

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Wisdom of Crowds in Decision Making: Enhancing
Kickstarter Valuation through Optimized Decision-Making
Processes“

verfasst von / submitted by

Mirza Gazdic, BSc (WU)

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Betriebswirtschaft

Betreut von / Supervisor:

Assoz. Prof. Steffen Keck, PhD

Abstract

The "Wisdom of Crowds" concept, well-known in the field of judgment and decision-making, states that aggregated estimates frequently outperform individual estimates and tend to emerge as the best estimate. This experiment at hand investigates the "wisdom of many in one" by obtaining a second estimate from the same person and averaging it, with the aim of improving the results. It investigates the cognitive foundations of why averaged quantitative estimates provide improved results. The key component of this analysis is an empirical study done using an online survey with three unique groups of randomly selected individuals where participants answered judgment-based knowledge questions while evaluating Kickstarter Projects. The study uses statistical tools to compare responses across groups and assess the influence of decision-making procedures on judgment accuracy and dependability. The experiment used three different decision-making techniques: Forced, Temporal Separation, and Dialectical Bootstrapping, to evaluate their application in crowdsourcing values. The study shows that judgment averaging has the potential to improve accuracy and dependability, which is especially obvious in the Forced and Temporal Separation groups. It also highlights the advantages of establishing gaps between assessments, implying the importance of temporal separation in decision-making situations. This experiment provides useful insights on utilizing the "wisdom of crowds" in crowdfunding, emphasizing the importance of bigger sample numbers and improved judgment processing to assure result dependability. The findings pave the door for further study into precisely predicting individual project valuations, which might assist investors by identifying ideas that are likely to get significant investment.

Kurzfassung

Das Konzept der "Weisheit der Vielen", bekannt im Bereich des Urteilens und Entscheidens, besagt, dass aggregierte Schätzungen häufig individuelle Schätzungen übertreffen und sich tendenziell als die beste Schätzung herausstellen. Dieses Experiment untersucht die "Weisheit vieler in einem", indem es eine zweite Schätzung von derselben Person erhält und diese mittelt, mit dem Ziel, die Ergebnisse zu verbessern. Es erforscht die kognitiven Grundlagen, warum gemittelte quantitative Schätzungen verbesserte Ergebnisse liefern. Der Schlüsselbestandteil dieser Analyse ist eine empirische Studie, die mittels einer Online-Umfrage mit drei einzigartigen Gruppen von zufällig ausgewählten Personen durchgeführt wurde, wobei die Teilnehmer wissensbasierte Urteilsfragen beantworteten, während sie Kickstarter-Projekte bewerteten. Die Studie verwendet statistische Werkzeuge, um Antworten über Gruppen hinweg zu vergleichen und den Einfluss von Entscheidungsverfahren auf Urteilsgenauigkeit und -zuverlässigkeit zu bewerten. Das Experiment verwendete drei verschiedene Entscheidungstechniken: Erzwungen, Zeitliche Trennung und Dialektisches Bootstrapping, um ihre Anwendung bei der Ermittlung von Crowdsourcing-Werten zu bewerten. Die Studie zeigt, dass das Mitteln von Urteilen die Genauigkeit und Zuverlässigkeit verbessern kann, was besonders in den Gruppen erzwungen und Zeitliche Trennung offensichtlich ist. Sie hebt auch die Vorteile der Etablierung von Lücken zwischen Bewertungen hervor, was die Bedeutung der zeitlichen Trennung in Entscheidungssituationen impliziert. Dieses Experiment liefert nützliche Einblicke in die Nutzung der "Weisheit der Vielen" im Crowdfunding und betont die Bedeutung größerer Stichprobenzahlen und verbesserter Urteilsverarbeitung, um die Zuverlässigkeit der Ergebnisse zu gewährleisten. Die Ergebnisse ebnen den Weg für weitere Untersuchungen zur präzisen Vorhersage individueller Projektbewertungen, die Investoren helfen könnten, indem sie Ideen identifizieren, die wahrscheinlich erhebliche Investitionen erhalten.

Table of Contents

1. Introduction	5
2. Literature Review	8
2.1 Wisdom of Crowds and its Origins	8
2.2 Optimized Decision-Making Processes	12
2.2.1 Averaging Judgments	13
2.2.2 Dialectical Bootstrapping	16
2.3 Improving Judgment Reliability	18
2.3.1 Debiasing Judgments	19
2.3.2 Improving Reliability of Judgmental Forecasts.....	23
2.4 Crowds Wisdom in Online Platforms	25
2.5 Social Dynamics and Decision Making in Crowdfunding.....	27
3. Research Gap.....	31
4. Research Question & Hypothesis	32
5. Experiment Design & Methodology	33
5.1 Survey Structure.....	33
5.2 Experimental Groups	35
5.3 Variable Description	36
6. Results.....	38
6.1 Demographics	38
6.2 Error Analysis	40
6.3 Statistical Analysis	42
6.3.1 Mean Squared Error.....	42
6.3.2 One Sample T-test	45
6.3.3 Median Gain	47
7. Conclusion	50
8. Limitations and Future Research	51
9. References.....	52
10. List of Figures.....	56
11. List of Tables	56
12. Appendix – Survey Questions.....	57

1. Introduction

In recent years, the idea of the "Wisdom of Crowds" has attracted a growing amount of attention. It implies that a group's decisions are more likely to be correct and well-informed than those made by any one member acting alone. This phenomenon has been seen in a variety of contexts, including politics, sports, and the entertainment industry. Francis Galton, an English polymath who lived in the 19th century, is credited with developing the idea after realizing that group estimates of an object's weight were generally far more accurate than individual estimates (Fiechter & Kornell, 2021).

Despite the fact that Francis Galton was the first to explicitly present the theory of the Wisdom of Crowds in 1907, the concept of group decision-making has a considerably longer history. In the "psephism" method of crowdsourcing, for instance, residents would cast votes using pebbles or shells in ancient Greece. Jean-Jacques Rousseau, a French philosopher, famously asserted in the 18th century that "the general will is always in the right," arguing that group decision-making was preferable to individual judgment (Waldman, 1960). John von Neumann, an American mathematician of Hungarian descent, hypothesized that large organizations may make more accurate forecasts than small groups or individuals in the early 20th century. Other economists and sociologists refined this theory in the decades that followed. (Von Neumann, 1947).

Since then, the Wisdom of Crowds theory has been thoroughly investigated, and other scholars have looked into the elements that influence how well a crowd makes decisions. For instance, diversity of thought is important because it guarantees that the group has access to a variety of viewpoints and knowledge. It is also crucial for people to be able to learn from one another because this enables the group to gradually come to a more correct conclusion. The size, character, and structure of the group can also have an impact on how well judgments are made

(Surowiecki, 2005). A clear and well-defined task, the presence of good communication channels, and the degree of trust among group members are other crucial elements to take into account when it comes to successful crowd decision-making. Trust is essential because it encourages members to openly express their thoughts and opinions, which can result in better decision-making. Another reason why a task must be clear and well-defined is to guarantee that everyone in the group is working toward the same objectives and that everyone is aware of what needs to be done (Seifert & Weinhardt, 2006). The efficient flow of ideas and information among group members is made possible by effective communication channels (Surowiecki, 2005). The skill and knowledge of the group members can also play a significant role in deciding how well the decisions are made. Members who have specific knowledge or skills that are pertinent to the job at hand should be included in the group since their contribution can substantially improve the group's ability to make decisions. The degree of commitment and motivation among group members can also influence the outcome of crowd decision-making, since a highly dedicated and driven group is more likely to collaborate well and accomplish their objectives (Surowiecki, 2005).

In relation to the wisdom of crowds, the idea of optimal decision-making procedures has received much attention. According to certain research, for instance (Surowiecki, 2005; Vul & Pashler, 2012), diversity of opinion might be a key element in streamlining group decision-making procedures. According to research, various views and knowledge can facilitate better decision-making by enabling the group to take into account a variety of viewpoints and potential consequences (Herzog & Hertwig, 2014). Additionally, the optimization of group decision-making processes may be significantly aided by good communication channels. Effective information and idea exchange among group members is made possible by good communication, which may assist prevent misunderstandings and raise the standard of decision-making as a whole (Surowiecki, 2005). Optimized decision-making procedures may be essential in the crowdfunding environment for obtaining accurate project values and guaranteeing the success of campaigns.

The role of efficient decision-making processes in crowdfunding has been the subject of several research. For instance, studies have shown that successful crowdfunding campaigns depend on clear lines of communication and a consensus on the project's aims and objectives (Mollick, 2014; Belleflamme et al., 2014). Similar to this, the presence of a varied set of investors can improve decision-making processes and raise the chance of successful results (Herzog & Hertwig, 2014; Mollick, 2014). Additionally, using algorithms and decision-making tools can aid in the optimization of decision-making processes in crowdfunding by facilitating the effective analysis of massive volumes of data and offering insights into investor behaviour and preferences (Greenberg et al., 2013). Overall, improving decision-making procedures can be essential to the success of crowdsourcing campaigns, and more study in this area may assist to discover best practices and methods for raising the precision and effectiveness of collective decision-making in crowdsourcing.

As already seen, there are variety of studies that explore the concept Wisdom of Crowds, some of them even deal with decision-making processes. However, the premise of this work will come from investigating wisdom of crowds and different decision-making processes in business world, something that researchers are still to cover and investigate. Results of this work should also provide insights in development of best practices and methods for improving the precision and effectiveness of collective decision making in crowdfunding, that would hopefully have an impact on correct judgment of crowdfunding campaigns.

2. Literature Review

2.1 Wisdom of Crowds and its Origins

One can say that wisdom of crowds as a concept has existed ever since the existence of the mankind, however it was first pioneered by that name in 2004 by Surowiecki. Nevertheless, researchers like Francis Galton and C.H. Waddington have already conducted interesting experiments that would analyse the crowd's impact on one's judgment of a certain topic. Additionally, researchers like Bokouras, Sudgen, Page, Chen, Vaughan and many others have further investigated upon wisdom of crowds after Surowiecki's publications.

In 1907 Francis Galton conducted an interesting study in which he created a weight-judging competition of a rather heavy ox. In order to avoid biases and joking at the competition, Galton requested each participant to put a six-penny fee, which also ensured that competitors would try and give a best guess possible. He also argued that through this example he would be able to show how people judge the value of a politician, since the knowledge on the structure of the ox is rather similar to the knowledge of a voter before the elections. After collecting all of the results, it was found that the average judgment for the weight of the ox was 1997 lb and the original weight was just as close coming at 1198 lb. With this he concluded that the average estimate off a group is often more accurate than one of a particular individual. Francis called this state "Vox Populi" meaning voice of the people. This experiment is often cited as one of the earliest examples of the wisdom of crowds phenomenon and was later a pillar for further research on crowds judgment (Galton, Fp. 1907).

Since the pillar for the concept wisdom of crowds was established, many researchers have later embarked on exploring it further. A British developmental biologist by the name of Conrad Hal Waddington was one of the firsts that followed on Galton's research. In 1940 an experiment by the name "the estimation of the number of beans in a jar" was published in the journal of the

royal statistical society. In the experiment 787 undergraduate students were divided into smaller groups and each one of them was asked to provide an estimate of how many beans are in the jar, without having any prior knowledge or the ability to count it. Conrad then proceeded to analyse the data and found out that individual judgments were rather inaccurate, however when he averaged all of those judgments on a group level all of them were within a 0.8% margin of the actual number. The results suggested that the average estimate of larger groups of individuals could be highly accurate, even if the individual judgments were widely scattered or inaccurate (Waddington, C.H., 1947).

Another interesting case happened in 1986 in the explosion of Challenger Space Shuttle. On January 28, 1986 space shuttle exploded in the air which was followed by an immediate stock markets reaction. Investors started liquidating their holdings in the launch's four primary contractors, Morton Thiokol, Lockheed, Martin Marietta, and Rockwell International, as soon as the explosion occurred. The worst hit was Morton Thiokol's shares, which fell by about 12% by day's end. The other stocks only marginally recovered. Thiokol's shares fell sharply, suggesting that the market had swiftly identified the corporation as the cause of the catastrophe. Indeed, six months later, the Presidential Commission on the Challenger confirmed that the O-ring seals made by Thiokol, which lost resilience in cold weather, were the cause of the explosion. The market managed to identify four conditions that usually go with wise crowds. Those four conditions were later also discussed by James Surowiecki in his book. When it came to the Challenger accident, the market's collective assessment of Morton Thiokol as the main cause of the disaster was demonstrated by the rapid decrease in Thiokol's stock price relative to the other firms. This decision was made despite the fact that on the day of the disaster, Thiokol was not publicly accused, and the Presidential Commission did not confirm the exact cause—the O-ring failure—for months after the incident. This example demonstrates how the market's collective wisdom isn't merely the view of the majority; rather, it's the outcome of integrating a large number of individ-

ual judgments, each of which provides a tiny bit of knowledge or insight. Results from such a procedure may be remarkably accurate. This was demonstrated by the market's quick recognition of Morton Thiokol's responsibility in the Challenger accident. This phenomenon supports the theory that, in some situations, a group's collective judgment may be extremely insightful and, in some circumstances, even more so than the knowledge of the group's most knowledgeable member (Surowiecki, 2004).

