initial views. In both of these studies, a direct comparison is made between the accuracy levels of advice discounting and equal weighting and the latter proves to be superior. Soll and Larrick (2009) define equal weighting in terms of computing a simple average for the given options. They also find that this technique is especially useful when dealing with estimates of quantity and probability. In fact, Clemen's (1989) research on forecasting suggests that the uncomplicated method of averaging is capable of eliminating errors stemming from overestimating or underestimating the correct answer. For example, if both the initial judgment and advice overestimate the exact answer, averaging them will improve performance. Connecting this back to JAS, if individuals behaved rationally, they would practice equal weighting of advice and own opinion, however, in actuality this behavior is seldom observed.

It is important to clarify what is the rational benchmark of egocentric discounting or how much of the advice do people take into consideration and how this standard is constituted. Advice taking literature tends to refer to the work of Harvey and Fischer (1997) who establish such a benchmark at an average of 20 percent adjustment from the initial judge's opinion. This means that the subject's initial judgment is shifted only about 20 percent of the difference between the initial estimate and the provided advice. This shows that on average individuals put more weight on their initial judgment than on the advisor's estimate and therefore end up making only minor adjustments to their original opinion to produce their final answer. In

support of this, Yaniv and Kleinberger (2000) reach a similar conclusion suggesting that individuals shift their opinions up to about 30 percent towards the advice.

As the outlined literature suggests, egocentrism is a common occurrence and it has been observed to have adverse effects on advice weighting (Yaniv & Kleinberger, 2000; Lim & O'Connor, 1995; Mannes, 2009; Soll & Larrick, 2009). As suggested by Yaniv (2004), individuals deal with a discrepancy between advice and own opinion by mostly remaining loyal to one's own judgement and making token adjustments in the direction of the given advice. Hence, this thesis aims to confirm this discounting effect within its own experiment with contemporary data. This research utilizes the widely used JAS where subjects are in the position of judges who are presented with a complex task and need to form an initial estimate, compare it to the provided advice and potentially revise their final judgement. For the purpose of this experiment, it is necessary to assume the advisor's accuracy is more or less the same as the judge's, meaning that the given advice is neither inferior nor superior. This assumption is necessary as the advice in the conducted experiment was generated by individuals with similar background to the actual subject pool. Having taken the above into account, the following hypothesis aims to reconfirm prior academic results:

H1: *When faced with a discrepancy between their own opinion and given advice, subjects tend to discount the advisor's opinion in the combination process.*

2.3 Advice discounting drivers

Understanding the roots of advice discounting is key for identifying an effective remedy that diminishes its effect on revising judgments. Initial theories attribute this discounting behaviour to the anchoring heuristic, which is defined as the notion that a person's initial opinion behaves similarly to an anchor which prohibits the adjustment of said opinion when advice is introduced (Tsversky & Kahneman, 1974). The effect of anchoring produces a final estimate that is more closely adjusted to the original judgment than to any provided advice. However, Soll and Larrick (2009) disagree with the aforementioned theory as they failed to reproduce the anchoring effect in their own experiment. In their case, they expected the amount of participant's judgement adjustment influenced by an anchor to vary, yet subjects opted for using other strategies such as choosing a precise option or averaging the alternatives (Soll & Larrick, 2009). Other opponents to the anchoring theory are Harvey and Harries (2004), who suggest

that being conservative in judgement formation is a more probable cause for advice discounting than being influenced by an initial information source.

The founding fathers of behavioral economics, Tsversky and Kahneman, also attribute egocentric advice discounting to one prevalent human characteristic - overconfidence in one's own abilities and knowledge (Tsversky & Kahneman, 1974). Additionally, Lim and O'Connor (1995) confirm this theory and identify the subject's overestimation in their own ability as a cause for advice discounting. In their experiment, a variety of useful visual statistical forecasts were ignored by individuals who fail to correctly assess their own perceived performance against their actual one. Moreover, Soll and Larrick (2009) reach a similar conclusion by demonstrating that overconfidence can also result in misidentifying inferior opinion as an expert one. Likewise, overconfidence does not only impact judgement formation but it is universal as shown by Svenson (1981), who states that individuals routinely overestimate their driving skills but underestimate their odds of experiencing a car accident.

Another notable mention is the mental distinction made by individuals between what is one's own and what is someone else's. As defined by Yaniv and Kleinberger (2000), the self/other effect causes advice discounting, since this effect is defined as the act of undervaluing the additional opinion just because it is coming from a third party. Egocentric advice discounting is driven by the belief that the own estimate is superior to the provided opinion (Harvey & Harries, 2004). In support of the self/other effect, Yaniv (2012) suggests that prior beliefs bias the weighting of options in favor of the own judgement in the mental combination process. Finally, Gardner and Berry (1995) expose the extent to which people distance their opinion from that of others with an experiment where helpful advice was only given upon request, which resulted in only 44 percent of subjects asking for assistance.

An additional explanation for advice discounting is the dissonance effect, which suggests that individuals have a tendency to show conservatism when faced with vastly opposing advice and are prone to look for information that confirms their prior beliefs (Klayman & Ha, 1987). This theory is supported by Cialdini (1993) who recognizes that people are in a constant struggle to maintain consistency in beliefs. Hence, advice closer to one's own judgement is more likely to produce an adjustment compared to advice whose numerical estimate is more distant. Decades prior Aronson, Turner and Carlsmith (1963) establish a similar notion that an opinion is less likely to change if there is an increase in distance of the new information with regards to one's own attitude. Therefore, an advice that highly contradicts a person's initial belief is less likely to be taken into consideration compared to an advice that does not.

2.4. Decreasing advice discounting via advice reasoning

Seen as egocentric advice discounting is a problem, it is important to explore how it's negative effect can be reduced. In the past 30 years, academic researchers have identified several conditions in which said phenomenon has been shown to diminish. Predictably, Harvey and Fischer (1997) find that when an advisor is perceived to have expert knowledge, individuals discount their advice to a lesser extent. Nevertheless, advisors with up to eight times more practice in task completion still have their opinion discounted by vastly less experienced judges, which only shows how prevalent and irrational advice discounting can be. Moreover, Yaniv and Kleinberger (2000) suggest that decision makers are sensitive to advice quality, meaning that opinion perceived to be of higher quality is taken into more consideration compared to a lower quality one. In their study, advice quality is interrelated to the reputation of the advisors as with the help of feedback screens across questions subjects can review how accurate the given opinions are. However, in a later study Yaniv (2004) disproves this notion by stating that a 20 percent increase in accuracy is not due to quality of advice but to its independent origin, meaning that the additional piece of information is helpful as long as it is coming from an independent advisor. The influence of advice quality is further opposed by Harvey and Fisher (1997) who claim that the sensitivity to advice is highly overstated, as shown in their experiment where varying caliber advice produces similar improvements in judgement. They advocate that advice perceived to be closer to the truth indeed produces an improvement in accuracy, however, even an incorrect advice has the same effect for poor initial judgements (Harvey & Fischer, 1997). An additional condition that reduces advice discounting is introduced by Soll and Larrick (2009) who suggest that egocentrism is moderated by task difficulty so that harder problems diminish an individual's self-confidence and advice discounting tendency. Additionally, opinion discounting can be directly impacted if the additional information is not provided freely but is being paid for (Gino, 2008).

Having covered some of the successful efforts of prominent researchers in the field to reduce the underweighting of helpful information, little attention has been given to one of the primary causes of egocentric advice discounting. Yaniv and Kleinberger (2000) and Yaniv (2012) suggest that egocentrism originates from differential information, which is the notion that the decision maker's own perceptions and experiences shape the manner in which they weight their initial judgment and the provided advice. Thus, individuals have privileged access

to their private reasons for holding a specific opinion but lack the same level of access to the internal reasons behind the advisor's judgment. Differential information is at the core of advice discounting as it is suggested that the self/other effect originates exactly from this information asymmetry between the two agents of the JAS (Yaniv, 2004). Additionally, it is reasonable to assume that having more evidence to back up one's own estimate is related to increased confidence, which explains why decision makers exhibit overconfidence in their own abilities to solve the given problem. Consequently, advice discounting is driven by differential access to the personal justifications for each opinion (Yaniv & Kleinberger, 2000). Precisely, evidence-based justification determines the weight distribution between opinions in the combination process between advice and initial judgement (Tversky & Koehler, 1994). This means that when the judge's access to evidence is not on par with that of the advisor, equal weighting of opinions is unlikely to occur.

Prominent researchers in the field of advice taking recognize the effect of differential information but few focus directly on reducing it (Gino, 2008; Yaniv, 2004; Yaniv & Choshen-Hillel, 2012; Yaniv & Kleinberger, 2000;). Hence, the following hypothesis aims to test a potential method of minimizing the aforementioned effect. More precisely, it would uncover whether differential information is decreased by a more transparent judgment formation process between the advisor and the judge. As elaborated above, the judge has personal evidence that supports their own opinion and so does the advisor. Having said this, it is questionable whether this information asymmetry can be reduced by having the advisor provide the reasoning behind their opinion. Advice reasoning in this thesis will refer to the cognitive process which leads advisors to their estimation. By incorporating advice reasoning in the JAS, this study aims to uncover whether this is a viable solution to the differential information phenomenon. Whilst previous research on the topic has developed options that tackle advice discounting, they do not directly address differential information that exists between a judge and advisor. Hence, the following hypothesis poses the question of whether disclosing the reasoning behind the advisor's judgment would lead to a reduction in egocentric discounting of subjects.

H2: *Provision of the reasoning behind an advisor's opinion decreases egocentric discounting in the judge.*

2.5 Conceptual framework

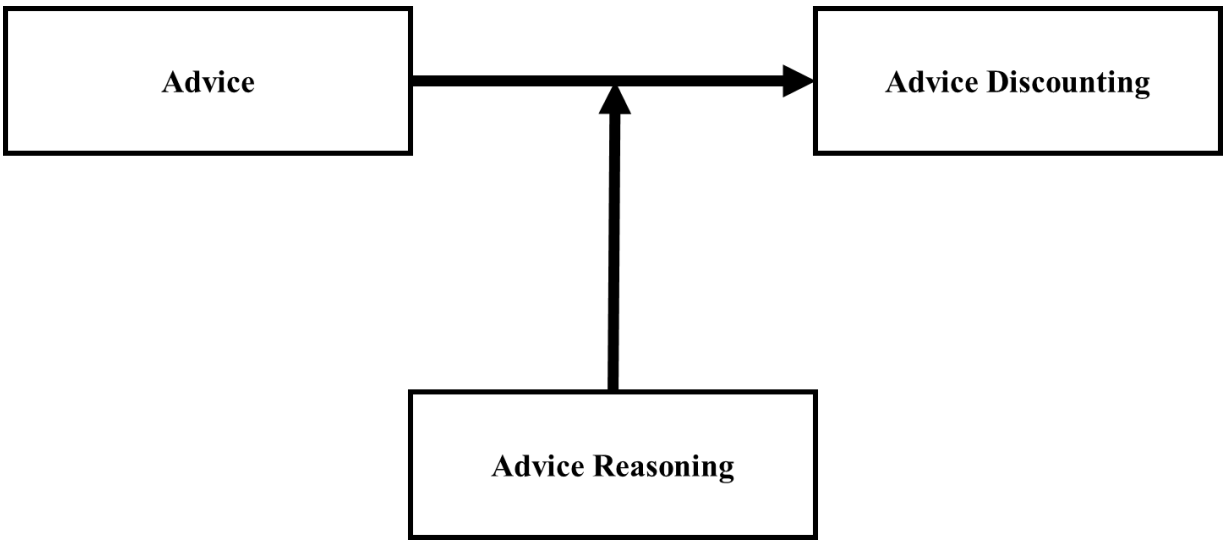
The conceptual framework comprises the constructs of advice discounting and advice reasoning. In this study advice discounting is the dependent variable of interest, while advice reasoning is the independent one. The latter also takes the form of the administered treatment that divides the study's subjects into two groups of roughly equal size.

The first hypothesis' goal is to confirm that individuals do not weight advice and their own opinion equally as seen in the discussed literature. My prognosis is that subjects will exhibit egocentrism in the evaluation of advice and own judgement. A subject is egocentric if he or she puts greater weight on own opinion rather than on provided advice. Hence, regardless of the experimental condition, the hypothesis expects individuals to overrely on their own beliefs when making a decision.

The second hypothesis aims to explore whether the provision of advice reasoning will have a decreasing effect on the egocentric advice discounting. In the experimental design, this effect will be achieved when advisors not only provide their best quantitative estimate but also include their own reasoning behind said estimate in short written form as explained in the next chapter. As outlined above, this treatment is expected to diminish the differential information between the judge and the advisor. Hence, I predict that subjects in this treatment exhibit lower advice discounting compared to the control condition, which does not receive advice reasoning.

Figure 2.2 presents the theoretical model with the variables and their relationships outlined above.

Figure 2.2 Conceptual framework



Chapter 3 - Research methodology

The following chapter focuses on this study's research design and data collection method. The section is organized around the choice of research technique, the statistical tests and measurements, the validity and reliability of the study. This academic work relies on self-collected data to test existing theoretical constructs in a novel approach with up-to-date information. An experiment in the form of an online survey was designed to assess the judgment revision process of conveniently sampled of group of subjects.

3.1. Experimental method

The two aforementioned hypotheses were tested with an experiment involving the creation of own data via the participation of subjects. The sample consisted of a randomly selected group of individuals whose demographic indicators such as age, gender, education level, nationality or location did not limit their participation in the study. The experiment had a between-subject design meaning that each individual participates in only one condition, either treatment or control. A between-subject design's main advantage is its ability to eliminate the opportunity for subjects to learn from the experiment and anticipate its goal, which can occur if an individual completes both conditions (Wooldridge, 2016). Furthermore, a one-shot observation design was adopted as it is easy to perform and it is in line with the limited research budget. Having the experiment consist of a single period gives subjects a single opportunity to fill in the survey and thus motivates them to perform the task optimally (Davis & Holt 1993). Similarly, convenience sampling was used to maximize the total amount of gathered responses. The electronic survey was circulated across the researcher's personal and professional networks.

A quantitative research method in the form of an electronic questionnaire was selected as it allows for the collection of large amounts of quantifiable data. Another advantage of gathering this type of information is the relatively straightforward analysis with the aid of statistical software. Nevertheless, when compared with qualitative techniques this research approach empowers subjects with complete anonymity, which can result in the provision of inadequate answers.

The experiment was distributed in the form of a structured online questionnaire, which was created via Qualtrics, a university-approved software platform. It was selected due to its user-friendliness, variety of question and answer options, and easily analysable output. In

addition, the software possesses a randomizer tool that randomly assigned relatively equal numbers of subjects to the two conditions. To correctly identify the causal effect of providing advice reasoning on subject's advice discounting, the experiment needed to include two groups of participants – those who undertook the treatment of advice reasoning and those who did not. The necessary assumptions which must hold for control to be obtained are that the groups should be similar in absence of the treatment and how they react to it (Gertler, Martinez, Premand, Rawlings & Vermeersch, 2016). The subjects should also not be influenced by an external factor and they must be randomly assigned in the different conditions. By designing a structured experiment with the only difference between groups being the treatment and by having individuals automatically assigned in conditions, the aforementioned assumptions automatically hold. This research method was also chosen since the question order in structured surveys remains constant. For these reasons, experiments in the field of judgement revision often utilise a survey format as seen in the research of Mannes (2009), Yaniv and Kleinberger (2000), Yaniv et al. (2009), and Yaniv and Choshen-Hillel (2012).

The experimental design of my study was heavily influenced by an experiment by Yaniv et al. (2009), where subjects were tasked with estimating a numerical measure of the caloric value of various food items. My experiment deviates from the aforementioned study design by using ten non-hypothetical historical questions. To clarify, subjects were tasked with making an exact estimation of the year in which a notable event from the last 300 years has occurred, which is an experimental question set-up also used by Yaniv (2004). The significant historical events were specifically selected so that subjects have a vague notion of them yet are difficult to date precisely without specialized knowledge. This type of questions is preferred when measuring advice discounting since they have the precise right or wrong answers. The full list of enquiries is exhibited in Appendix A.

The survey started with an introductory screen, which thanks subjects for their participation and informs them of the researcher's compliance with EU GDPR regulations ("General Data Protection Regulation (GDPR) Compliance Guidelines", n.d.) as seen in Appendix B. The next screen contained overall instructions about the format in which each yearly estimate should be written in. It also informed that the advisors differ across questions and most importantly requests subjects to not aid their responses with any internet search engines or other information sources as seen in Appendix C. The advisors were anonymous and they varied across questions as Yaniv and Kleinberger (2004) argue that subjects form opinions on the reputation of distinguishable advisors from the quality of their advice. In other words, the reputation of the advisor directly influences the participant's weighting policy of the advice.

Each of the ten historical questions consisted of two stages, which need to be completed for a subject to move on to the next question. In the first phase of the experiment, subjects were asked to provide their initial precise numerical estimation to a non-obvious historical question. Additionally, my experiment adopted a feature from Yaniv and Choshen-Hillel (2012) experiment, where participants were also asked to indicate how confident they were in their initial estimate. This rating illustrates the extent to which a subject assesses their initial estimation's accuracy. Individuals were presented with a scale and a movable pointer, which started at zero and could be moved up to a hundred. A confidence rating of 0% would mean "not confident at all" in the initial estimate and a rating of 100% would mean "completely confident". This scale identified subjects who believe to possess expert historical knowledge. Thus, in rare cases where the precise historical date was already known by a subject, that answer was excluded from the overall analysis.

This was done because it is rational for such individuals to discount advice when they assume to know the correct answer with certainty and hence expect to be better informed than any anonymous advisor. As explained in Chapter 2, people are predisposed to seek and consume advice in situations that produce uncertainty or anxiety, and when the problem is complex (Brooks et al., 2015). Hence, if a subject knows the problem's solution, then they are not put in such a situation and it is reasonable for them to discount incoming advice, which will skew the results. Similarly, as done in the studies of Harvey and Fischer (1997) nonsensical values, initial estimates given by subjects that are identical to advice or answers that overshoot or undershoot the correct date and advice multiple times over were all excluded from the analysis. This procedure resulted in less than four percent of the data to be omitted.