Research and papers written by Francis Galton, Conrad Hal and Irving Janis as well as the interesting cases were all essential for the concept of wisdom of crowds. James Surowiecki, an American journalist was one of the first who collected all of the opinions and cases together and used them as a pedestal to define what the concept “wisdom of crowds” actually is and how it can be used in further research. Besides using their experiments, he also further explored the conditions under which groups are more likely to be wise or have even more accurate valuation. One of the conditions that he argued can have a major impact was diversity. James stated that it is important to include people from various experiences and viewpoints in order to increase the likelihood that the group would make appropriate conclusions as opposed to including people from similar backgrounds and viewpoints. Additionally, he explored the concept of independence, referring to the degree to which the members of the group are influenced by another member. He contends that when group members are overly affected by one another, they can be more inclined to adopt the views of the majority, which might result in incorrect conclusions. Decentralization, which refers to the division of decision-making authority among group members, was another concept investigated by Surowiecki. He contends that when decision-making authority is distributed across a group, rather than concentrated in the hands of a few people, the collective is more likely to make informed conclusions. Finally, Surowiecki also looked at the concept of aggregation, which is the act of merging the different viewpoints of group members to reach a conclusion as a group. He contends that the aggregation technique can significantly affect how accurately the

group decides. Investigation of these conditions was crucial to the wisdom of crowds since it helped to pinpoint the variables that are most likely to produce sound judgments in a collective context. Surowiecki was able to offer a more complex view of the wisdom of crowds and the elements that contribute to its efficacy by concentrating on the circumstances of variety, independence, decentralization, and aggregation (Surowiecki, 2004).

After Surowiecki many papers have researched the concept of wisdom of crowds, some notable ones were by Scott Page, Aristedis Boukouras and Rober Sudgen as well as Yiling Chen and Jennifer Wortman. In the paper called “Expertise, Diversity, and the Limits of Contemporary Wisdom” Scott Page examines the role of diversity and expertise in the wisdom of crowds. He contends that because they frequently have similar opinions and may exert significant influence on decision-making, specialists can sometimes restrict the wisdom of crowds. On the other side, a range of viewpoints and experiences can result in wiser choices. Additionally, Page contends that the intricacy of the issue at hand can have an impact on the wisdom of crowds, with more complicated issues needing a wider range of specialized knowledge (Page, 2007). Furthermore, Boukouras and Sudgen explored the factor of deliberation and how it can enhance the accuracy of the crowds decisions. The wisdom of crowds effect, according to the authors, may be enhanced when individuals are free to express their ideas and learn from one another. In an experiment, subjects were asked to make estimates of the worth of an unknowable number both on their own and after deliberating in groups. The findings demonstrated that group estimates fared better than individual estimates and that the most deliberative groups outperformed the others. This, according to the authors, emphasizes the significance of dialogue and cooperation in the decision-making process, supporting the notion of the wisdom of crowds (Boukouras & Sudgen, 2014). Two years later, authors Chen and Vaughan have investigated the impact of restricting participation in the wisdom of crowds. They found that in some cases, limiting the number of participants might be advantageous in order to increase the decision-making accuracy of the crowd. They take into account a

situation where people may choose whether or not to engage in a decision-making process, and they demonstrate how restricting participation to a select group of the best knowledgeable participants can provide superior results. This study expands on Surowiecki's findings by looking at how participation restrictions might maximize the wisdom of crowds (Chen & Vaughan, 2016). Of course, there were many other experiments conducted after Surowiecki defined the wisdom of crowds, however these were some that explored factors explained by him in further detail.

2.2 Optimized Decision-Making Processes

When considering the wisdom of crowds, according to the research *optimized decision-making processes* may be utilized to combine the assessments and viewpoints of several people in order to raise the standard of decision-making. They can assist in identifying the most accurate and trustworthy views or estimations by employing a structured and systematic strategy to balance the contributions of individual decision-makers. A systematic and organized strategy is used in optimized decision-making processes to raise the standard of judgment. It often entails locating and evaluating a variety of feasible possibilities, considering the advantages and disadvantages of each option, and choosing the optimal plan of action in accordance with a specified set of criteria. To aid decision-makers in understanding the possible outcomes of various alternatives, they could entail the use of analytical tools like decision trees, simulation models, and other methods. These methods may be used to estimate the likelihood of various outcomes, identify important variables that can influence the choice, and assess trade-offs between various possibilities. (National Academies of Sciences, Engineering, and Medicine, 2018).

2.2.1 Averaging Judgments

Papers written by Stefan Herzog and Ralph Hertwig as well as Edward Vul and Harold Pashler both used decision-making processes to assist in identifying the accuracy of collective wisdom. In the paper called “Measuring the Crowd Within” by Edward Vul and Harold Pashler research on wisdom of crowds was conducted. The research itself was rather similar to those conducted in the past. Approximately 428 participants were asked eight questions on general real-world knowledge. Interestingly enough participants were divided in two groups one which made a second guess on the given questions immediately (*immediate condition*) and the other which was asked to answer the questions for the second time only after three weeks’ time (*delayed condition*). In order to avoid participants researching answers none of the groups knew they were going to be asked to make a second guess on the questions. These two conditions helped the researchers identify the best possible ways of utilizing the average of both answers. What was rather interesting is that in both conditions the average of two guesses was more accurate than both guesses alone. In the Figure 1 one can see the impact of both guesses on the bar graph as well as the result of averaging. Vul and Pashler concluded that most individuals think their initial estimate about a fact provides all the information they could ever need, a forced second guess adds more details, making the average of the two guesses more accurate than either guess alone. This apparent advantage of averaging replies from the same subject shows that responses are sampled from an internal probability distribution rather than being deterministically chosen based on all available information. Edward and Harold found that temporal separation of guesses (an optimized decision-making process), which occurs when guesses are separated in time, a second estimate from the same person becomes more like to a guess from a different person, boosting the value of within-person averaging. These findings show that, in addition to having theoretical implications for the probabilistic nature of knowledge, the advantage of averaging two estimates from a single person can act as a quantitative measure of the value of “sleeping on it.” (Vul, E. &

Pashler, H., 2008). It also shows how optimized decision-making processes such as that one can be used to further boost the accuracy of the results.

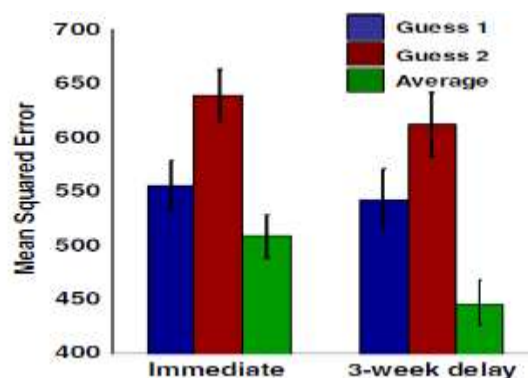


Figure 1 - MSE of all 3 conditions (Vul, E. & Pashler, H., 2008)

Vul and Pashler were not the only ones who explored the averaging of judgments in wisdom of crowds. In 1994 researchers Erev et.al have developed a model in which they suggested that trial by trial error in the judgment process may contribute significantly to the observed *overconfidence* in probability estimates. When discussing subjective probability estimations, the term "*trial-by-trial error*" refers to the variety or inconsistencies that arise in individual assessments over several trials or occurrences. This idea suggests that evaluations may vary from one trial to the next when someone is repeatedly judging or estimating the possibility of certain outcomes. Variations in the judgment process itself are more likely to be the cause of these fluctuations or mistakes than shifts in the underlying probability of the events. This kind of error can be influenced by several factors, one of them being cognitive biases where individuals are affected by their judgment differently in each trial. Information processing limitations is another one where most of us recall and process information differently across trials, on top of that contextual factors can have an impact since the framing or context of the issue can change the judgments. Lastly, emotional, or psychological states of an individual can impact the decision-making process as well (Eve, et al., 1998). Using this concept researcher Dan Airley and his colleagues have tried to test them over certain predictions. First of the predictions discussed the concept of

judges not influencing each other's judgments, thus allowing the average of the estimates to provide a clear and accurate assessment of the probability of an event. After the analysis, researchers found that mean subjective probability estimates of multiple judges are closely in line with the true statements. Adding to that they predicted that even when estimates are not entirely independent, but the error judgments are that the averaging process can still give an accurate probability estimate. The results gave an insight into the prediction as the judges were increased the reliability curve stabilized itself more, showing that the estimates should not be affected by independent errors and can result in well-calibrated estimates. Lastly, the third prediction analysed the contrasts between and within-subject averaging. In this case between-subject averaging would mean comparing different individuals while within-subject averaging would compare multiple estimates from the same individual. The results showed that between-subject averaging should give more accurate and well-calibrated results than within-subject averaging, however these correlations are quite close. Same as with Vul and Pashler this research suggests that averaging of estimates can significantly improve the accuracy and calibration of probability estimates (Ariely, D., et al., 2000).

Other researchers such as Thomas Wallsten and his colleagues argue that criteria for evaluating estimates should depend on the type of uncertainty represented by the decision-maker, which is in turn based on the decision makers model of the event space. Two types of uncertainty are presented: *aleatory uncertainty* and *epistemic uncertainty*. Aleatory are those events that are exchangeable, these events are similar in nature and can be considered equivalent in terms of their probability of occurrence. If such events are analysed, decision makers or judges can benefit from accurate estimates. On the other hand, epistemic are those events which are distinct and not interchangeable with others. In these cases, diagnostic estimates are more helpful as they provide insightful or revealing information about the specifics and unique nature of the event. Thomas et.al argue that it is important to understand the way decision makers perceive and categorize

events being exchangeable or unique, this will also influence the results achieved. As seen for exchangeable events the focus should be on accuracy since the events are similar and can be analysed in aggregate. Unique events on the other hand focus on gaining deeper insights or underestimation about that specific event, that is why diagnostic estimates are more reliable here. Averaging of estimates even when different types of events are considered will again increase the reliability of judgments, in their study researchers saw improvement in judgment reliability by averaging the probability estimates of as few as five subjects. Exchangeable events when averaged do have a more reliable and correct judgments that can give a precise estimation or likelihood of an event. However, unique events also do benefit from averaging of judgments, but those results give more diagnostic and deeper insight about the nature of the event that can serve as a basis for further estimates (Wallsten T, et.al, 1997).

2.2.2 Dialectical Bootstrapping

Another interesting study was conducted by Stefan Herzog and Ralph Hertwig in 2009. Two of them suggested that by averaging one estimate with another of the same person can result in improvements of the accuracy of their quantitative judgments, or so called a *dialectical estimate*. In that case a dialectical estimate would be the one coming from the same person but is based on different information and assumptions resulting in different inaccuracies than a first estimate. The process would of development would then compromise out of three stages: *thesis* (initial estimate), *antithesis* (dialectical estimate), and *synthesis* (aggregation of both). From that Herzog and Herwig derived the concept of *dialectical bootstrapping*. In order for average of two estimates to be more accurate than the first one has to compare errors of both. Errors in this case would be the distance of the first estimate from the true value of the result, which is also defined as the upper boundary. Lower boundary of this range would be on the opposite side, and it is 3 times ($3 \times (\text{first estimate} - \text{true value})$) as far from the truth as the first estimate. Even in the case where a dialectical estimate is located on the lower boundary the average of two estimates will

still lie exactly far from the true value as the first one. In the figure 2 below “a” would state the conditions for the dialectical bootstrapping, while “b” describes the situation in which a second estimate or a dialectical estimate in this case has the same value as the lower boundary. Second scenario is shown under “c” in which a dialectical estimate is closer to the true value or in this case has a value of 80 the average would then result in the value of 95, having an absolute error of just 5. This just goes to show that even if the dialectical estimate is 3 times as large as the error of the first estimate, averaging the values would still achieve better or at least the same results than just the first estimate. Researchers concluded that as long as the errors of the first and dialectical estimate have nonidentical random or systematic errors and the dialectical estimate is not too far off the mark, the dialectical average will likely be more accurate than the first estimate (Herzog & Hertwig, 2008).

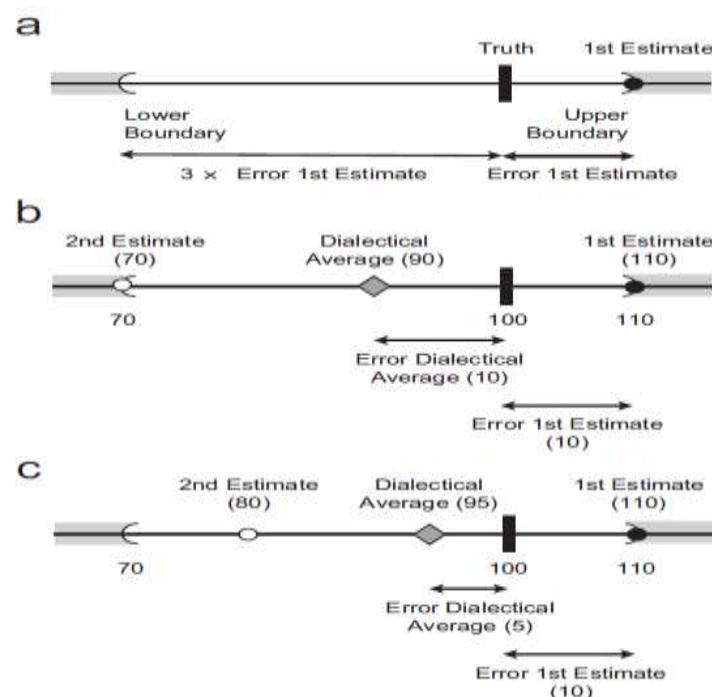


Figure 2 - Dialectical Bootstrapping (Herzog & Hertwig, 2008).