In the second phase of each question, the historical enquiry was shown again along with the subject's initial yearly estimate from the first phase. What is more, an anonymous advice to the same inquiry was displayed for reference on the same screen. At the bottom of the screen, subjects were tasked with entering their final possibly revised answer. The advice came in the form of precise numerical estimation. The benefits of numerical advice are that it is widely used to solve problems in real-life settings. Take for example, the tendency of professionals in financial analysis, legal, pharmaceutical and other consulting industries to communicate their opinions in numerical estimates (Yaniv, 2004). Also, numerical values allow for direct and exact measurement of the subject's advice weighting as it can be clearly calculated by how much the final opinion moves away or toward the advice (Yaniv, 2004). Subjects were expected to make a comparison between their initial estimate and the advice before the revised judgment occurs. This type of experimental design aimed to uncover to what extent respondents revised

their initial opinion after consulting with the advisor's best estimation on the same problem. This technique of obtaining revised judgment has been adopted in the research of Mannes (2009), Yaniv and Kleinberger (2000), Yaniv et al. (2009), and Yaniv et al. (2012). The second phase of each question ended with a screen that revealed the correct answer, which in this case was the actual year of the historical event. The addition of this experimental feature was motivated by experiments of Yaniv and Kleinberger (2000) and Harvey and Fisher (1997), who included a feedback screen after the provision of the subject's final estimate. The reasons for doing so were that firstly subjects can personally assess how accurate both their own estimate and that of the advisor were. Secondly, according to Yaniv et al. (2009) giving away this additional piece of information acts as a motivator for participants to think carefully about the question and to make the best estimate possible. The observation of the advice accuracy was expected to not have an effect on subject's decision making as they were informed that each piece of advice is coming from a different, independent and anonymous person. This design was implemented so that reputation formation about the advisors, as seen in the experiments of Yaniv and Kleinberger (2000), cannot occur. The feedback screen for each question can be seen in Appendix D.

To further incentivize the optimal completion of the survey, I incorporated a monetary incentive to the experimental design. It is widely believed in the research community that incentives are necessary for the production of reliable data (Cubitt, Starmer & Sugden, 1998). The labour framework of economic experiments was taken into account and it states that when subjects are rewarded for their mental effort, they perform the task better (Bardsley, Cubitt, Loomes, Moffatt, Starmer & Sugden, 2010). Hence, a monetary incentive of €10 was distributed randomly to a single participant for the completion of the experiment. Thus, the reward did not depend on the subject's task performance. There was an equal positive probability of a reward for every individual in the sample, which acted as additional encouragement. Nonetheless, the experiment still relied on intrinsically motivated subjects. Upon starting the experiment, all subjects were informed of the randomly assigned monetary incentive as seen in Appendix C. This aims to comply with the experiment labour framework and motivate participants to put more effort into their decision-making process. At the end of the survey participants who wish to enter the lottery were provided with the option to leave an email as seen in Appendix E.

Lastly, in order to account for additional characteristics of the subject sample, three demographic questions regarding gender, age and the highest attained education level were included after the ten estimation questions as seen in Appendix F.

The randomly assigned treatment was the provision of advice reasoning and it acted as the key independent variable in the statistical analysis. Advice reasoning took the form of a brief written explanation of how the anonymous advisor had reached their precise point estimate to the given problem. While in the first phase of each question all participants were asked to provide their best initial estimate to a question, the treatment is implemented in the second phase along with the given advice. The treatment subsample received advice with the reasoning behind the estimate as seen in Appendix G, while the control subsample only received advice with no reasoning behind it as seen in Appendix H. In the second phase subjects were expected to consult with the provided advice and to potentially revise their initial judgment. The advisor estimates and advice reasoning for each survey question were obtained prior to the actual experiment in a pre-experimental survey, which was distributed to a smaller convenience sample of ten intrinsically motivated volunteers. They were instructed to produce an estimated answer complemented by a detailed explanation of their thought process in a free entry text box as seen in Appendix I. The aim of this data gathering task was to record various opinions in sufficient detail which were later used as advice reasoning in the second phase of the questions. The use of so-called ecologically valid advice, collected from actual individuals and not artificially created by the researchers, has the benefits of being more authentic, allowing for more generalization from study to reality, since the researcher cannot control the advice's distance from the truth (Yaniv, 2004).

My survey was designed to be in line with the necessary conditions for obtaining control while creating own experimental data. By financially incentivizing subjects to invest mental effort in completing the questionnaire, I partially satisfy the nonsatiation and dominance precepts, which require researchers to sufficiently reward subjects for their effort with something people would like to have more of than less of it – most commonly cash (Smith, 1982). Secondly, subjects were not deceived in any way, which satisfied the salience precept (Smith, 1982). Additionally, the privacy precept is satisfied as participants were not made aware of other individual's pay-offs and could not be influenced by them. The winner of the reward was contacted privately after the survey was closed. By making considerable effort to comply with Smith's precepts (1982), I have obtained experimental control and thus keep other elements constant.

3.2. Construct measurements

For testing purposes, the variable of interest in this study was a measure of advice discounting. To uncover how subjects weigh their own judgment against advice, the "weight of own estimator" method (WOE) was used as introduced by Yaniv and Kleinberger (2000). This measure represented the final estimation as a combination of the initial estimate and the provided advice so that the weights were proportional to the direction in which the final answer moves - either toward or away from the advice. Hence, the measure of WOE represented how much weight a subject placed on their own opinion and how much advice was discounted. The WOE was calculated by the following formula:

$$WOE = \frac{|advice - final\ estimate|}{|advice - initial\ estimate|}$$

The WOE was measured as a proportion between one and zero, where one indicated that a subject does not alter their final answer from their initial opinion and advice was completely discounted. Alternatively, a WOE value of zero showed that a subject has completely adopted the advice, meaning that advice was not discounted at all. Therefore, any WOE values between one and zero showed a partial discounting of advice and thus both initial opinion and advice were weighted to some extent. For example, a WOE value of 0.5 suggests equal weighting of advice and own opinion, but a value of 0.8 demonstrates egocentric discounting of advice and preference toward own opinion. When the assumption of neither superior nor inferior advice is taken into account, a WOE value closer to one would represent egocentric discounting. It is important to point out that WOE was computed for each question for each participant and for testing purposes the mean WOE across questions per participant is taken into account. Averaging at the subject level was decided as an appropriate technique since key studies on the topic followed this method (Yaniv, 2004; Yaniv et al., 2009; Yaniv & Choshen-Hillel, 2012). Additionally, I adopted Yaniv and Kleinberger's (2000) experimental WOE technique as it has been widely accepted in the research community.

The questionnaire also collected information about a number of control variables such as gender, age and education. Gender is a dummy variable with a value of zero for females and one for males. In the data set one respondent preferred not to disclose his or her sex which results in one missing value. The "age" variable is continuous and the "education" variable is categorical as it takes the value of one for obtained secondary education only, two stands for

bachelor's degree, three represents master's degree and finally four indicates that the subject has obtained a doctorate degree.

With my first hypothesis, I aim to test whether subjects tend to discount advice and overly on their own opinion. To test the first hypothesis, I performed one sample t-test to uncover whether the whole sample mean of WOE significantly differed from a hypothesized value of 0.5. This cut-off point was used by Yaniv and Kleinberger (2000) and it indicates equally weigh of own judgment and the provided opinion. Hence, the subject's final estimate is a product of averaging the distance between the initial estimate and the given advice. Hypothesis 1, claiming that individuals do not weight advice and their own opinion equally, will be rejected in case the test results suggest that advice and initial opinion are given equal weight in the combination process. If I fail to reject Hypothesis 1, then the result will be consistent with prior literature in the field of advice taking and judgment revising, and will support the notion of egocentric advice discounting.

To test the second hypothesis, I utilized a two independent sample t-test for equal means to assess whether the provision of advice reasoning behind the given advice had an effect on advice discounting (WOE). I used this test as my aim is to understand if the mean WOEs in the treatment and control groups are equal or differ significantly. As stated before, the control group received only advice while the treatment group had both advice and the reasoning behind it at their disposal. The two-sample t-test would answer the question whether receiving advice and its reasoning reduces egocentric discounting. The null hypothesis states that the mean WOE difference between the two groups is zero. Hence, if I reject the null hypothesis then advice reasoning is contributing to a difference in the mean WOE, which means that individuals discount the advice less egocentrically when provided with advice reasoning. This would mean that providing additional information about the thought process behind the advice can minimize the effect of differential information and decrease advice discounting. Meanwhile, the very opposite would be true in case I fail to reject the null hypothesis.

Lastly, for statistical completeness I ran a linear regression that included all of the aforementioned control variables in its model. This test showed whether age, gender or education had a statistically significant influence on the amount of advice discounting. Furthermore, the regression allowed for the interpretation of the effect of one unit increase in the independent variable on the dependent variable, mean weight of own estimator.

3.3. Validity and reliability of methodology

Before covering the results from the hypothesis testing it is meaningful to discuss the overall reliability and validity of the study.

The experiment's internal validity is defined as the ability to recognize causal relationships between the theoretical constructs (Angrist & Pischke, 2008). That is to say that the results should not be influenced by any confounds, which are variables that directly influence the dependent variable of interest. An effort was made to eliminate one such confounding effect - selection bias. This bias was neutralized by having subjects automatically assigned between the two conditions via the randomization tool in the Qualtrics software. The conducted treatment randomization ensures that the individuals in the two groups are similar, which makes the comparison between groups possible. That is to say, the only difference between the groups is the treatment and the outcome in absence of treatment will be identical (Duflo, Glennerster & Kremer, 2007). Additionally, confounding effects arising from the context that subjects bring to the experiment are counteracted by presenting all individuals with the same order of questions and the same instructions. Moreover, by adhering to Smith's (1982) nonsatiation, dominance, salience and privacy precepts experimental control is obtained, which ensures proper research design and complements internal validity.

When it comes to external validity, the main concern is whether the conclusions of this study are generalizable for the remainder of the population. In order for any identified causal relationships to hold outside of the experiment, the sample included not only student subjects but also other individuals with varying age, education and gender. Despite this, the representativeness of the sample remains in question due to some situational factors. The survey was distributed mainly across South Holland and location factors may be at play. On another note, some sample features like the survey distribution across a network of higher education multinational students may lead to limited generalizability. Additionally, the study's small sample size may have an impact on the statistical power of the conducted tests and thus reduce the study's external validity.

It is imperative for the reliability of academic studies to withstand being replicated under similar circumstances by other researchers who come to comparable conclusions. Firstly, to ensure reliability this paper adopted a widely used experimental mechanism for gathering revised judgment by obtaining an initial estimate, presenting a stimulus and collecting final revised estimates as used by several researchers in the field such as Mannes (2009), Yaniv and

Kleinberger (2000), Yaniv et al. (2009), and Yaniv et al. (2012). Secondly, the ten survey questions are a matter of fact since they are of historical nature and thus are not debatable. Thirdly, the use of quantitative tools is associated with greater accuracy of the findings compared to qualitative research methods. Despite this, the convenience sampling of subjects and the very impersonal nature of electronically administered surveys leave room for the influence of various environmental factors. One cannot account for the setting in which the survey was completed, the level of subject's task comprehension or how much thoughtful consideration was invested in each answer.

3.4. Descriptive statistics of the data

The data set consisted of 118 unique independent observations on the subject level. The average survey participant was 28 years old and 53 percent of the sample was male. The remainder of the descriptive statistics can be observed in Table 3.1 below.

Table 3.1 Descriptive statistics

	Treatment condition	Control condition	Total
	n = 58	n = 60	118
Gender			
Female	27	27	54
Male	30	33	63
Prefer not to disclose	1	0	1
Age			
Min	21	20	20
Max	59	58	59
Mean	27.67	28.05	27.86
Standard Deviation	9.27	9.00	9.10
Education			
Secondary	0	3	3
Bachelor	24	33	57
Master	32	22	54
Doctorate	2	2	4

Additionally, Table 3.2 presents a correlation matrix that shows the correlation between the variables included in the statistical model. It can be observed that there is no

multicollinearity between the variables since no correlation coefficient is above 0.80. Unsurprisingly, it can be concluded that there is a positive relationship between age and education, which is reasonable to assume as with the passing of time individuals are likely to progress in their education. Interestingly, mean WOE as an indicator of advice discounting does not have a significant relationship with the treatment, advice reasoning, which suggests that the provision of advice reasoning did not significantly alter advice discounting. This contradicts what was expected as the treatment was anticipated to have an effect on the dependent variable. This result potentially casts doubt on the effectiveness of advice discounting. Remarkably, there is a significant positive relationship between the correlation coefficients of age and mean WOE, which may hint at increased levels of advice discounting with the increase of age. This could be explained by the tendency of people to acquire more knowledge with age and this may lead to a stronger belief in own judgment.

Table 3.2 Correlations between variables

	Mean WOE	Advice reasoning	Gender	Age	Education
Mean WOE	1.0000				
Advice Reasoning	0.0273 0.7690	1.0000			
Gender	0.0432 0.6440	-0.0237 0.7994	1.0000		
Age	0.2212*** 0.0161	-0.0208 0.8228	-0.1700* 0.0669	1.0000	
Education	0.0984 0.2896	0.1954** 0.0339	-0.1063 0.2540	0.4037*** 0.0000	1.0000

***, **, * indicate level of statistical significance at the 0.01, 0.05, and 0.10.

Chapter 4 - Results

This section covers the outcomes of the performed statistical test and directly addresses the aforementioned hypotheses. Initially, I report the results of the one-sample t-test, the two-sample t-test and the multiple linear regression, which includes control variables to the model. Then, I discuss the randomization checks and the necessary assumptions for the conducted t-tests and linear regressions. Lastly, a brief summary of the results is provided and followed by a discussion on how they fit in the established literature on the subject.

4.1 Hypotheses testing

To test the first hypothesis which predicts that people discount the advisor's opinion egocentrically, I conducted a one-sample t-test, whose output is presented in Table 4.1 below. The test's null hypothesis equated the mean weight of own estimator across the whole subject sample to 0.5, which implies an even distribution of weight between the advice and own estimate. I hypothesized that individuals have a tendency to discount incoming advice and overweigh own opinion, which makes their mean weight of own estimator take a value closer to one. The result of the t-test showed that the individual's mean weight of own estimator (M=0.647, 95% CI, 0.612 to 0.681) was higher than the equal weighting score of 0.5, t-statistic (117) = 8.399, p-value = 0.000 (significant at 1% significance level). This outcome supports my first hypothesis that subjects did not weigh advice and own intuition equally but tended to overrely on own initial prediction.

Table 4.1 Testing Hypothesis 1

One sample t-test						
Dependent Variable	Observations	Mean	Standard Errors	Standard Deviation	95% Conf.	Interval
Weight of own estimator	118	0.647	0.017	0.189	0.612	0.681
mean = mean (mean_woe)						t = 8.39
H0: mean = 0.5						df = 117
Ha: mean < 0.5		Ha: mean != 0.5			Ha: mean > 0.5	
p-value: 1.000		p-value: 0.000***			p-value: 0.000***	

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

My second hypothesis focused on examining the role of advice reasoning on the mean weight of own estimator of subjects. I hypothesized that the provision of advice reasoning with the numerical advice can reduce the effect of differential information and thus minimize the overreliance on own opinion. To test this hypothesis, I conducted a two sample t-test to determine if there was a difference in the means of the dependent variables between the two conditions. Each group consisted of about 60 randomly assigned participants. The results in Table 4.2 show that individuals that received advice reasoning had no statistically significant lower mean weight of own estimator (M=0.641) compared to the subjects that received advice only (M=0.652), p-value = 0.796. This outcome failed to find support for my second hypothesis. That is to say, the provision of advice reasoning for the decision-making process did not have a statistically significant effect on the judgement revision.

Table 4.2 Testing Hypothesis 2

Two sample t-test for equal means						
Dependent variable: Weight of own estimator						
	Observations	Mean	Mean difference	df	t	p-value
Control	60	0.641	-0.010	116	-0.2943	0.769
Treatment	58	0.652				

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

The use of parametric t-tests is appropriate if their integral assumptions hold. Despite having reviewed said assumptions in the next section, nevertheless I decided to test this hypothesis with the t-tests' non-parametric counterpart – Mann Whitney U test for comparing the medians between the population. This non-parametric test has fewer necessary assumptions compared to the used above parametric tests and these assumptions are met as they are linked to the study design (Laerd Statistics, "Mann Whitney U test assumptions", n.d.). It hypothesizes that the mean weight of own estimator is equal between the control and treatment conditions. The results in Table 4.3 indicate that the mean WOE of the two groups are equal as the null hypothesis cannot be rejected, p-value = 0.895. Hence, this non-parametric test's finding is consistent with the parametric one and further demonstrates that there is no evidence that the applied treatment produced its predicted effect.

Table 4.3 Non-parametric testing of Hypothesis 2

Mann Whitney U test for comparing means				
Dependent variable: Weight of own estimator				
	Observations	Rank sum	Expected	p-value
Control	60	3545.5	3570	0.895
Treatment	58	3475.5	3451	

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

4.2 Randomization checks

After testing the aforementioned hypotheses, it is useful to assess the extent to which randomization is successfully implemented across the control and treatment groups. Despite having used the automatic respondent randomization feature in the Qualtrics software, one can never be certain that perfect randomization is achieved. I had little reason to suspect that the distribution of subjects in either control or treatment condition may not be fully random. However, I conducted an analysis of variance to determine whether the mean of the demographic control factors (age, gender and education) was the same across the two independent groups of control and treatment. Hence, I ran one-way ANOVA tests to examine whether there were statistically significant differences in the group's means. If such differences were detected, then I can conclude that randomization was not successful. This test compared the means of the control variables of age, attained education level and gender of the subjects across the two groups.

The outcomes of ANOVA test indicated that variables of age and gender were randomized effectively (p-value > 0.05) while the education variable was not (p-value < 0.05). This result meant that there was a statistically significant difference in the means of attained education between the two conditions, p = 0.0339. By looking at the means in Table 4.6 of Appendix J, it can be seen that the individuals in the control group (M = 2.3833) have obtained less education compared to the treatment group (M = 2.6207). Therefore, I have reasons to suspect that education was not approximately equally distributed so that the treatment subjects were on average higher educated than those in the control condition.