After defining dialectical bootstrapping methods, both Stefan and Ralph tried to implement it in the experiment they conducted in 2008. Approximately 100 students out of which 76% female from the university of Basel were participating in the experiment. In the experiment par-

ticipants were asked to give estimation on dates of 40 different historical events ranging from 16th – 19th century. Participants were divided in two groups, both of which would make a second assumption, however one of the groups would have a dialectical-bootstrapping condition connected to them. Dialectical-bootstrapping group was shown their first guess and asked a few defining questions that would guide them towards better assumptions as well as given a stimulant in the form of reward if their assumptions were more accurate or decreased their absolute errors. After all the results were submitted, Herzog and Hertwig would calculate the median absolute deviation between estimates of the participants and the correct dates and after average the measure across all participants. First estimates were off by 130.8 years, repeated estimates were off by 126.5 years and dialectical estimates were off by 123.2 years, lowest of them all. Researchers noticed that around 72% of participants in the dialectical-bootstrapping group have benefited from it, meaning that their results were more accurate. However, participants would actually benefit more from asking others on their averages and estimating their results based on that rather than using a dialectical-bootstrapping method. Researchers concluded that dialectical bootstrapping can indeed increase the wisdom of the person however not of the whole crowd thus it is better to average repeated estimates in order to gain faster accuracy or let participants share their estimates and draw conclusions based on them (Herzog & Hertwig, 2008).

2.3 Improving Judgment Reliability

Shifting the attention from the topics of decision-making processes and the wisdom of crowds to how biases might be understood and minimized to improve forecasts and choices. This change is significant because it clarifies the relationship between eliminating bias and enhancing judgment. Whether we're making decisions individually or as a group, knowing about these concepts can help us make smarter, more accurate decisions. These ideas are connected in the next part, which demonstrates their significance in both individual and group decision-making.

2.3.1 Debiasing Judgments

Using various techniques such as averaging judgments and dialectical bootstrapping can certainly lead to improvements in the reliability of ones or even groups judgments. However, for those judgments to be completely reliable one must consider biases that occur as these judgments are made. Cognitive biases are very common in any of the decision-making processes we go through. They are regular patterns of judgmental deviations from rationality or the norm, where people construct their own "subjective social reality" based on how they interpret the information. These biases frequently arise from the brain's desire to streamline information processing and are a typical feature of human psychology. There are many techniques which help debias judgments, and these techniques are particularly important where objective decision-making is necessary. However, it is often challenging or even impossible to completely debias a certain judgment (Kennedy J., 1995).

One of the biases that is often seen would be *explanation bias*. It is a cognitive phenomenon where explaining or imaging hypothetical future events increases the perceived likelihood of those events. These bias states that people judge the likelihood of an event based on the memory with which they can recall or imagine examples of it. In 1995 two researchers Edward Hirt and Keith Markman investigated the concept of explanation bias further. They designed three experiments in which they considered that suggesting any alternative outcome can lead to the simulation of multiple scenarios, thus creating debiased judgments. All three experiments showed similar trendlines when it comes to their results. In study 1 participants were asked to consider a relationship between two firefighters, one of which was successful while the other was not. One group was asked to explain one of the three possible relationships: positive, negative, or no relationship while the other group was asked to explore alternative relationships, resulting in six additional explanation conditions. Those who were asked to estimate only one possible type of relationship remained in that condition or changed slightly, while those with judgment exploration

were debiased and gave different judgments the other time. Further in the study 2 similar experiment was conducted where participants were asked to suggest hypothetical scenarios of an upcoming high school game. Following the pattern this study gave insights into exploration biases, where participants who were engaged in explanation task had debiased judgments but also those who were simply asked to explain a different version of the same outcome. Study 3 added on the both, here participants were asked to suggest why a top contender of NL East teams was expected to win the division based on the evidence provided. Suggesting to participants to explore alternative options (in this case contender teams) effectively debiased their judgments. However, when they were asked to do the same with a team that was not contender, their judgments remained the same, suggesting that alternative being considering is also crucial to have an effective biasing. As seen in all three studies, *counterexplanation* can lead to debiased judgments, especially when alternative explanations considered are possible. Engaging in counterexplanation tasks will mostly likely result in multiple alternatives to be considered and effectively reduced any biases present (Hirt, E.R., & Markman, K.D., 1995).

Besides not exploring additional possibilities within judgments many of us also tend to formulate a standard for certain opinions and never revise it again. In psychology this effect is called *Anchoring effect* which occurs when estimates or opinions are influenced by initially considered standard, also known as anchor. That is why it is important to keep in mind that such effects do occur more often and find ways to reduce them as much as possible. In year 2000 three researchers Thomas Mussweiler, Fritz Strack and Tim Pfeiffer have conducted research where they tested the anchoring effects and suggested a model that can mediate the effect itself. Three of them have found out that anchoring effects are extremely common and usually occur even if the anchor values are clearly uninformative for critical estimate. On top of that they are also independent of participants motivation, in some cases participants of other studies were given incentives to improve accuracy of best estimates, but even this had no change in the anchoring effect of

the estimate. What is even more astonishing is the fact that experience in the field and explicit instructions to correct potential anchor also cause no change in this effect. This also comes from a *Selective Accessibility Model* which suggests that the anchoring effects are mediated by the increased accessibility of anchor-consistent knowledge. Meaning that when individuals make judgments, they selectively retrieve information from memory that aligns with the anchor, influencing their final estimates. Researchers suggest that a person should be more exposed to anchor-inconsistent knowledge in order to reduce or even completely correct the anchoring effect. Besides that, Mussweiler, Strack and Pfeiffer suggest an interesting strategy called *Consider the Opposite*. Through this strategy participants are asked to actively generate reasons or arguments that contradict the anchor. With this the decision makers perspective can be broadened, making them more aware of information that is inconsistent with the anchor. In one of the studies three of them conducted they asked car experts to estimate the value of a used car. This was done in a real-world environment and was not controlled. They found out that experts who were prompted to consider the opposite of the anchor provided estimates that were less influenced by it and more accurate, thus reducing its biasing effect. In the second study researchers analysed anchoring effects in a more controlled environment. Recruiting 31 students, they asked them to indicate whether likelihood of particular Chancellor winning the next election is higher or lower than either 20% or 80%. Those who were asked whether this percentage is higher were also asked to list three arguments that implied that chancellor would win the next election. While those who were asked if the percentage is lower were also asked to list three arguments that spoke against those possibilities. The results indicated that participants who considered reasons against the anchor were less influenced by it in their judgments which was very similar to the first study in the real-world setting. These studies just go to show that it is important to encourage a more balanced consideration of information. It also helps to mitigate one of the common cognitive biases that affect decision-making processes, making the suggested idea valuable tool in various estimation and judgment scenarios (Mussweiler, T., 2000).

Authors like Hal R. Arkes argue that it is not just enough to recognize that biases exist and implement techniques that can tackle them, but it is also important to realise what kind of errors are occurring. Categorizing them or setting a proper taxonomy system, can lead to higher success rates of certain debiasing techniques. In 1991 Arkes suggested that errors can be divided into three categories: Strategy-Based Judgment Error, Association-Based Judgment Error and Psychophysical-Based Error. *Strategy-Based Judgment Errors* mostly occur when individuals use a suboptimal strategy for decision-making. This typically happens because the effort or cost of employing a more sophisticated, optimal strategy exceeds the perceived benefits. An example of this would be a person using a quick way to solve a specific problem since the stakes of his judgments are low, thus the time and effort saved by using simpler strategy can outweigh the cost of potential errors. To tackle this research shows that people will often shift to a more sophisticated strategies when they recognize higher stakes. Furthermore, *Association-Based Judgment Error* is connected to system of associations within our semantic memory. These associations are generally beneficial but can become counterproductive in some contexts. Hal gives an example of associating hometown with sports players success, where people will rate a player higher just because another successful player comes from the same city. Unfortunately, there are no specific techniques to tackle these errors as they occur spontaneously and unconsciously hence traditional debiasing techniques such as incentives or warnings tend to be less effective. Lastly *Psychophysically Based Errors* will result from how people perceive and psychologically interpret physical stimuli. This psychophysical adaptation can influence our decision-making processes. One example would be economic decisions, where people might not accurately perceive the difference between very high costs or gains, leading to errors like the sunk cost effects. In this case people might irrationally continue investing in a losing aspect just because psychological impact of additional losses is not present. In order to tackle this error, Arkes purposes various techniques ranging from simple awareness and education of individuals and feedback as well as calibration to

complete system redesign and environmental modification as well as bringing in additional technology and tools can reduce the rate of these errors occurring (Arkes, H.R., 1991).

2.3.2 Improving Reliability of Judgmental Forecasts

Debiasing judgments is not the only possible way to improve judgments. Researcher Thomas R. Stewart has published a paper in 2001 called “Improving Reliability of Judgmental Forecasts”. In this paper Stewart addresses the inherent unreliability of judgmental forecasts. According to him unreliability originates from the two main factors: the acquisition and the processing of given information. On top of that judgments also tend to be less reliable as the task complexity increases, humans tend to find tasks more difficult as their processing power is much lower than the one of a computer for example. Environmental uncertainty such as rapidly changing conditions or lack of clear data can also lead to poor judgments. As we have seen before creating patterns or relying on memory can definitely impact the quality of one’s judgment, as well as the use of intuition rather than a proper analysis.

To tackle mentioned factors Stewart suggests five principles that can improve the reliability of judgmental forecasts (Stewart T.R., 2001):

1. **Organize and Present Information Clearly:** Emphasize relevant information and limit reliance on short-term memory and complex pattern recognition.
2. **Limit Information Used:** Focus on a few important cues to prevent cognitive overload and distractions.
3. **Use Mechanical Methods for Information Processing:** Employ computerized models, even simple ones, to process information more reliably than human intuition.
4. **Combine Several Forecasts:** Aggregate multiple forecasts to average out unsystematic errors and improve reliability.

5. **Require Justification of Forecasts:** Encourage forecasters to justify their predictions verbally, especially in low predictability scenarios, to improve consistency and move towards more analytical processes.

Besides five principles suggested by Thomas Stewart, a useful framework model has been developed by Egon Brunswick. One way to understand this model is to compare it to the camera lens focusing on the scene. On one side, there is real world with various cues or signals, while on the other there is a person making a decision or judgment based on those cues. The person would go to predict what is happening based on those cues and formulate a judgment. However, not all cues are seen or interpreted accurately. According to Brunswick the accuracy depends on how well the cues in the environment align with the persons interpretation of them. Reliability would refer to how consistently a person can make the same judgment when presented with the same or similar cue.

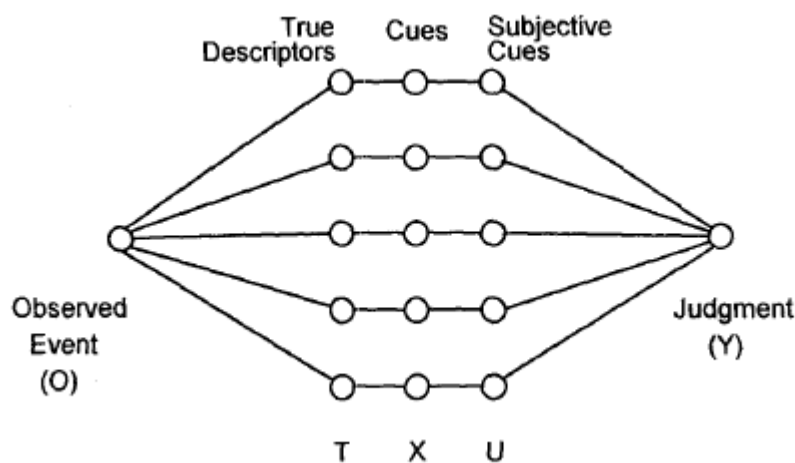


Figure 3 - An expanded version of Brunswiks Lens Model (Stewart T.R., 2001)

In the figure above we can see three factors that can have an effect on the observed event (O). True descriptors (T) would be cues available in the environment that are directly observable or measurable. They are mostly factual and are not influenced by a personal perception. Subjective Cues (U) are cues that individuals integrate into their judgments based on their perception. They are usually influenced by individuals experience, biases, and interpretations.

Objective Cues (X) are cues that are presented by factual data at hand (for example sales figures or weather measurements). According to Brunswick, using Lenses model can help quantify the relationship between forecast accuracy and reliability (Stewart T.R, 2001). However, little is still known on the reliability of judgments as most of the data can be hard to quantify and transfer into numerical figures.

2.4 Crowds Wisdom in Online Platforms

It was clearly noted how more accurate the average judgments are rather than the individual ones (wisdom of crowds). These judgments are often connected through different environments. One of those environments that has exponentially evolved in the past decade are online platforms. Online platforms offer a different approach to communication that can lead to all kinds of new and interesting ways for humans to interact. This then leads to crowds being able to work together in order to formulate valuable, more credible and more accurate opinions and statements that can be shared publicly (Johnson, 2022). Some examples of those would be Wikipedia, Reddit, Amazon and even Kickstarter.

In some cases, there are limitations to the statement that a group will collectively be able to formulate a more accurate opinion than an individual. One interesting research performed by S. Bhatt et al. has used Twitter as the source in order to find out whether it was possible to take out a measure of a crowd's diversity based on its members as well as whether those measures can have a selection of wiser crowds. To be more specific this research has focused itself on the fantasy sports and the crowds that select players based on tweets of other users and compared the performance of those relevant to those who do not use tweets and are randomly sampled (S. Bhatt et al. 2017). Interestingly the research has given support for the stated point to a certain extent. For example, it was interesting to see that the crowds of people who were in a group of 10 or larger had lower average pick diversity of their player than the smaller groups (less than 6), which would

have more unique picks in their roster. Furthermore, the research seemed to show that randomly diverse groups tend to have lower pick diversity than those who are more or less diverse (S. Bhatt et al. 2017). This could indicate that random selection of crowds tends to lead to similar opinions rather than those that are more or less diverse, seemingly diversity influences the connection between crowds in a more negative manner.