4.3 Tests for t-test assumptions

As the two key hypotheses are tested with t-tests, it was vital to assess the necessary assumptions for reaching an unbiased coefficient result with correct standard errors. The initial three t-test assumptions were directly addressed by the experimental design. They required the use of a continuous dependent variable, a categorical independent variable and the collection of only independent observations (Laerd Statistics, "Independent t-test in Stata", n.d.). This study complied with said assumptions by having the dependent variable of weight of own estimator take a continuous value between zero and one. Additionally, the independent variable of advice reasoning took the form of a dummy for the presence or the lack of a treatment condition. The independence of observations was achieved by designing an experiment that was completed individually with limited opportunity for subjects to be swayed by the opinions of other participants. What is more, subjects were evenly assigned on a random principle to either of the two experimental conditions.

The remaining t-test assumptions required additional statistical examination as their validity cannot be guaranteed by the experimental design alone (Laerd Statistics, "Independent t-test in Stata", n.d.). Firstly, the data should not contain significant outliers. I predicted that there might be outliers in variables of age and mean WOE and thus created boxplots to visualize the data to check for outliers, which can be seen in Appendix K. These visualizations indicated that there are no outliers for mean WOE, however the boxplot for age displayed approximately twelve subjects above the usual pattern. As seen in the descriptive statistics section above, the subject's mean age was 27.8 years old and the aforementioned boxplot illustrates roughly a dozen participants whose ages lied in the range between 30 and 60. This can be attributed to the fact that the data was gathered via convenience sampling across a network with a majority of student population. Despite the assumption not holding entirely and potentially undermining the accuracy of the results, outliers are common for small sample studies. A viable solution to this would be the use of a larger data set which is less likely to register outliers.

The next assumption required a homogeneity of variances which was formally tested with Lavene's test (Laerd Statistics, "Independent t-test in Stata", n.d.). As seen in Table 4.7 of Appendix L, the results demonstrated that the variance of the independent variable of advice reasoning in the control and treatment groups was homogenous. Thus, the difference in means of the two conditions did not statistically differ from another, which means that the homogeneity assumption holds.

Finally, there needs to be a normal distribution of the residuals of the dependent variable of weight of own estimator (Laerd Statistics, "Independent t-test in Stata", n.d.). Since the variable is a range between zero and one, its residuals have values within the same range. This was tested with a linear regression with the independent variable which was followed by predicting the residuals of the variable of interest. Afterward, a Shapiro-Wilk test of normality identified whether the residuals are normally distributed. As seen in Table 4.8 of Appendix M, the residuals of weight of own estimator for the treatment variable are normally distributed ($p=0.068$) as the p-value is not under 0.05, which is regarded as the standard significance level. The null hypothesis cannot be rejected, hence we do not have enough evidence to state that the residuals are not normally distributed. As a result, the assumption holds when a significance level of 0.05 is used. Nevertheless, it is worthwhile to mention that the dependent variable is only approximately normally distributed since WOE and its residuals both are values between zero and one, which technically cannot be normally distributed.

4.4 Additional testing

One of the drawbacks of conducting t-tests is that they do not account for other variables, but linear regressions do. Performing a multiple regression allows for incorporating more than one independent variables, which on its own allows to control for more factors that simultaneously can affect the dependent variable (Platt, 1997). Additionally, regressions make for richer and easy to interpret statistical models. In the case of this study, I collected information not only regarding the subject's judgments but also on an individual's characteristics such as age, gender and education. These factors were accounted for in the following regression. A linear regression allowed to test for an effect of one unit change in the independent variable on the variable of interest, WOE.

Before examining the result of the linear regression, it is vital to consider the necessary assumptions for reaching an unbiased coefficient. Firstly, it is natural for linearity in parameters to not hold as multitude of factors are included in the model (Wooldridge, 2016). Secondly, the assumptions for lack of outliers, no perfect collinearity and normal distribution of residuals coincide with the t-test assumptions and hence have been discussed above and those tests can be seen in Appendices J, K, L and M (Laerd Statistics, "Multiple regression in Stata", n.d.). Thirdly, the zero conditional mean assumption cannot be formally tested, however a number of steps have been taken to make sure it holds (Wooldridge, 2016). For instance, a more complete

model was developed by adding more variables to the regression. Additionally, the key explanatory variable or the treatment condition was randomized evenly across the sample.

Finally, a formal test was needed to assess the constant error variance assumption or homoscedasticity, which means that the error term has the same variance given any value of the independent variables (Wooldridge, 2016). After creating variables in the raw data set for fitted values, unstandardized and standardized, and studentized residuals, a formal test for constant error variance showed no significant evidence for heteroscedasticity (p-value = 0.304). Hence the homoscedasticity assumption holds as seen in Table 4.9 in Appendix N.

Having covered the necessary assumptions, I ran a linear regression with mean weight of own estimator as a dependent variable, which is a continuous interval variable. The independent variable was the provision of advice or lack thereof, which was a dummy variable. I also added the factors of age as a continuous variable, gender as dummy variable and education as a categorical variable to the model.

The formula for the multiple linear regression is presented below along with its results in Table 4.10:

$$WOE = \beta_0 + \beta_1 \times Advice\ Reasoning + \beta_2 \times Gender + \beta_3 \times Age + \beta_4 \times Education$$

Table 4.10 Multiple linear regression results

Dependent variable: Mean Weight of Own Estimator				
	Coefficient	Standard error	t	p-value
Advice Reasoning	0.019	0.035	0.53	0.596
Gender	0.031	0.035	0.89	0.373
Age	0.005	0.002	2.30	0.023**
Education	-0.002	0.032	-0.08	0.940
Intercept	0.495	0.083	5.93	0.000***
Observations = 117, R^2 = 0.0558, Adjusted R^2 = 0.0221				
F statistic = 1.65 (df = 4, 112), p = 0.1658				

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

The multiple regression was run to include possible effects of the control variables. The whole model did not statistically significantly predict mean WOE, R^2 = 0.056 (p = 0.166). From the value of R-squared, it can be estimated that only 5.58% of the variance is explained by the regression model. Thus, the model had low explanatory power and did not fit the data well.

From the aforementioned control variables, only age was statistically significant (p-value = 0.023) at 5% significance level, ceteris paribus. This means that one additional year in age increases the mean WOE by 0.005 points. This finding suggested that the older an individual is, the higher their mean WOE score would be, which translates to an increased reliance on own initial judgment instead of the given advice. Despite this, by looking at the magnitude of the coefficient of the variable age ($\beta_3 = 0.005$), one could argue that the coefficient has debatable scientific relevance due to its small size. More importantly, the regression outcome confirmed that the treatment variable of advice reasoning has no statistically significant effect on the mean value of the weight of own estimator.

4.5 Result summary

The aim of this research was to inspect the phenomenon of egocentric advice discounting and examine the effect of provision of advice reasoning on the judgment revision process. This study's results have provided additional evidence to support the pre-existing notions of advice discounting. On the one hand, the conducted experiment confirmed the prediction that individuals do not weigh advice and own initial opinion evenly but rather overweigh their initial judgment. This finding is consistent with the conclusions of Yaniv (2004), Lim and O'Connor (1995), Mannes (2009), Soll and Larrick (2009) and others. Finding support for my first hypothesis contributes to existing literature and highlights the prevalence of advice discounting in revising judgments, where judgements are point estimates on a continuous quantity. This finding does not diverge from those of previous experiments in the field as this study utilized a widely prevalent experimental design for eliciting revised judgments.

On the other hand, the conducted experiment was unsuccessful in finding evidence that the provision of advice reasoning reduces the amount of the judge's advice discounting. Hence, there was no support for the prediction that the reasoning decreases the effect of differential information, which was defined as a primary contributor to egocentrism by Yaniv and Klenberger (2000). No significant effect of the randomly applied treatment was uncovered by either one of the two alternative tests performed – parametric and non-parametric. In conclusion, the conducted experiment was unable to answer the imposed research question regarding the ability of advice reasoning to influence judgment revision. The potential reasons for the lack of such effect will be discussed further in the next chapter.

This study offers meaningful conclusions that are consistent with established literature's findings regarding the prevalence of overweighing one's own judgment. Additionally, this paper acts as further evidence for the effectiveness of the utilized multiple-stage experimental design's ability to elicit egocentric advice discounting in subjects. However, the applied treatment was unable to reduce advice discounting and it cannot be concluded that the provision of advice reasoning is an effective method for reducing differential information.

Chapter 5 - Discussion

The subject matter of advice discounting remains of academic interest due to the prevalence of advice seeking in social and professional settings. As emphasized above, requesting an additional opinion regarding a problem offers both a meaningful increase in accuracy and social benefits. However, individuals fall victim to their own cognition and perform in a suboptimal manner in applying the given opinion to their decision making process. The primary reason is advice discounting, which is driven by multiple factors and originates from differential information. This study contributes to the existing literature as the chosen research method confirms the notion shared by Yaniv and Kleinberger, (2000), Lim and O'Connor (1995), Mannes (2009) and Soll and Larrick (2009) that people have egocentric tendencies when they assess additional pieces of information. My experiment failed to find support for the hypothesis that the provision of advice reasoning decreases differential information which is recognised by Yaniv (2012) as the ultimate driver of advice discounting. The applied treatment failed to reduce egocentric discounting. The following sections discuss the direct implications of my findings, the study's shortcomings and avenues for future research on the topic.

5.1 Implications

Taking into account the findings of the conducted experiment two takeaways are clear. Indeed, subjects discount incoming advice egocentrically. However, there was no sufficient evidence to suggest that the addition of reasoning to the advisor's opinion had a decreasing effect on the judge's advice discounting.

First of all, this study provides additional evidence for the prevalence of advice discounting among individuals who face a discrepancy between initial opinion and given advice. Thus, the key implication is the confirmation that this area of human cognition and behaviour remain ripe for further research. A better comprehension of these complex mental processes may lead to improvements in judgement formation and advice provision. More clarity on such advice discounting behaviour and its negative effect on performance can directly benefit how advice is provided in business, medical, education and other professional settings. As a result, individuals in the position of advisors should be more aware of the tendency of decision makers to overvalue their own opinion.

Secondly, my research did not find sufficient evidence to confirm the provision of advice reasoning as a viable tool for overcoming differential information. This bears implications for business organisations, consulting bodies and policy makers. Hence, the provision of more information and transparency along with the advice may not necessarily resolve the opinion discounting problem.

By not finding support for the use of advice reasoning, my study highlights the lack of understanding of the topic of differential information. Moreover, this outcome may serve future researchers in identifying more novel ways to minimize information asymmetry in the JAS.

5.2 Limitations

It is imperative to also address this study's limitations when considering its outcome and discussion of its results. Due to the limited research possibilities, a relatively small subject sample participated in the electronically distributed experiment. Consequently, it is likely that my findings have less statistical power and it is difficult to generalize a specific behaviour across the whole population. This puts into question the inability to find evidence for significance of the thesis' second hypothesis as one can argue that a larger data set of observations could have produced a different result. Additionally, I also recognize that it is in the nature of electronically distributed surveys to limit the researcher's control over environmental factors. For instance, subjects have the freedom to partake in the experiment whenever it is convenient for them. Therefore, the researcher cannot account for the time of day, the setting or any stimuli in the environment, which can be controlled for in a conventional lab experiment.

While previous academic work is inconclusive on the topic of advice quality, little is being said about the quality of the advice reasoning. This is because this technique albeit straightforward has received little academic attention. My study relies on the underlying assumption that the provided advice is neither inferior or superior to the opinion of the judge. However, the same cannot be said about the argumentation that advisors provide in the treatment condition. It remains open to discussion whether the differential information can be minimized just by providing advice reasoning or by providing value-adding reasoning behind the given opinion.

It is also notable to mention that the experimental design depends on numerical judgements. While they are statistically reliable and easy to interpret, they are not entirely

realistic when professional and social settings are taken into account. In reality, the majority of problems do not only consist of the estimation of quantities or probabilities. In fact, usually situations that require advice do not come in the form of textbook questions. Additionally, the historical nature of the questions used in the survey-style experiment may have been familiar ground for subjects. As elaborated above, highly confident and knowledgeable observations, which made up roughly four percent of the data set, were omitted from the analysis. However, one may argue that the participants may have been familiar with the historical events in question and have not experienced a state of anxiety or uncertainty necessary for active advice seeking to occur.

5.3 Recommendations for future research

It is unreasonable to think that my study completely covers all the nuances of the topic of advice discounting. Thus I provide a number of suggestions for future research in the topic. First of all, a larger and more diverse subject sample may further validate the conclusions that were reached in my study. Additionally, this bears the potential to reveal evidence to overturn the conclusion about the significance of the applied treatment. To collect my pool of observations I heavily relied on convenient sampling across student colleagues and personal networks, which is inherently prone to lead to doubts about the generalizability of the findings to the whole population and there is a possibility of overrepresentation in the data set.

Additionally, further attention on the quality of advice reasoning may yield meaningful results. Perhaps future experiments can incorporate advice argumentation of varying worth, as this characteristic did not receive considerable scrutiny in my own experimental design. It is not unreasonable to suggest that the quality of advice reasoning may be a topic of debate very much like the discussion revolving around the quality of advice. This may very well be a potential explanation for the lack of significant effect of advice reasoning on advice discounting in my own experiment.

Another avenue for improvement is the context of the experimental design itself. There are distinct benefits to researching judgments and advice as numerical estimates on a continuous scale. However, advice of qualitative nature has received little investigation. This is likely because such advice would not fit the weighting assessment methods that are widely practiced in the field. For instance, this study's WOE method would not be computable if the formula inputs are not quantitative. Additionally, rather than focusing on matter of fact questions, which

are the experimental standard, an interesting approach would be evaluating questions of taste or attitude. Such type of judgments is especially prevalent in the fields of art, law and politics, where subjective beliefs trump matters of fact.

Finally, this study's statistical analysis relied upon t-tests and regression models, which could have accounted for more subject characteristics. It is probable that there may be other factors that directly influence advice taking and judgment revision which were not included in my experimental design. Their addition could have explained some the variance and enriched the statistical models. Perhaps character traits such as compliance or stubbornness may impact the extent to which subjects follow advice. In addition, some aspects of a person's professional environment and experience may affect their propensity to accept or reject incoming opinion. It may be the case that an individual who works within a team and engages in daily discussions with others regarding the delivery of a product or a service is more open to weighing advice equally compared to a person who works on their own.

5.4 Conclusion

The use of advice benefits decision making and it is imperative to comprehend how people combine advice with their own intuition. My study found further evidence that individuals discount external opinions and overweight their own judgments. My experiment's results remain inconclusive with regard to whether the provision of advice reasoning along the advice has an impact on advice discounting. Nevertheless, I presented potential explanations for this outcome and avenues for further research. In summary, the topic of advice discounting and revising judgements is a multifaceted one and it is increasingly evident that a lot remains unknown.

References

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Aronson, E., Turner, J. A., & Carlsmith, J. M. (1963). Communicator credibility and communication discrepancy as determinants of opinion change. *The Journal of Abnormal and Social Psychology*, *67*(1), 31.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). Experimental Economics: Rethinking the Rules. Princeton; Oxford: Princeton University Press.

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, *61*(6), 1421-1435.

Cialdini, R. B. (1993). Influence: The psychology of persuasion (Rev. ed.). *New York: Morrow*.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, *5*(4), 559-583.

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, *1*(2), 115-131.

Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton university press.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, *4*, 3895-3962.

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, *7*(2), 117-140.

Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied cognitive psychology*, *9*(7), S55-S79.

General Data Protection Regulation (GDPR) Compliance Guidelines. Retrieved 19 July 2020, from https://gdpr.eu/

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. The World Bank.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational behavior and human decision processes*, *107*(2), 234-245.

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of personality and social psychology*, *102*(3), 497.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational behavior and human decision processes*, *70*(2), 117-133.

Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting*, *20*(3), 391-409.

Hofmann, D. A., Lei, Z., & Grant, A. M. (2009). Seeking help in the shadow of doubt: the sensemaking processes underlying how nurses decide whom to ask for advice. *Journal of Applied Psychology*, *94*(5), 1261.

Hütter, M., & Fiedler, K. (2019). Advice taking under uncertainty: The impact of genuine advice versus arbitrary anchors on judgment. *Journal of Experimental Social Psychology*, *85*, 103829.

Kennedy, J., Kleinmuntz, D. N., & Peecher, M. E. (1997). Determinants of the justifiability of performance in ill-structured audit tasks. *Journal of Accounting Research*, *35*, 105-123.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, *94*(2), 211.

Krueger, J. I. (2003). Return of the ego--Self-referent information as a filter for social prediction: Comment on Karniol (2003).

Laerd Statistics. (n.d.). *Independent t-test in Stata*. Retrieved July 16, 2020 from https://statistics.laerd.com/stata-tutorials/independent-t-test-using-stata.php

Laerd Statistics. (n.d.). *Mann Whitney U test assumptions*. Retrieved July 16, 2020 from https://statistics.laerd.com/statistical-guides/mann-whitney-u-test-assumptions.php

Laerd Statistics. (n.d.). *Multiple regression in Stata*. Retrieved July 16, 2020 from https://statistics.laerd.com/stata-tutorials/multiple-regression-using-stata.php

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*(3), 149-168.

Mannes, A. E. (2009). Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Science*, *55*(8), 1267-1279.

Platt, R. W. (1997). Logistic regression and odds ratios. *Injury prevention*, *3*(4), 294.

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an Island: integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, *19*(1), 43-60.

Sloman, J., & Garratt, D. (2010). *Essentials of economics*. Pearson Education.

Smith, V. (1982). Microeconomic systems as an experimental science. The American Economic Review, 72 (5).

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in Judge-Advisor decision making. *Organizational behavior and human decision processes*. *62*(2), 159–174.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers?. *Acta psychologica*, *47*(2), 143-148.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological review*, *101*(4), 547.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational behavior and human decision processes*, *93*(1), 1-13.

Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making*, *25*(5), 427-434.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 558.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, *83*(2), 260-281.

Yates, J. F., Price, P. C., Lee, J. W., & Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting*, *12*(1), 41-56.

Appendices:

Appendix A: Qualtrics Survey questions

Question 1:

In what year did Charles Darwin's release “On the Origin of Species”?

Your best estimate:

Question 2:

In what year was insulin first used to treat diabetes?

Your best estimate:

Question 3:

In what year did Gandhi's civil disobedience movement result in the Indian independence act?

Your best estimate:

Question 4:

In what year did Albert Einstein announce his theory of relativity?