In the last paper it was clearly stated how individuals can benefit from wisdom of crowds to improve their own sphere and generate personal benefit. However, there are cases where a wisdom of crowds can be used to generate improvements of organisations and/or teams. In a paper published by Kim & Mijung in 2019 research was conducted using text mining and cluster analysis to identify opportunities to improve team performance in soccer. In the study a researcher has used online fan comments on a team performance opportunity during FIFA World Cup qualification (Y. Kim & M. Kim, 2019). The paper wanted to take a different point of view while using big data. The algorithm used a fan-centric view that was left as a comment on various social networking platforms with a goal of identifying opportunities for improvement. This of course takes another perspective as in these cases the data that would be used usually is not coming from fans but rather from the field or experts. Using fan related data has enabled researchers to give an overview on how crowds itself can positively serve to give better insights in the problem. By grouping the comments in different topics based on which topic was most common and often mentioned, the researchers have managed to identify that highest opportunity lays in governance body specifically in management practices (Y. Kim & M. Kim, 2019). Opportunities that were identified in the study are also considered as the “Wisdom of Crowds” since they are judgments of significant number of people/groups and are argued to be valuable advice to the coaches that can later improve performance of the team.

Sport and fantasy football is only one of the topics where social media and wisdom of crowds have been connected. Another interesting topic that has been getting a lot more attention in the past five years with the rise of different applications and ways of investing is certainly the investing sector. Unlike in the case of sports and fantasy football, where wisdom of crowds tends to lead to more reliable opinions and positive outcomes, investment branch is one of those cases where crowds' opinion could lead to negative outcomes for the individuals. In the research done by Chen and Xie it was found that 82% of consumers tend to make decisions for purchase based on peer opinions found online (Chen and Xie, 2008.). Further research has evolved in the past years and one interesting paper written by Chen et al. finds strong support on the statement that opinions that are publicly relieved on social networks tend to strongly predict the future of stock returns and earnings (Chen et al , 2014). Additionally, the researchers seem to give insights similar to the analysis of S. Bhatt et al. where crowds that are more experienced are able to differentiate between authors who historically offer good advice versus those who offer bad advice (Chen et al., 2014). This would then lead to investors achieving more positive results and making good investment decisions based on public opinions shared by users and reactions of crowds. Nevertheless, according to many sources online most of the investors (approximately 90%) still manage to lose money by relying on public or crowd opinions (S. Josh, 2022).

2.5 Social Dynamics and Decision Making in Crowdfunding

Crowdfunding platforms like Kickstarter are not only spaces for individuals to showcase their projects but also vibrant communities where social interactions and network effects play a significant role. In order for a project to succeed or raise enough money, besides marketing and having a quality product, one has to focus on the decision-making processes of crowds and how they influence each other in order to fund a certain project. Researchers like Ethan Mollick have explored the dynamics of crowdfunding while his colleagues Paul Belleflamme, Thomas Lambert, and Armin Schwienbacher have tried to explore the network effects in crowdfunding. These

dynamics of crowdfunding as well as the network effects can help identify factors that can influence crowds ability to judge a certain project but also decide what project to support.

Since very little was known what factors do actually influence a crowdfunding project, in 2013 Ethan Mollick has decided to conduct a research project in order to find out the dynamics of crowdfunding. According to Mollick's analysis, project parameters have a big influence on the profitability and value of a campaign. According to the data, campaigns with strong narratives are 2.3 times more likely than those with weaker ones to reach their donation targets. Additionally, initiatives that successfully use multimedia content, such as captivating movies and high-quality photos, have financing success rates that are increased by 50%. These results emphasize how crucial it is to create a compelling project presentation in order to maximize the attractiveness to potential funders. The study also underlines the crucial part that social factors play in the success of crowdfunding. According to the statistics, projects with more project updates obtain, on average, 3.5 times more financing than those with fewer updates. Furthermore, initiatives with more comments have a 1.8 times better chance of achieving their financing targets. These figures highlight the importance of active participation in and communication with the crowdfunding community in influencing campaign outcomes. Additionally, the study shows that initiatives with substantial social capital—as shown by connections and endorsements—raise, on average, 2.6 times more money than those without it. This is especially true for campaigns started by artists. These figures highlight the value of creating and utilizing social networks to raise Kickstarter worth. Mollick's study demonstrates unique trends in funding accumulation over time in terms of temporal dynamics. According to the statistics, successful campaigns saw a spike in fundraising at the beginning, with an average of 30% of the total cash received in the first 48 hours. Additionally, during particular times throughout a campaign, such as the midway or the final days, financing is sometimes expedited. Project developers may take advantage of these crucial financ-

ing milestones by carefully planning their campaign efforts and streamlining their decision-making processes with the aid of an understanding of these temporal patterns (Mollick E., 2013).

Knowing the basis of dynamics of crowdfunding projects, researchers like Paul Belleflamme, Thomas Lambert, and Armin Schwenbacher have investigated at whether network effects are present and how they affect crowdfunding, concentrating on Ulule, a prominent reward-based crowdfunding site in Europe. The authors' goal is to comprehend user interdependencies and the value creation that results on digital platforms powered by network effects. The study demonstrates the pervasiveness of network effects in crowdfunding, which result from both increasing usage and engagement. On the crowdfunding site, it is discovered that network effects associated with greater utilization are prevalent. The authors present evidence of intra-project network effects, wherein past supporters' contributions to a given project have a favourable impact on current backers' donations. It's interesting that the brand-new discovery of favourable inter-project network effects is presented. It demonstrates how recent donations to other concurrent projects on the crowdfunding platform grow with recent contributions to one project. This demonstrates how interrelated crowdfunding campaigns are and implies that financing for one project may have an impact on donations to other initiatives. Additionally, researchers have looked at how network effects affect diverse user groups, including investors and business owners. The authors discover that the presence of extra entrepreneurs has a favourable impact on backers' inclination to make more pledges on the crowdfunding site. This suggests the existence of cross-group network effects, whereby businesspeople draw in more investors, creating a positive feedback loop that encourages platform expansion. However, the research also finds within-group network effects that act the other way around. Within-group network effects describe how other supporters affect a backer's likelihood of making another contribution. The results of the investigation show that the probability of repeated donations is negatively impacted by the number of backers in the group. The study emphasizes the significance of project selection and the

mix of projects on a platform to optimize contributions and take advantage of synergies. As initiatives with a larger percentage of recurring supporters tend to produce more donations, it also underlines the need of keeping current backers. According to the Paul, Thomas and Armin, reward-based crowdfunding displays a "winner-takes-all" market dynamic from the perspective of competitiveness. The research has found that the favourable network effects enhance successful crowdfunding platforms while potentially harming lesser ones. Smaller platforms may need to specialize or merge in order to increase their chances of surviving. This implies that while platform expansion is driven by cross-group network effects, its size may be constrained by within-group network effects (Belleflamme, P., Lambert, T. and Schwienbacher, A. 2018).

After carefully reviewing both papers one can see how certain factors like strong narratives, multimedia content, engagement of funders, supporters' contributions and many others can change the course of a single crowdfunding campaign. Companies that can identify such factors ahead of their projects can certainly benefit from a faster start as well as increased popularity. However, these factors can also be beneficial to the investors or those who support crowdfunding projects. It would be interesting to see what values individuals would give about certain projects after knowing some of the factors that social dynamics factors and how they affect the decision-making processes of those crowds.

3. Research Gap

Taking a look at the literature presented averaging judgments should lead to more precise results, especially if these judgments are averaged on a group level. This concept is called Wisdom of Crowds and was already investigated in research by many individuals, two of them being Edward Vul & Harold Pashler in 2008. Two of them found out that even though averaging groups judgment can lead to improvement of the judgment in general additional techniques such as a temporal separation where judgments are separated by certain time frame will lead to even higher accuracy of judgments. Furthermore, researchers Stefan Herzog and Ralph Hertwig have conducted research in 2009 in which they found out that additional decision-making tool called dialectical bootstrapping can as well improve the accuracy of judgments. With this technique participants are asked to make judgments that are at least partially different from their first, thus reducing the systematic error and completely cancelling random error (Stefan & Herzog, 2009).

Considering all of this it would indeed be interesting to find out which of the decision-making methods is the most effective and on top of that whether they can also help with estimating Kickstarter valuations without any prior knowledge. Up until now most of the research has focused itself on investigating the concept wisdom of crowds in so called ever-day situations and research has usually focused on judgments that were made on general topics. Research on other topics such as funding projects valuations has not been conducted up until now. That is why being able to create valuations of certain funding projects would certainly help both investors and creators have a picture of what is to be expected from that specific project. In this research, the focus will be on investigating Kickstarter valuations with three different decision-making methods: temporal separation, forced second judgments and dialectical bootstrapping as presented in the past research.

4. Research Question & Hypothesis

In order to explore a relationship of averaging judgments in the context of crowd wisdom the following research question and hypothesis have been formulated.

Research Question:

How do optimized decision-making techniques enhance the accuracy and reliability of collective judgments within the framework of the wisdom of crowd's theory in evaluating Kickstarter projects?

H₀ (Null Hypothesis) – The use of different decision-making techniques does not significantly affect the accuracy and reliability of collective judgments in Kickstarter project evaluations.

H₁ (Alternative Hypothesis) – The use of different decision-making techniques significantly enhances the accuracy and reliability of collective judgments in Kickstarter project evaluations.

H_{1a} – Forced technique significantly enhances the accuracy and reliability of collective judgments in Kickstarter project evaluations.

H_{1b} – Temporal separation significantly enhances the accuracy and reliability of collective judgments in Kickstarter project evaluations.

H_{1c} – Dialectical bootstrapping significantly enhances the accuracy and reliability of collective judgments in Kickstarter project evaluations.

To evaluate the precision of a certain judgment a specific true value will be taken as benchmark. Since the investors can continue to support a certain project, a representative of all the true values was taken on 27th November 2023. Thereafter the true values will be subtracted from the averaged estimate showing the value of the error. A minimal error indicates that the participants estimation closely aligns with the true value, thereby denoting a high degree of accuracy and reliability in their judgment.

5. Experiment Design & Methodology

Data collection for this project involves distinct obstacles in real-world settings. As a result, data collection will be handled using an online survey created with Google Forms. The structure of the experiment is shown in the picture below. While each group will complete a unique survey tailored to the many decision-making approaches, they will all follow a single overall methodology. The architecture and complexities of the survey are explained in the next section.

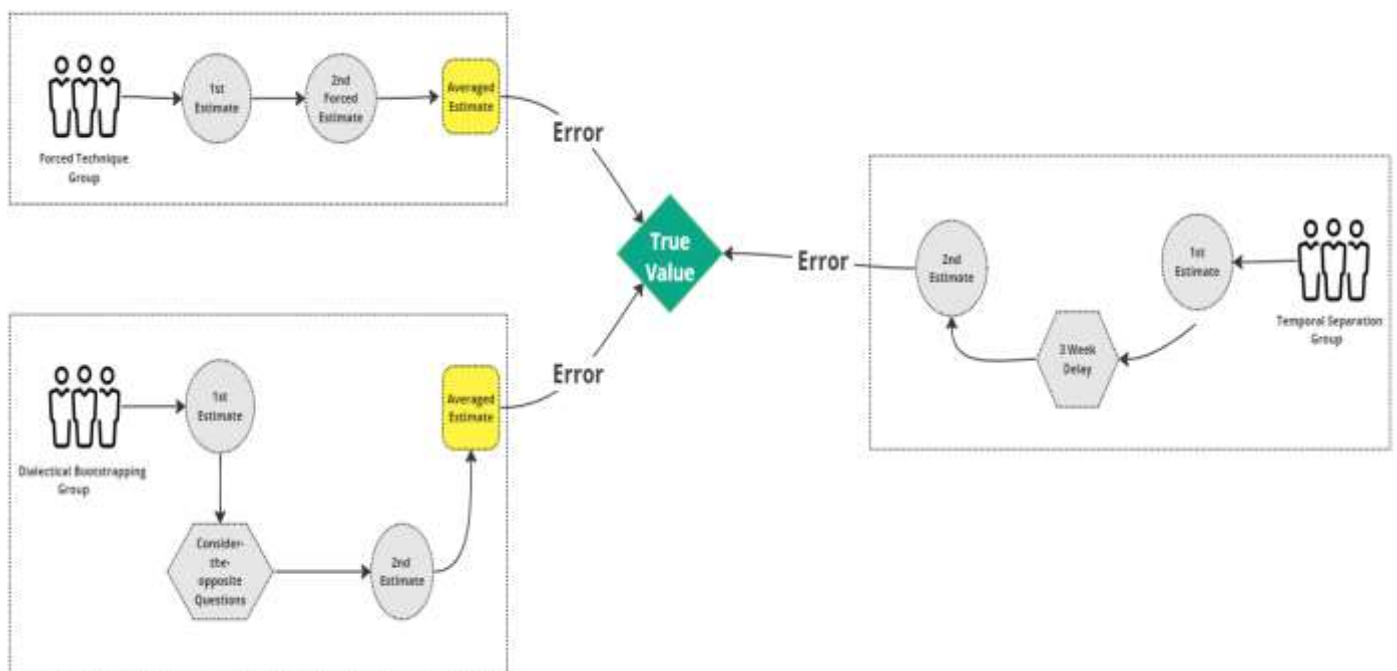


Figure 4 - Experiment Design

5.1 Survey Structure

1. **Introduction** – Participants will be introduced to Kickstarter; in this section they will read a short paragraph on Kickstarter and the whole concept behind it. This is mostly due to a fact that not all participants are familiar with Kickstarter or even the concept of crowdfunding. Besides that, they will also get to see statistics of all successfully funded

projects in Kickstarter until now. This will enable them to get a grasp of values of all successfully funded projects and ensure that the data is reliable, it should also make participants stay away from assuming that they are asked about valuation of product prices. Additionally, they will also get a short explanation on what their role in this project is and how helpful they can be.