Your best estimate:

Question 5:

In what year did New Zealand become the first country in the world to grant women the vote?

Your best estimate:

Question 6:

In what year was smallpox officially declared as eradicated worldwide?

Your best estimate:

Question 7:

In what year was NATO founded?

Your best estimate:

Question 8:

In what year was the construction of the Empire State Building completed?

Your best estimate:

Question 9:

In what year did Alexander Graham Bell invent the telephone?

Your best estimate:

Question 10:

In what year did the first living being reach space (low earth orbit)?

Your best estimate:

Appendix B: Qualtrics survey's introductory screen

Greetings,

My name is Simeon Georgiev and I want to thank you for taking part in this survey! You are helping my master thesis research at Erasmus University Rotterdam.

This survey should take you less than 10 minutes to complete.

Your responses are completely anonymous. Data collection is for research purposes only and it is in line with the EU GDPR privacy regulations. If you have any questions about the survey, please feel free to contact me at 545586sg@eur.nl.

→

Appendix C: Qualtrics survey's instruction screen

You are given ten questions and you need to provide your best estimate for certain events by typing a single year in the text entry box e.g. 1994.

For each question, you will receive advice that is provided by a different advisor.

Please do NOT use any search engines (Google, Bing, Yandex, etc.) **or other information sources to look up answers**. This is NOT an examination and your responses are completely anonymous.

One participant will be chosen at random and rewarded €10 for completing this survey!

→

Appendix D: Feedback screens with correct historic date

Question 1:

The actual year is 1859.

Question 2:

The actual year is 1922.

Question 3:

The actual year is 1947.

Question 4:

The actual year is 1915.

Question 5:

The actual year is 1893.

Question 6:

The actual year is 1979.

Question 7:

The actual year is 1949.

Question 8:

The actual year is 1931.

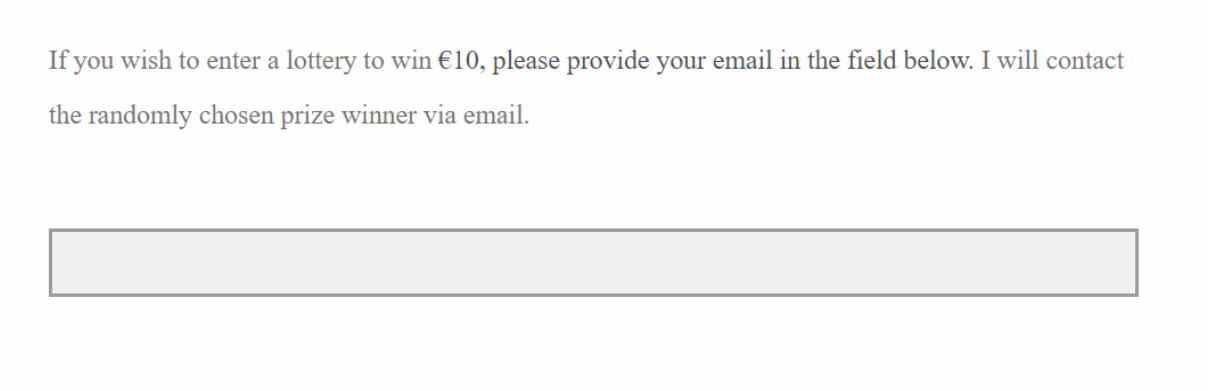
Question 9:

The actual year is 1876.

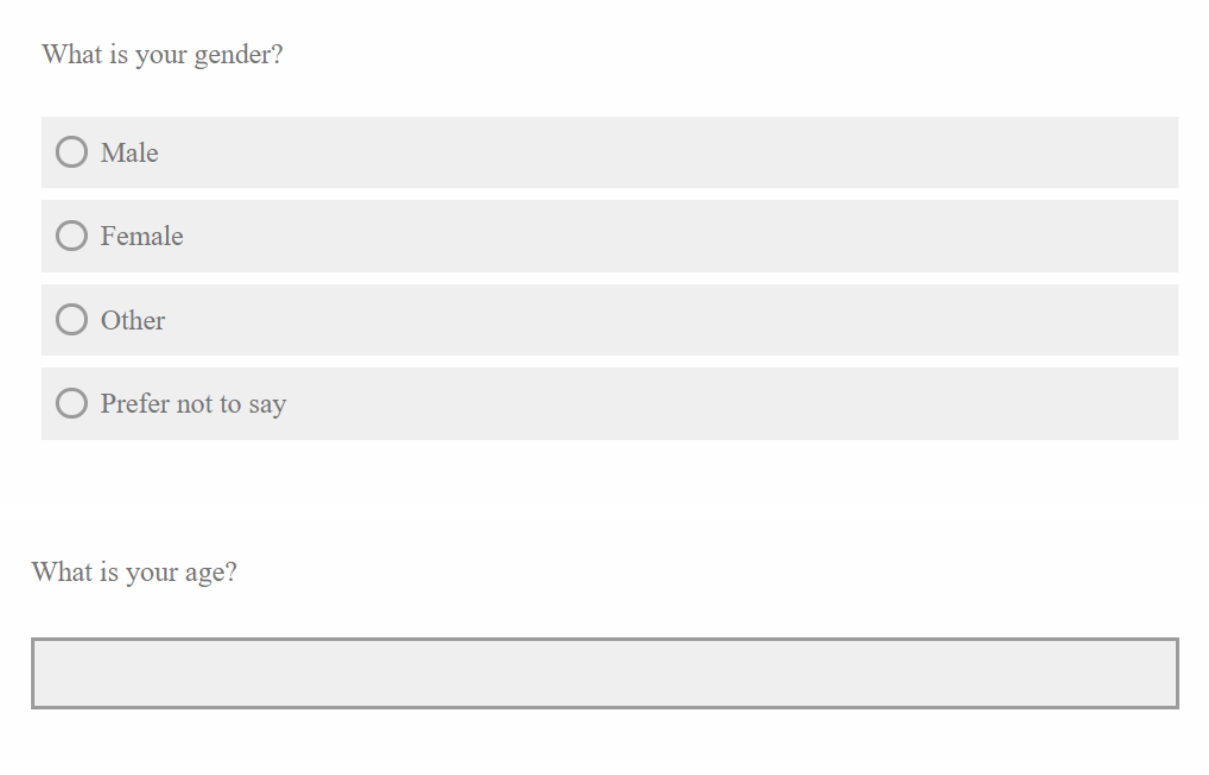
Question 10:

The actual year is 1957.

Appendix E: Random lottery survey question



Appendix F: Demographic survey questions



What is your highest attained education level?

- ○ Secondary education
- ○ Bachelor's degree
- ○ Master's degree
- ○ Doctorate degree / PhD

Appendix G: Advice in the treatment condition

Question 1:

In what year did Charles Darwin's release "On the Origin of Species"?

Your previous estimate was ${q://QID37/ChoiceTextEntryValue}

The best estimate of advisor **A** was 1885, because advisor **A** has seen only black and white photos of Darwin and he lived before the Great World Wars.

Your final best estimate:

Question 2:

In what year was insulin first used to treat diabetes?

Your previous estimate was ${q://QID40/ChoiceTextEntryValue}

The best estimate of advisor **B** was 1955 because advisor **B** thinks that most medicines were developed in the early 20th century and diabetes is a more recently studied disease.

Your final best estimate:

Question 3:

In what year did Gandhi's civil disobedience movement result in the Indian independence act?

Your previous estimate was ${q://QID43/ChoiceTextEntryValue}

The best estimate of advisor **C** was 1925 because advisor **C** thinks that India fought in World War I on the side of England but the county did not participate in World War II.

Your final best estimate:

Question 4:

In what year did Einstein announce his theory of relativity?

Your previous estimate was ${q://QID46/ChoiceTextEntryValue}

The best estimate of advisor **D** was 1921 because advisor **D** thinks Einstein was already a famous scientist before his escape from the Nazi regime.

Your final best estimate:

Question 5:

In what year did New Zealand become the first country in the world to grant women the vote?

Your previous estimate was ${q://QID49/ChoiceTextEntryValue}

The best estimate of advisor **E** was 1870 because advisor **E** thinks that New Zealand was among the first to gain independence from the British empire and develop its own parliament.

Your final best estimate:

Question 6:

In what year was smallpox officially declared as eradicated worldwide?

Your previous estimate was ${q://QID52/ChoiceTextEntryValue}

The best estimate of advisor **F** was 1995 because advisor **F** thinks that mass vaccination became possible due to the globalization of the late 20th century.

Your final best estimate:

Question 7:

In what year was NATO founded?

Your previous estimate was ${q://QID55/ChoiceTextEntryValue}

The best estimate of advisor **G** was 1953 because advisor **G** thinks that many world organizations and initiatives emerged after the World Wars.

Your final best estimate:

Question 8:

In what year was the construction of the Empire State Building completed?

Your previous estimate was ${q://QID58/ChoiceTextEntryValue}

The best estimate of advisor **H** was 1912 because advisor **H** thinks that before the Great Depression there was plenty of cheap immigrant labor from Europe.

Your final best estimate:

Question 9:

In what year did Alexander Graham Bell invent the telephone?

Your previous estimate was ${q://QID61/ChoiceTextEntryValue}

The best estimate of advisor **I** was 1832 because advisor **I** thinks that the industrial revolution of the 19th century introduced a lot of new inventions to humanity.

Your final best estimate:

Question 10:

In what year did the first living being reach space (low earth orbit)?

Your previous estimate was ${q://QID64/ChoiceTextEntryValue}

The best estimate of advisor **J** was 1950 because advisor **J** thinks that animal test flights occurred prior to human flights.

Your final best estimate:

Appendix H: Advice in the control condition

Question 1:

In what year did Charles Darwin's release “On the Origin of Species”?

Your previous estimate was ${q://QID3/ChoiceTextEntryValue}

The best estimate of advisor **A** was 1885.

Your final best estimate:

Question 2:

In what year was insulin first used to treat diabetes?

Your previous estimate was ${q://QID9/ChoiceTextEntryValue}

The best estimate of advisor **B** was 1955.

Your final best estimate:

Question 3:

In what year did Gandhi's civil disobedience movement result in the Indian independence act?

Your previous estimate was ${q://QID12/ChoiceTextEntryValue}

The best estimate of advisor **C** was 1925.

Your final best estimate:

Question 4:

In what year did Einstein announce his theory of relativity?

Your previous estimate was ${q://QID15/ChoiceTextEntryValue}

The best estimate of advisor **D** was 1921.

Your final best estimate:

Question 5:

In what year did New Zealand become the first country in the world to grant women the vote?

Your previous estimate was ${q://QID18/ChoiceTextEntryValue}

The best estimate of advisor **E** was 1870.

Your final best estimate:

Question 6:

In what year was smallpox officially declared as eradicated worldwide?

Your previous estimate was ${q://QID21/ChoiceTextEntryValue}

The best estimate of advisor **F** was 1995.

Your final best estimate:

Question 7:

In what year was NATO founded?

Your previous estimate was ${q://QID24/ChoiceTextEntryValue}

The best estimate of advisor **G** was 1953.

Your final best estimate:

Question 8:

In what year was the construction of the Empire State Building completed?

Your previous estimate was ${q://QID27/ChoiceTextEntryValue}

The best estimate of advisor **H** was 1912.

Your final best estimate:

Question 9:

In what year did Alexander Graham Bell invent the telephone?

Your previous estimate was ${q://QID30/ChoiceTextEntryValue}

The best estimate of advisor **I** was 1832.

Your final best estimate:

Question 10:

In what year did the first living being reach space (low earth orbit)?

Your previous estimate was ${q://QID33/ChoiceTextEntryValue}

The best estimate of advisor **J** was 1950.

Your final best estimate:

Appendix I: Results from advice reasoning gathering

Question	Anonymous advisor id	Advice	Advice reasoning
1. Darwin	001	1815	Darwin lived in the early 19th century.
	002	1850	Darwin's ship had Victorian era sails.
	003	1885	I have only seen black and white photos of Darwin and he lived before the Great World Wars.
2. Diabetes	004	1955	Most medicines were developed in the early 20th century and diabetes is a more recently studied disease.
	005	1936	Medical discoveries occurred in the early 20th century.
	006	1926	Such medicinal advancements did not occur before World War I.
3. Gandhi	007	1948	Gandhi was active in the early 30s political movement and Indian's independence happened around the 40s of the 20th century.
	008	1925	India fought in World War I on the side of England but the country did not participate in World War II.
	009	1935	His daughter Indira Gandhi was in power in the 1970's.
4. Einstein	010	1921	Einstein was already a famous scientist before his escape from the Nazi regime.
	001	1918	His work was used in the development of the atomic bombs.
	002	1925	Einstein was relatively young when it got published.
5. Rights	003	1910	The civil rights movement occurred in that era.
	004	1870	The county was among the first to gain independence from the British empire and develop its own parliament.
	005	1918	Europe gave women rights some time later in the 20th century.
6. Smallpox	006	1995	Mass vaccination became possible due to the globalization of the 20th century.
	007	1965	Vaccination rates can spike in more stable times without wars.
	008	1977	It must have taken some time for it to be eradicated across African countries.
7. NATO	009	1949	It was formed in mid-20th century.
	010	1947	Countries needed to combine collective security against communist regimes.
	001	1953	Many world organization and initiatives emerged after the World Wars.
8. Empire	002	1926	Skyscrapers are a product of the early 20th century.
	003	1912	Before the Great Depression there was plenty of cheap immigrant labor from Europe.
	004	1938	It must have been built before World War II.
9.Telephone	005	1832	The Industrial Revolution of the 19th century introduced a lot of new inventions to humanity.

	006	1872	The principles of signals and electromagnetic waves were understood in the late 19th century.
	007	1889	Telephones were already common in the early 20th century.
10. Space	008	1965	It happened before the moon landing.
	009	1956	Gagarin's flight occurred in the early 1960's.
	010	1950	Animal test flights occurred prior to human flights.

Appendix J: Randomization checks

Table 4.4 Randomization check - Age

Mean age of the two conditions

	Mean	Standard Deviation	Frequency
Control	28.050	9.004	60
Treatment	27.672	9.267	58
Total	27.864	9.097	118

ANOVA

	Sum of squares	df	Mean square	F	p-value
Between groups	4.205	1	4.205	0.05	0.8228
Within groups	9677.626	116	83.428		
Total	9681.831	117	82.751		

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table 4.5 Randomization check - Gender

Mean gender of the two conditions

	Mean	Standard Deviation	Frequency
Control	0.550	0.502	60
Treatment	0.526	0.504	57
Total	0.538	0.501	117

ANOVA

	Sum of squares	df	Mean square	F	p-value
Between groups	0.016	1	0.016	0.06	0.7994
Within groups	29.061	115	0.253		
Total	29.077	116	0.251		

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table 4.6 Randomization check - Education level obtained

Mean education of the two conditions

	Mean	Standard Deviation	Frequency
Control	2.383	0.640	60
Treatment	2.621	0.556	58
Total	2.5	0.610	118

ANOVA

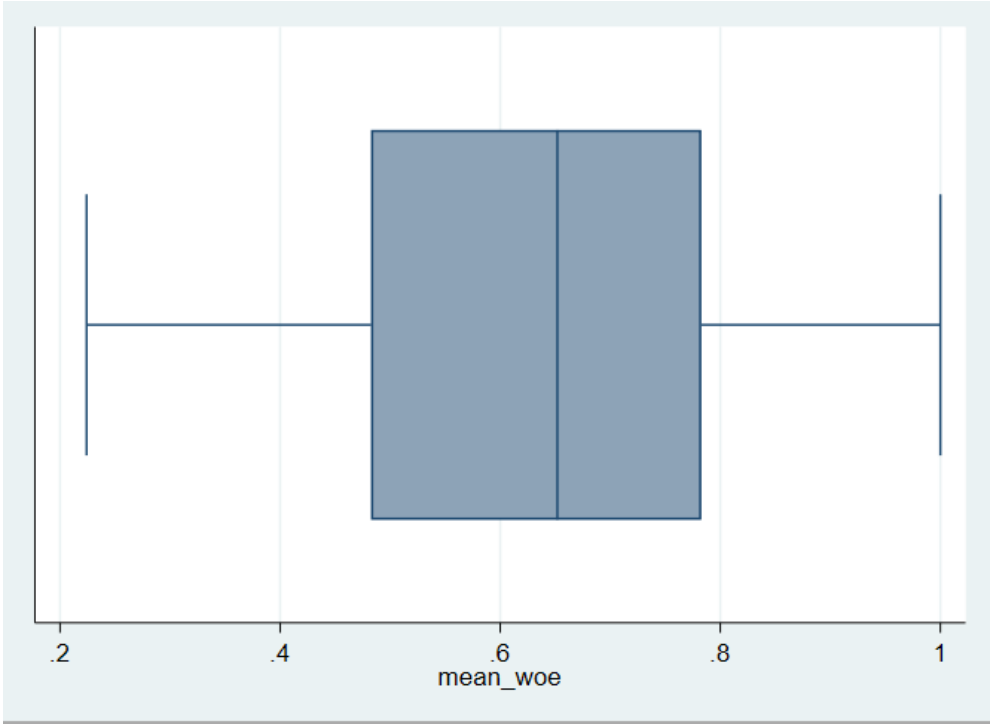
	Sum of squares	df	Mean square	F	p-value
Between groups	1.661	1	1.661	4.61	0.0339**
Within groups	41.838	116	0.361		
Total	43.500	117	0.372		

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Appendix K: Testing for outliers

Boxplot 1. Outliers in the continuous variable of age

Boxplot 2. Outliers in the dependent variable of weight of own estimator



Appendix L: Testing for homogeneity

Table 4.7 Tests for equality of variance of the independent variable

	Levene's test for equality of variance		
	W0	df	p-value
Advice reasoning	0.396	1, 116	0.5303

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Appendix M: Testing for normal distribution

Table 4.8 Tests for normal distribution of residuals of the independent variables

	Shapiro-Wilk test of normality				
Residuals	Observations	W	V	z	p-value
Regression for advice reasoning	118	0.979	1.945	1.489	0.068*

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Appendix N: Testing for Heteroskedasticity

Table 4.9 Testing for Heteroskedasticity

Source	Chi 2	df	p-value
Heteroskedasticity	13.95	12	0.304
Skewness	4.44	4	0.350
Kurtosis	8.55	1	0.004
Total	26.94	17	0.059

***, **, * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.