2. **General Questions** – in this section all participants will have to fill in general questions such as their age group, their sex and their highest education level. These questions will help with data analysis and open a possibility to analyse any trends, patterns, or correlations.
3. **Kickstarter projects funding valuations 1st judgment** – the main part of the survey. Participants will be introduced to 10 different Kickstarter projects, all of them varying in funding collected, specifications and design. There is no trend in these projects, and they were ordered completely randomly in the survey to ensure validity of data and also let participants have clear judgments without any pattern recognition. Some of the projects have already collected their target funding while others did not. However, participants will now know which so that the judgments are only based on their intuition.
4. **Break section** – this section is different depending on the group a participant was assigned to. Further explanation is discussed in the next paragraph.
5. **Kickstarter projects funding valuations 2nd judgment** – after participants have wrapped up a 1st section and were either asked to go to the next one or asked a set of questions to consider the opposite or have wrapped up the survey, they will then follow either to give judgments immediately afterwards or give judgments 3 weeks after on the same projects as in the first section.

5.2 Experimental Groups

The experiment will be separated in three groups:

1. **Forced Technique (*controlled group*)** – This group will follow a simple structure of being asked to give judgments twice immediately after they have done it once. There will be no additional instructions given to this group. It will be used as a *control group* for data comparison.

2. **Dialectical Bootstrapping (*target group*)** – Participants of this group will be asked a set of consider-the-opposite questions, these questions will ensure that confirmation bias is reduced, critical thinking is enhanced and hopefully the accuracy of judgments is also increased. This group will act as a *target group*. The questions that are given in the survey are following:

- a) Assume your first estimate is off the mark. Why might this be the case? Reflect on this possibility.
- b) Think about a few reasons why your initial estimate could have been inaccurate. Which assumptions and considerations might have been wrong?
- c) What do these new considerations imply about your first estimate? Was it perhaps too high or too low?
- d) Based on this new perspective, make a second, alternative estimate on the funding each project has raised.

3. **Temporal Separation (*target group*)** – In this group participants will be asked to give judgments on 10 projects and will end the survey. However, they are not aware that they will be asked to provide another judgment on the same 10 projects three weeks after. To ensure that the same participants are giving judgments and to be able to compare them, participants of this group will be asked to provide their email address before starting the survey. As seen

in the research this process will increase the independence of guesses, making a second guess feel like a guess from completely different person.

5.3 Variable Description

Having constructed the experimental design and the survey, a foundation for identifying key variables was established. These variables that are crucial for the experiment at hand are as follows:

- **Independent Variable**

- Decision-Making techniques – being the primary focus of the study and variables whose effect is observed. All three of the techniques: temporal separation, dialectical bootstrapping and forced technique based on both studies of Herzog and Hertwig (2009) and Vul and Paschler (2008) will be observed to see what effects they have on judgment accuracy and reliability.

- **Dependant Variable**

- Judgment accuracy and reliability – depending on each decision-making technique the judgment accuracy and reliability will be investigated. Participants will be able to give judgments in a form of numbers, or total funds collected in Euros by each Kickstarter project to be more exact. All surveys are restraining participants to give answers in any form besides numeric. Would a participant try to type a non-numeric answer they would see the following message: “Please enter an estimate in euros as a whole number”.

- **Controlled Variable**

- Kickstarter project characteristics – in order to ensure uniformity all Kickstarter projects will have a same description. The characteristics will consist out of two

paragraphs describing the design and features of the product with specific features being bolded in the text. Additionally, all of the projects will have two pictures of the product, in most cases this will consist out of one photo and one gif which is basically few pictures of the product in one to give participants a better grasp of the product and its design. The description and pictures are all the same in each of the groups as well as in questions of the second judgments.

- **Mediator and Moderator Variables**

- Cognitive processes (mediator) – each participant has a unique way of processing information and based on that information they will make judgments about Kickstarter projects. Different decision-making techniques could in this case influence these cognitive processes which can also influence the accuracy and reliability of judgments. This was already seen in research by Hirt E.R. and Markman D (1995), where they used counterexplanation to debias judgments.
- Familiarity with Crowdfunding/Kickstarter (moderator) – the exact influence of different decision-making techniques could definitely vary depending on how well participants understand crowdfunding platforms in general. Participants who are more familiar with this concept might have a better grasp of which figures could be correct. Furthermore, participants who are even more experienced with Kickstarter could even know some of the products and their fundings collected possibility giving completely correct answers.

6. Results

To gather data for the current experiment, surveys were distributed across various sources. The first two groups, Forced and Dialectical, predominantly sourced their data through diverse social networks such as LinkedIn, university groups, forums, and also through sharing with work colleagues. The Temporal Separation group, however, employed a different strategy. Given that this group's data relied on participants responding again within three weeks, email addresses were collected to ensure that the initial responses could be matched with the same individuals. To address concerns about reliability, friends and family were primarily approached to respond to the survey for this group.

6.1 Demographics

There was a total of 103 participants responding to all the survey questions. 33 of them came from Forced technique group, other 35 were part of the dialectical bootstrapping group while 35 participated as the part of the temporal separation group. Demographically speaking, 67

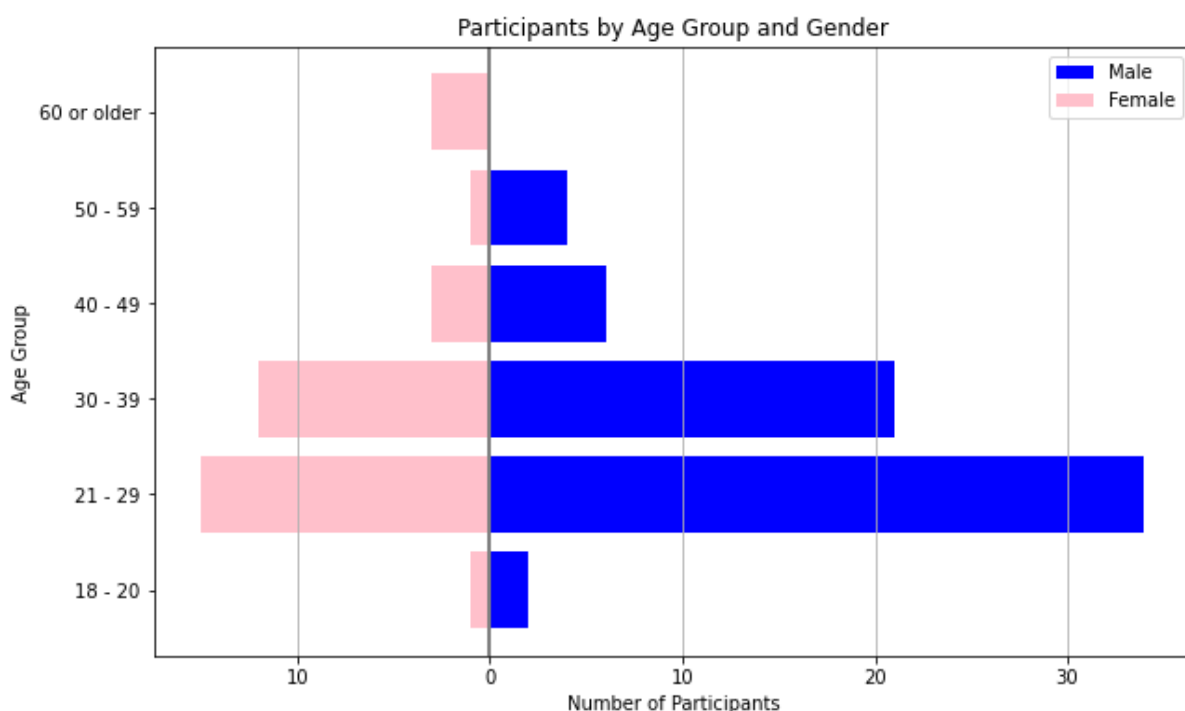


Figure 5.1 - Participants by Age and Gender

participants were male while 35 were female. Participants were also asked to answer which age group they belong to. As seen in the previous figure most of the participants were aged from 21-29, 35 of them were male and 15 females. Additionally, most of the participants came from the age group 30-39, out of which 21 were males and 12 were females. Only 2 males were in the age group of 18-20 while 6 males were aged from 40-49 and 4 were aged 50-59. On the other hand, only one female was in the age group of 18-20 and 50-59 while 3 of them were aged 40-49 and interestingly 3 of them were in the age group of 60-69.

Furthermore, participants were asked to answer an additional question on their highest educational level obtained. Most of the participants had some sort of higher education. This data is important as it shows that most of the participants should be familiar with the concept of crowdfunding and funding in general as well as with Kickstarter as a whole.

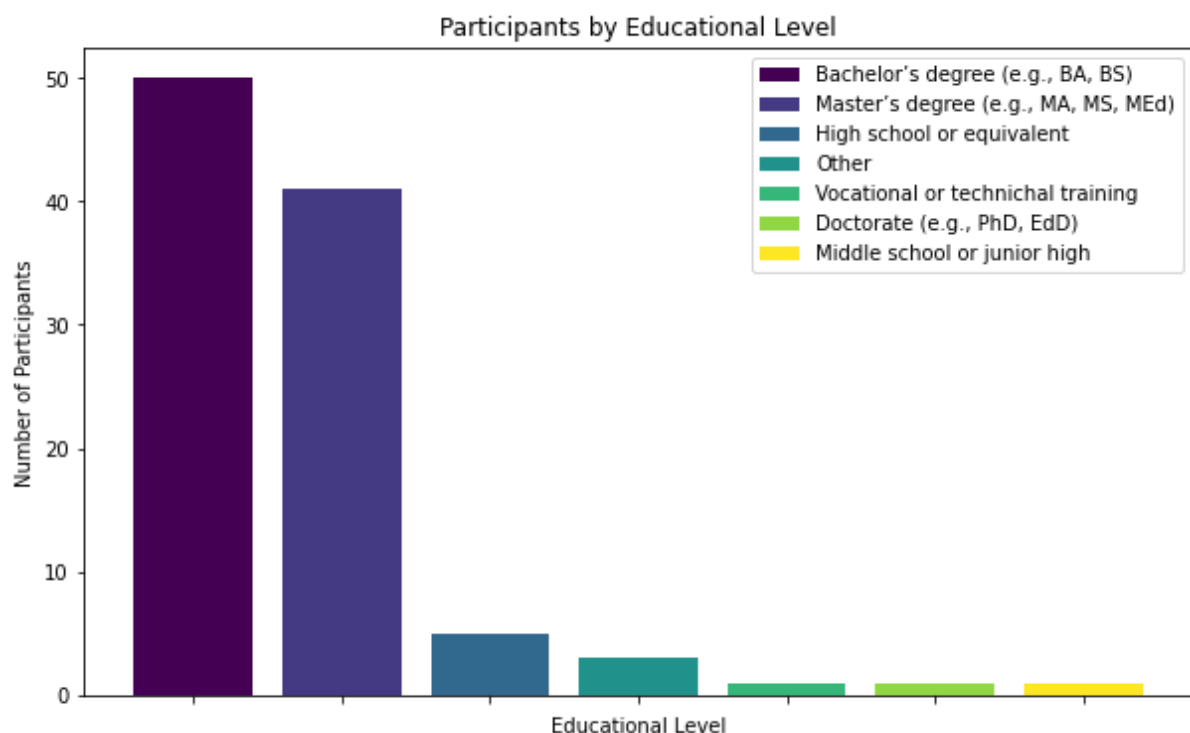


Figure 5.2 - Participants by Educational Level

According to the graph above 50 participants have already obtained their Bachelor's degree, while 41 of them obtained a Master's degree as well. Doctorate, Middle school and technical training was obtained by 1 participant only. Furthermore, 3 participants obtained other form of education while 5 of them have high school or equivalent as their current educational status.

It should be noted that 13 participants were excluded from the analysis due to ambiguous results. Some participants misinterpreted the required value, thinking it referred to the item's selling price rather than the total amount raised for the crowdfunding project, leading to their estimates being significantly low (mostly in the hundreds of euros). Conversely, a few participants provided estimates in the billions, which was entirely unrealistic for a single project. This discrepancy is likely due to a lack of careful attention when completing the survey or an overestimation by the participants.

6.2 Error Analysis

To gain a clearer understanding of the guesses and their averages, a simple table was constructed for each group. This table displays the actual amount raised for each project alongside the average of all guesses made by participants in each group (Forced, Temporal, and Dialectical). The total average guess was calculated by summing the first guess for each participant and dividing by the total number of participants, then repeating this process for the second guess. This means that for each guess, approximately 35 values per project were collected and averaged. Subsequently, the averages of the first and second guesses were combined to determine the overall average guess for each project. These average guesses were then compared to the actual values to calculate the error margins. The tables below detail these calculations and their outcomes.

Forced	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
True Value	8.668	34.909	79.081	63.028	205.574	9.558	3.485.826	70.040	5.942	2.868
Total Guess Average	21.438	47.766	52.296	102.591	152.453	15.181	1.742.703	165.276	36.694	22.526
Error	-12.770	-12.857	26.786	-39.563	53.121	-5.623	1.743.124	-95.236	-30.752	-19.658

Table 1 - Error Analysis Forced Technique

Temporal	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
True Value	8.668	34.909	79.081	63.028	205.574	9.558	3.485.826	70.040	5.942	2.868
Total Guess Average	9.137	26.390	35.892	67.064	143.572	8.602	2.035.215	94.539	13.757	6.262
Error	-469	8.519	43.190	-4.036	62.003	957	1.450.612	-24.499	-7.815	-3.394

Table 2 - Error Analysis Temporal Separation

Dialectical	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
True Value	8.668	34.909	79.081	63.028	205.574	9.558	3.485.826	70.040	5.942	2.868
Total Guess Average	23.534	36.026	46.623	101.690	117.272	16.092	1.206.543	67.014	23.884	14.738
Error	-14.866	-1.117	32.459	-38.662	88.303	-6.534	2.279.283	3.026	-17.942	-11.870

Table 3 - Error Analysis Dialectical Bootstrapping

As shown in the tables, each one follows the same format with the first header row listing each project numerically. True values are recorded for each project to facilitate comparison and provide a clear overview, as visible across all three tables. It is observable that error values may be either positive or negative; negative values indicate an overestimation compared to the true values they were intended to approximate, while positive values suggest an underestimation, indicating that the average guesses were lower than the actual figures.

Upon examining the tables, it is evident that the group with temporal separation achieved the best performance, exhibiting relatively minor errors close to the actual values. This is particularly noticeable in projects one, six, four, nine, and ten. Moreover, project seven had a significant im-

pact on the results and the magnitude of errors. Given its substantial funding relative to other projects, the error associated with project seven is markedly larger, reaching into millions of euros, whereas for other projects, it remains below one hundred thousand euros. This discrepancy is likely to affect the overall results due to its high variance. The forced group exhibited the highest error rates in several projects, notably one, two, four, eight, and nine. In contrast, the temporal group generally provided more accurate guesses, except for project three. The dialectical group, however, had the largest errors, particularly in project seven and also a significant error in project five. These results suggest that the temporal separation group is the most effective, followed by the forced technique group, and lastly, the dialectical bootstrapping group. However, further statistical analysis is required to examine these observations in depth.

6.3 Statistical Analysis

Using the results and variables from the previously analysed tables, further statistical analysis will be conducted. Initially, the analysis will focus on Mean Squared Error (MSE) values to evaluate whether the average guesses are more accurate than individual first and second guesses. Additionally, several tests, including one-sample t-tests, will be performed as part of the hypothesis testing. Moreover, a metric known as median gain, introduced by Stefan Herzog and Ralph Hertwig in their research, will be assessed across all groups. This metric is designed to determine whether averaging guesses results in an improvement or reduction in accuracy.

6.3.1 Mean Squared Error

To assess the accuracy of participants' predictions or estimates, we will use the Mean Squared Error (MSE) as a statistical indicator. In this study, MSE serves as a measure of the precision of participants' guesses compared to the actual fundraising amounts of crowdfunding projects. It quantifies the average of the squares of the errors—where an "error" represents the discrepancy between a participant's guess and the actual total collected by each project.

Technique	Forced	Temporal	Dialectical
First Guess	0,514	0,649	0,691
Second Guess	0,541	0,440	0,614
Average Guess	0,464	0,413	0,625

Table 4 - Mean Squared Error Values

It should be highlighted that the Mean Squared Error (MSE) values presented in the table, as well as in the figure above, are expressed in exponential notation (e12). This is primarily because the MSE values are relatively high, reflecting the significant differences between the actual values, as observed in the previous section.

When comparing the different techniques, it is evident that the first guess of the forced technique had the lowest MSE, suggesting that this group's initial estimations were closest to the actual values. However, the temporal separation technique showed marked improvement in accuracy with the second guess, and even more so when the guesses were averaged. This implies that the three-week interval between guesses was advantageous for participants, enhancing their judgment and positively affecting the average results.

By comparison, the dialectical bootstrapping technique consistently showed the highest MSE values across all three metrics, signifying that participants using this method were the furthest from the actual values. Additionally, the second guesses generally reveal an improvement in participants' judgment. This improvement suggests that the process of engaging with dialectical questions or experiencing a three-week delay may have aided participants in refining their estimates. In contrast, the forced technique showed no significant improvement on the second guess, potentially due to its methodology requiring participants to simply repeat their valuation of the project without additional intervention or time for reflection.

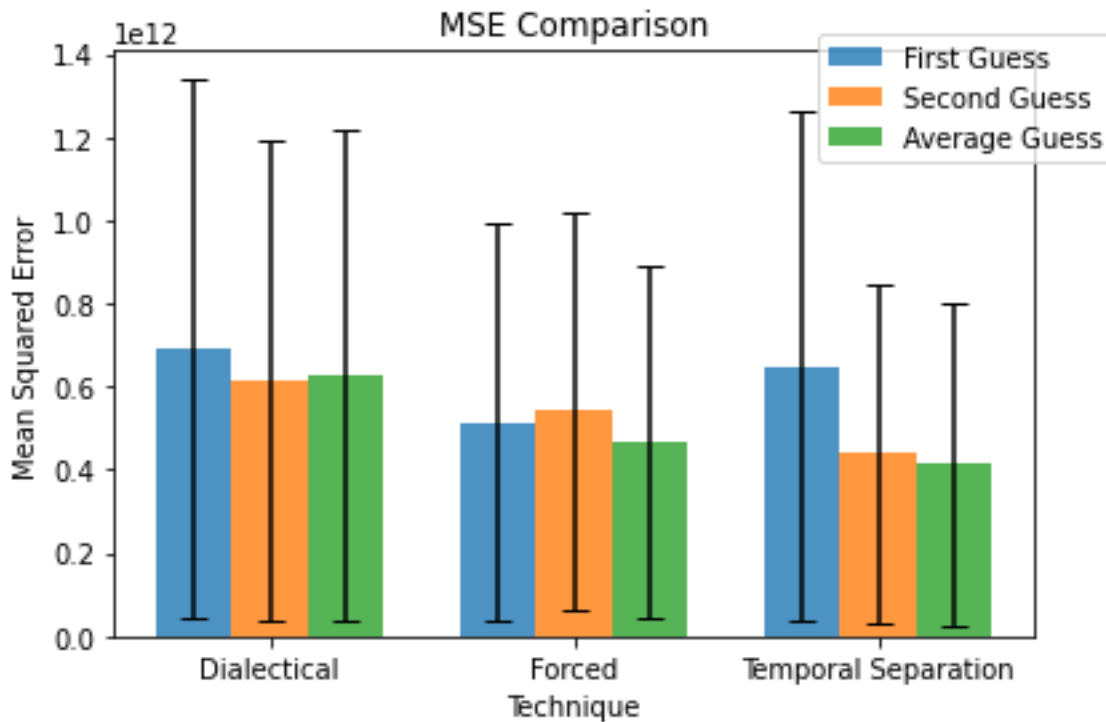


Figure 6 - MSE Comparison

Figure 6 above provides a visual representation of the overall performance of each guess, accompanied by error bars. These error bars indicate the standard error and reveal substantial variability in the guesses across all techniques and guess types, suggesting a lack of consistency in the participants' estimates and a high degree of variability in the results. The Temporal Separation group appears to have the most favourable performance, this is particularly noticeable in the second (MSE = 0,440) and average guesses (MSE = 0,413), when taking into account the error bars. Following the pattern of temporal separation, forced technique has also managed to show substantial improvement due to the process of averaging (MSE = 0,464). This was however not the case with dialectical bootstrapping technique (MSE = 0,625).

Moreover, certain projects, notably projects three, five, seven, and eight, exhibited larger error values than the rest. These considerable discrepancies from the original values undoubtedly contributed to the variability of the outcomes. Significantly, project seven's error values were in the

millions, which heavily contrasts with those of other projects and thus affects the error analysis. In addition, notable outliers in some participants' guesses could have had a substantial impact on the Mean Squared Error, these values are depicted in the figure above. These observations provide insights into the efficacy of each technique. Nonetheless, to establish statistical significance, further analysis using a one-sample t-test is necessary.

6.3.2 One Sample T-test

The mean squared error serves as a valuable indicator for analysis when testing whether the presented hypotheses are accepted or rejected. To conduct this analysis, a one-sample t-test was employed. Given that the experiment involved comparing a certain value to a known (true) value, the one-sample t-test was ideally suited as a statistical method. In this context, the one-sample t-test measures the difference between the participants' guesses (mean of the sample) and the true collected value of a project (known population mean), which is then divided by the standard error of the mean. Larger t-values suggest a greater difference between the two. Additionally, these t-values are accompanied by p-values, which also indicate whether these results are statistically significant or not. Like the previous tests, all these analyses were conducted using a Python script.

Technique	Guess Type	t-value	p-value
Forced	First	-0,71	0,310
	Second	-0,58	0,363
	Average	-0,84	0,351
Temporal Separation	First	-0,70	0,310
	Second	-0,82	0,311
	Average	-0,80	0,297
Dialectical Bootstrapping	First	-1,04	0,331
	Second	-0,82	0,339
	Average	-1,02	0,340

Table 5 - One Sample T-test

One common trait among all projects and their t-values is the presence of a negative sign as a prefix to the number, indicating negative values exclusively. This demonstrates that participants' guesses tend to be lower than the actual values when utilizing any of the given techniques. At first glance, it's apparent that the t-values for the Forced Technique, especially its second guess, exhibit the lowest figures of -0.51. This suggests that participants using the Forced Technique were closest to the actual values while also exhibiting the lowest data variability. Furthermore, on average, the Temporal Separation Technique has performed the best according to the presented t-values, with a first guess t-value of -0.71 and an average guess measuring at -0.80. Dialectical Bootstrapping as a technique has fared the worst out of the three. This is particularly noticeable in the first guess with a t-value of -1.04 and the average guess with a value of -1.02.

Reviewing the p-values presented in Table 5, it's apparent that none of the p-values fall within the significant range ($p < 0.05$). The Forced Technique group has managed to show lower t-values across all three guess types, but it has also produced one of the highest p-values, especially in the second guess ($p = 0.363$) and average guess ($p = 0.351$). This could be attributed to the fact that the observed differences are likely due to random variability rather than a systematic effect of the decision-making techniques themselves. Temporal Separation has outperformed the other groups, achieving the lowest p-values across all three guess types, with the average guess ($p = 0.297$) being the best performing guess type. Since none of the groups managed to reach the significant range, the hypothesis: *'The use of different decision-making techniques significantly enhances the accuracy and reliability of collective judgments in Kickstarter project evaluations'* must be rejected. This also indicates that all alternative sub-hypotheses (H1a, H1b, and H1c) must be rejected as well.

6.3.3 Median Gain

According to the research conducted by Stefan Herzog and Ralph Hertwig, the term "median gain" refers to the enhancement in the accuracy of an individual's judgments after incorporating a second, different estimate and then averaging it with the individual's initial estimate (Herzog & Hertwig, 2008). Following what they pioneered, a median gain was calculated for all participants across three groups. This approach allows for the assessment of how many individuals benefited from making a second estimate using various techniques.

Among the groups, the Forced group demonstrated significant improvement in its second estimate. To calculate the median gain as a percentage, the absolute error of the initial judgment is subtracted from the absolute error of the averaged judgment. This result is then divided by the absolute error of the initial judgment and multiplied by 100 to obtain a percentage. Participants with positive values shown notable improvements through the process of averaging, which is specific to their techniques. The graph below illustrates that 21 participants within the forced technique have benefited from this process, with 3 participants experiencing over a 20% improvement. Additionally, 12 participants exhibited a decline in median gain, suggesting that the averaging process sometimes reduced the accuracy of participants estimates.

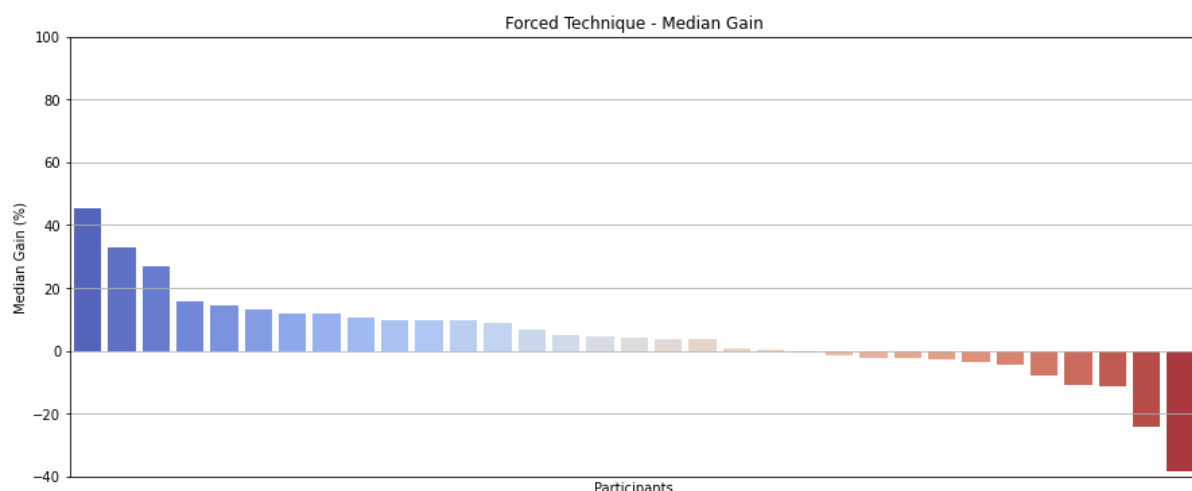


Figure 7.1 - Median Gain of Forced Technique

Building on the previous findings, the Temporal Separation group emerged as the top performer. This is further evidenced by the median gain, where 25 participants demonstrated significant improvement. As illustrated in the graph below, two participants achieved a median gain of over 60%, and nine participants (representing 25% of the entire group) achieved a median gain of more than 40%. The majority of the group members (70%) improved their results through the averaging process, confirming its potential to enhance guess outcomes.

However, a closer examination reveals that 10 participants experienced a negative median gain, suggesting a decline in the accuracy of their results, moving them further from the true value. It is noticeable that 7 out of these 10 participants saw a more substantial decrease than those in the Forced group, indicating that individuals in the Temporal Separation group were either very close to or significantly distant from the true value.

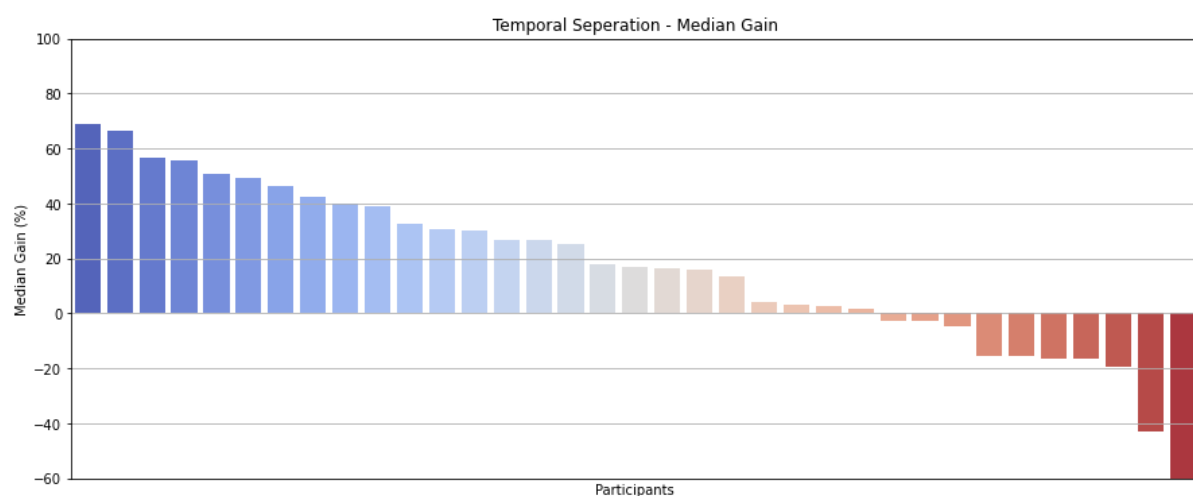


Figure 7.2 - Median Gain of Temporal Separation

The Dialectical Bootstrapping group exhibited some of the most notable results among the three groups. This group saw 19 participants improve their outcomes through the averaging process, each showing a positive median gain. Uniquely, this group also included 6 participants whose results neither improved nor worsened through this process, each registering a median gain of

0%. This indicates that for these individuals, the process of averaging and considering a second estimate was, on average, neither more nor less accurate than their initial guesses.

Additionally, 10 participants experienced a decline in their gains during the process. Notably, the participant with the highest median gain, approximately 78%, was from this group. Furthermore, within the context of negative scores, participants in the Dialectical Bootstrapping group performed relatively well; all 9 participants who observed a decrease in performance had a maximum median gain loss of only -10%.

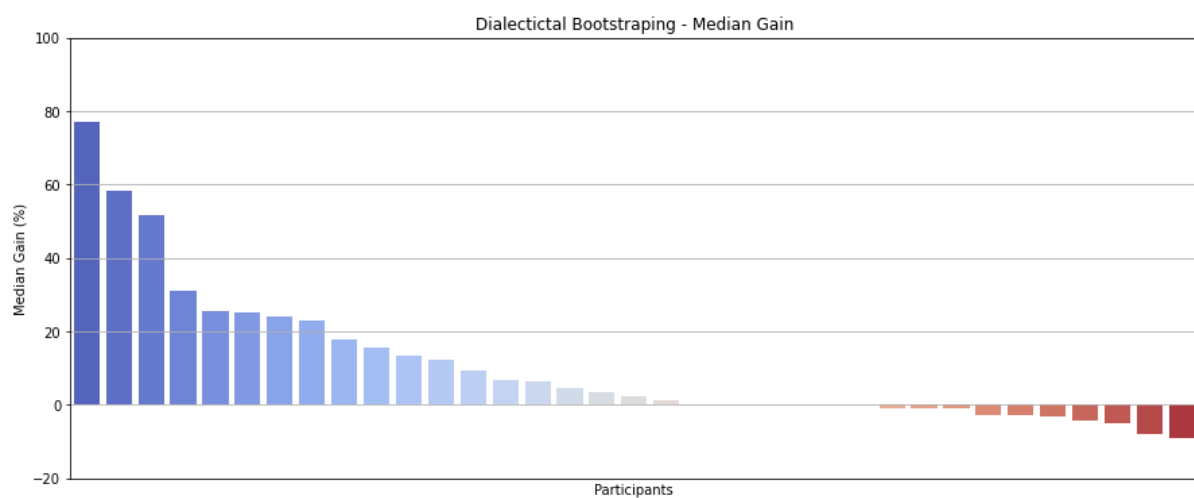


Figure 7.3 - Median Gain of Dialectical Bootstrapping

It is clear that for the majority of participants across all three groups, averaging the judgments has led to improvements. Although statistical analyses suggest that the results are not statistically significant due to significant differences in project valuations, the method still appears to be effective in refining judgments. This effectiveness is seen by the presence of at least one participant in each group, if not more, achieving over a 40% improvement, while the negative median gains remained relatively small across all groups.

7. Conclusion

The experiment conducted aimed to reveal the applicability of the wisdom of crowds in the context of crowdfunding, employing a variety of decision-making techniques. Specifically, three distinct methodologies were explored: Forced, Temporal Separation, and Dialectical Bootstrapping. These methods played a crucial role in assessing the reliability and accuracy of valuations for Kickstarter crowdfunding projects, while considering key insights from the past research (Vul & Pashler, 2008; Herzog & Hertwig, 2008).

Across all experimental groups, statistical analysis revealed that none of the methods achieved statistical significance, with p-values consistently exceeding the $p < 0.05$ threshold. However, the application of mean squared error as a statistical tool highlighting the potential for judgment averaging to enhance both accuracy and reliability. This effect was mostly present within the Forced and Temporal Separation groups, where averaged judgments demonstrated better performance compared to both first and second estimates provided by participants.

Further analysis focused on median gain, a metric designed to quantify the improvement in participants' judgments following the introduction and averaging of a second estimate. This analysis revealed that, across the board, averaged judgments tend to surpass initial participant judgments on an individual level, suggesting a broad applicability of judgment averaging across diverse decision-making techniques (Herzog & Hertwig, 2008; Greenberg et al., 2013).

An interesting aspect of the findings was the observed benefit of introducing a pause between judgments. This was reflected in the enhanced performance of the Temporal Separation technique across various statistical measures, including t-values and mean squared errors, pointing out the potential value of temporal separation in decision-making contexts (Herzog & Hertwig, 2008).

Contrary to initial expectations, the Forced technique, known for its simplicity, outperformed the Dialectical Bootstrapping method in all evaluated metrics. This outcome suggests that the straightforward approach of averaging two consecutive guesses, without the integration of more complex techniques, might offer a surprisingly effective means of capturing the wisdom of crowds in crowdfunding evaluations.

8. Limitations and Future Research

To ensure reliable results, a significantly larger sample size is crucial. In this experiment, 103 participants took part, which fell short of meeting the necessary criteria for the statistical tests applied. It is assumed that a sample size of 500 or more participants would significantly improve the study's reliability. According to research, the method of collecting and processing judgments critically influences the reliability of the outcomes (Stewart, T.R., 2001). It was noted that several participants struggled with processing judgments, with some offering estimates that were unrealistically low or high. Moreover, the values of the projects presented to participants varied widely, from \$2,000 to \$2,500,000, affecting the results. This variation led to high variability, complicating the achievement of statistical significance in hypothesis testing due to large standard errors.

Despite their lack of statistical significance, these findings offer insights into the potential application of the wisdom of crowd's concept in other fields, such as crowdfunding. Recognizing the limitations of this experiment and pursuing further research could allow for the exploration of accurately predicting the value of individual projects. This would be especially beneficial for investors, as it would enable them to identify projects likely to secure more funding and act as a valuable indicator for investment decisions.

9. References

- Fiechter, J.L., Kornell, N. How the wisdom of crowds, and of the crowd within, are affected by expertise. *Cogn. Research* 6, 5 (2021). <https://doi.org/10.1186/s41235-021-00273-6>
- Waldman, T. (1960). Rousseau on the General Will and the Legislator. *Political Studies*, 8(3), 221–230. <https://doi.org/10.1111/j.1467-9248.1960.tb01142.x>
- Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd rev. ed.). Princeton University Press.
- Surowiecki, J. (2005) *The Wisdom of Crowds*. New York: Anchor Books.
- Seifert, S. and Weinhardt, C. (2006) Group decision and negotiation (GDN) 2006 international conference, Karlsruhe, Germany, June 25 - 28, 2006; Proceedings. Karlsruhe: Univ.-Verl. Karlsruhe.
- Galton, F. (1907). Vox Populi. *Nature*, 75(7), 450-451
- Waddington, C.H. (1947). The scientific attitude. *Bulletin of the Atomic Scientists*, 3(2), 31-35
- Bhatt, S.P., Minnery, B.S., Nadella, S., Bullemer, B., Shalin, V.L., & Sheth, A. (2017). Enhancing crowd wisdom using measures of diversity computed from social media data. *Proceedings of the International Conference on Web Intelligence*.
- Boukouras, A., & Sugden, R. (2014). Deliberation, Learning, and the Wisdom of Crowds: An Experimental Study. *American Economic Journal: Microeconomics*, 6(3), 37-77.
- Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, 19(7), 645-647.

- Herzog, S.M. and Hertwig, R. (2008) ‘The wisdom of many within one mind: Making better judgments with dialectical bootstrapping’, *Psychological Science* , 20(2), pp. 231–237. doi:10.1037/e636742011-085.
- Page, S. E. (2007). Expertise, Diversity, and the Limits of Contemporary Wisdom. *The Journal of Economic Perspectives*, 21(4), 185-205.
- National Research Council. 2015. *Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance Potential for Military Accession*. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/19017>.
- Chen, Y., & Vaughan, J. W. (2016). Wisdom of Crowds with Restricted Participation. In *Proceedings of the 2016 ACM Conference on Economics and Computation* (pp. 405-422).
- Johnson, J., 2022. *Social networks and the wisdom of crowds*. [online] Network World. Available at: <<https://www.networkworld.com/article/2287661/social-networks-and-the-wisdom-of-crowds.html#:~:text=It%20refers%20to%20the%20fact,crowds%20can%20be%20self%2Dcorrecting.>> [Accessed 8 June 2022].
- Chen, Y. and J. Xie, 2008, “Online consumer review: Word-of-mouth as a new element of marketing communication mix,” *Management Science*, 54(3), 477-491.
- Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, Byoung-Hyoun Hwang, Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media, *The Review of Financial Studies*, Volume 27, Issue 5, May 2014, Pages 1367–1403
- Josh, S., 2022. *Shocking But True: 90% People Lose Money In Stocks – Research & Ranking – Blogs*. [online] Research and Ranking. Available at: <<https://blog.researchandraking.com/shocking-but-true-90-people-lose-money-in->

stocks/#:~:text=According%20to%20popular%20estimates%2C%20as,both%20new%20and%20seasoned%20investors.> [Accessed 8 June 2022].

Y. Kim and M. Kim, "'A Wisdom of Crowds': Social Media Mining for Soccer Match Analysis," in *IEEE Access*, vol. 7, pp. 52634-52639, 2019, doi: 10.1109/ACCESS.2019.2912009.

Mollick, E. (2013). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1), 1-16.

Belleflamme, P., Lambert, T. and Schwienbacher, A. (2018) 'Network effects in crowdfunding', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3259191.

Du, Q. (2019) Extracting the wisdom of crowds from crowdsourcing platforms. dissertation. Virginia Tech.

Polzin, F., Toxopeus, H. and Stam, E. (2017) 'The wisdom of the crowd in funding: Information heterogeneity and social networks of Crowdfunders', *Small Business Economics*, 50(2), pp. 251–273. doi:10.1007/s11187-016-9829-3.

Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *The American Economic Review*, 88(4), 848–881. <http://www.jstor.org/stable/117009>

Ariely, D., Au, W.T., Bender, R.H., Budescu, D.V., Dietz, C.B., Gu, H., et al. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied*, 6, 130–147.

Wallsten, T.S., Budescu, D.V., Erev, I., & Diederich, A. (1997). Evaluating and combining subjective-probability estimates. *Journal of Behavioural Decision Making*, 10, 243–268.

Kennedy, J. (1995). Debiasing the Curse of Knowledge in Audit Judgment. *The Accounting Review*, 70(2), 249–273. <http://www.jstor.org/stable/248305>

Hirt, E.R., & Markman, K.D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, 69, 1069–1086.

- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, 26, 1142–1150.
- Arkes, H.R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110, 486–498.
- Stewart, T.R. (2001). Improving reliability of judgmental forecasts. In J.S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 81–106). Norwell, MA: Kluwer Academic.

10. List of Figures

Figure 1 - MSE of all 3 conditions (Vul, E. & Pashler, H., 2008).....	14
Figure 2 - Dialectical Bootstrapping (Herzog & Hertwig, 2008).	17
Figure 3 - An expanded version of Brunswiks Lens Model (Stewart T.R., 2001)	24
Figure 4 - Experiment Design.....	33
Figure 5.1 - Participants by Age and Gender.....	38
Figure 5.2 - Participants by Educational Level.....	39
Figure 6 - MSE Comparison	44
Figure 7.1 - Median Gain of Forced Technique.....	47
Figure 7.2 - Median Gain of Temporal Separation.....	48
Figure 7.3 - Median Gain of Dialectical Bootstrapping.....	49

11. List of Tables

Table 1 - Error Analysis Forced Technique.....	41
Table 2 - Error Analysis Temporal Separation	41
Table 3 - Error Analysis Dialectical Bootstrapping.....	41
Table 4 - Mean Squared Error Values	43
Table 5 - One Sample T-test	45

12. Appendix – Survey Questions

Kickstarter Valuations

Introduction to the Kickstarter Funding Perception Survey

Welcome and thank you for participating in this insightful survey. Your contribution is vital as it helps me delve deeper into understanding how collective insights shape the perception of crowdfunding success.

Understanding Kickstarter

Kickstarter is a pioneering crowdfunding platform that transforms creative ideas into tangible realities, empowered by the support of a global community. It hosts a wide variety of projects ranging from innovative tech to groundbreaking art, with each project setting a clear funding goal and deadline. The platform operates on an all-or-nothing funding model, where projects must meet their financial targets within the set timeframe to receive the funds. In return, backers receive unique rewards or experiences, fostering a sense of shared accomplishment.

How Kickstarter Functions

1. **Initiation:** Creators launch their project on Kickstarter, establishing a funding target and a timeline.
2. **Community Support:** Backers pledge financial support in exchange for rewards or to champion the idea.
3. **Achieving Goals:** Projects that meet their funding goals by the deadline receive all pledged funds. Those that don't, don't receive any funds.
4. **Project Fulfillment:** Successful projects advance to the realization phase, often keeping their backers informed and involved.

Successfully Funded Projects

Category	Successfully Funded Projects	Less than \$1,000 Raised	\$1,000 to \$9,999 Raised	\$10,000 to \$19,999 Raised	\$20,000 to \$99,999 Raised	\$100 K to \$999,999 Raised	\$1 M Raised
All	249,927	32,128	132,230	36,160	38,097	10,528	784

Your Contribution to the Survey

In this survey, you'll explore 10 diverse Kickstarter projects. These vary in their objectives, amount raised, and funding status – some have achieved their goals, while others are in the process of fundraising. Keep in mind that **there is no limit to the funding** that a single project can raise.

Your Important Role

Your task is to assess each project based on the provided descriptions and estimate the amount of funding you think each has raised. This exercise isn't about right or wrong answers but about gauging collective judgment and perception in evaluating project success. Your input is a crucial piece of a larger puzzle, helping me explore how shared insights and perspectives can influence and enhance the understanding of crowdfunding dynamics.

I encourage you to take the time to consider each project thoughtfully and share your estimations based on your intuitive judgment. **Please base your judgment solely on the information provided and your intuition, without conducting any internet research or looking up additional details about the project.**

Thank you for your participation and for lending your perspective to this vital study.

Please continue to next page to answer a few general questions and afterwards give judgments on Kickstarter projects.

General questions

What is your age group? *

17 or younger

18 - 20

21 - 29

30 - 39

40 - 49

50 - 59

60 or older

What is your sex?*

Male

Female

Prefer not to say

Other:

What is the highest level of education you have completed?*

Primary or elementary school

Middle school or junior high

High school or equivalent

Vocational or technical training

Bachelor's degree (e.g., BA, BS)

Master's degree (e.g., MA, MS, MEd)

Doctorate (e.g., PhD, EdD)

Other

Bodyweight training device

The first full body workout device for multifunctional strength training anywhere, anytime. It is a **revolutionary fitness tool** uniquely combining traditional workout rings and parallettes, creating a multifunctional exercise solution for full-body training. This device stands out with its 7° ergo-

nomic angled handle, ensuring natural wrist alignment, vital for injury prevention and comfort. It also features a **quick-release buckle system**, allowing for swift transitions between various exercises, enhancing workout efficiency.

Handcrafted in Germany, this compact and versatile tool is designed for easy transport and use in any setting. Its **inside wood lock** reduces skin contact with the straps, offering enhanced comfort during use. Ideal for users at all fitness levels, from beginners to professionals, it's a comprehensive solution for those seeking a simple yet effective way to improve their strength and flexibility. The tool's adaptability makes it a perfect fit for a wide range of exercises, addressing the needs of today's dynamic fitness enthusiasts.

Example of the training device and its abilities



1. Bodyweight training device - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round

them to the nearest whole number, rely on both the information and your intuition for your judgment):*

Sharpest Micro Knife

Discover an innovative **micro knife** that rewrites the rules of portability and functionality. Crafted from **Grade 5 Titanium**, this tool is not only **waterproof and fireproof** but also nearly unbreakable. Its blade, a blend of **90% tungsten carbide** and other materials, achieves unmatched sharpness and durability.

Compact at only **5cm in length**, it can be easily attached to a keyring, slipped into a wallet, or clipped to a backpack. Unique from traditional designs, it features an **upside-down orientation** for immediate use without detachment from a keyring.

The blade's composition ensures it remains **sharp and effective** for longer, ranking high on the Mohs hardness scale. Available in four distinct designs, each style combines functionality with a modern aesthetic, ranging from intricate patterns to vibrant colors. This tool is a testament to efficient design, offering a blend of **strength and agility** in everyday tasks.

Example of the micro knife and its abilities



2. Micro Knife - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Lighting Device from your Wallet

This lighting device is **wireless-powered credit card-sized light designed** to revolutionize your smartphone and simplify your everyday lighting life. It effortlessly fits in your wallet or pocket, making it the most convenient light source for every situation. It is **easily attachable to magnetic phone cases** and **supports wireless charging** for cable-free experience. This device **delivers 5W of power** with superior color accuracy to enhance your photos and videos.

While smartphones have revolutionized photography, inadequate and inaccessible lighting solutions have held users back from realizing their full creative vision. This device removes the boundaries in portable lighting, providing unmatched quality and convenience in a device that fits right into your wallet.

Example of the lighting device and its abilities



3. Lighting Device - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Smart Ring

It is smartest and lightest wellness ring that tracks your vital signs such as your **heart rate and HRV**, monitors your **sleep quality**, counts your **daily steps and calories** burned, tracks

your **mindfulness** during a meditation session, keeps tabs on your **stress levels** and offers **personalized recommendations** for optimal and holistic wellness.

It gives **real-time insights** into your overall health and **receive personalized coaching** to improve your vital signs. This ring can track your **daily activity levels**, including **steps taken, calories burned, distance covered and so on**. Based on the quality of your sleep, the ring recommends personalized yoga & meditation sessions from the App, to ensure you feel refreshed and ready for the day.

Example of the smart ring and its abilities



4. Smart Ring - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Stand-alone milk steaming device

Elevate your beverage experience with this **advanced milk steaming device**, perfect for achieving cafe-quality results at home. With its state-of-the-art technology, it offers **precise control** over steaming, ensuring the smoothest and creamiest milk for any type of drink. This intuitive

device heats up in just 7 seconds and features a **revolutionary steam tip design** for precise and powerful steaming.

This versatile tool is compatible with all milk types and an array of coffee-making methods, from espresso to French press. Its **infrared laser technology** within the steam wand guarantees the perfect milk temperature every time, enabling you to create **professional-grade latte art** with ease. Compact and elegantly designed, it's a perfect combination of functionality and style, transforming every cup into a barista-level masterpiece.

Example of the milk steaming device and its abilities



5. Milk Steamer - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Gift Wrapping Tool

Experience a revolutionary approach to gift wrapping with this **innovative organizer**. Designed to streamline the gift wrapping process, it comes fully stocked with over \$45 worth of supplies, including **gift wrap cutters, rolls of tape, ribbons, scissors, bows, gift tags, and pens**. Each compartment in the organizer is thoughtfully arranged in the order you use the items, guiding you through the gift wrapping process step-by-step while keeping your supplies neat and organized.

This organizer not only makes gift wrapping fun and easy but also serves as a compact and beautiful container for all your wrapping needs. Whether placed under the tree or on the kitchen table, it adds aesthetic value to your space. When not in use, it's easily stored out of sight. With compartments specifically designed for each wrapping item and features like a quick access drawer, this organizer simplifies and enhances the entire gift wrapping experience.

Example of the gift wrapping tool and its abilities



6. Gift Wrapper - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Smart plant watering device

Revolutionize your indoor gardening with this advanced plant care system, an ideal solution for those who struggle with regular plant maintenance. With its **four smart soil sensors**, it precisely gauges soil moisture, ensuring your plants receive just the right amount of water. The system's intuitive **mobile app** provides real-time data and control, allowing for remote monitoring and adjustment of your plant's hydration needs.

This system also features **four autonomous water pumps** that meticulously manage water distribution, preventing both overwatering and underwatering. Complemented by an **extendable water tank** with a 1.3-gallon capacity, it's designed for long-term, worry-free plant care. Whether you're away on travel or simply busy, this system ensures your plants thrive with minimal effort on your part.

Example of watering device and its abilities



8. Plant Watering - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Application for life organization

Discover an all-encompassing **app** designed as a central hub for enhancing both personal and professional aspects of your life. It's more than just a digital tool; it's a **personalized platform** for nurturing passions, wellness, relationships, and career growth. This app stands out with its **customizable modules** that cater to a broad spectrum of needs, from **network management** and **goal setting** to **financial planning** and **self-reflection**. Additionally, it offers robust features for organizing daily routines, tracking hobbies, and monitoring wellness, making it a versatile tool for a balanced lifestyle.

This app distinguishes itself with a user-centric approach, providing a **distraction-free environment** to explore and manage life's essentials. Users can add main interests as 'Waffles,' delve into relevant 'Topping Categories,' and save them for **easy tracking and management**. Its **intelligent integration** with various apps and devices enhances its functionality and customization, making it a transformative tool for users to enrich every facet of their life in a private, effortless, and meaningful way.

Example of the app and its features



9. Life App - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole

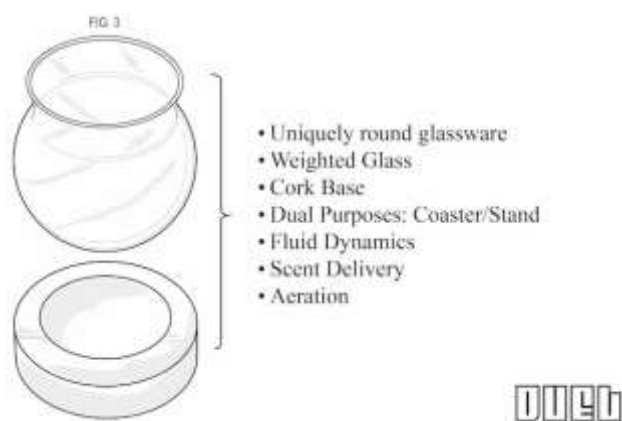
number, rely on both the information and your intuition for your judgment):

Whiskey Glasses Drink set

Introducing an exquisite **drinkware set** that transcends traditional design to elevate your drinking experience. This set is not merely functional but is crafted to create moments and enhance enjoyment with friends. The **Witches Cure Drink Set** is a fusion of design and practicality, intended to be a statement piece in any home. Each piece in this set is designed to be displayed proudly, like artwork enhancing your space.

Key features of this set include **hand-blown lead-free crystal glasses** with a unique round shape that allows for an enhanced swirling and nosing experience. This design amplifies the aromas of your elixir, creating a one-of-a-kind sensory experience. The glasses are complemented by an **iconic decanter**, continuing the theme of the perfectly round shape, complete with a cork base and stopper for an elegant display. This set is not just about serving drinks; it's about transforming the ritual of enjoying a drink into an art form. Each piece stands freely on a table, adding an organic and rustic charm to your drinkware collection.

Example of the set and its features



10 . Whiskey Set - Based on the provided data and image, please **estimate the total funds raised** by this Kickstarter project (assume that figures are in Euros (€) and round them to the nearest whole number, rely on both the information and your intuition for your judgment):

Dialectical Bootstrapping Questions:

Second Section - Alternative Judgment

Thank you for your thoughtful contributions in the first section of this questionnaire, where you provided your initial judgments on the Kickstarter projects.

As you progress to the second section, I invite you to engage in a reflective exercise, **re-evaluating your initial estimates** with a new perspective. Please consider the following questions as you revisit the same 10 projects:

1. **Assume your first estimate is off the mark. Why might this be the case? Reflect on this possibility.**
2. **Think about a few reasons why your initial estimate could have been inaccurate. Which assumptions and considerations might have been wrong?**
3. **What do these new considerations imply about your first estimate? Was it perhaps too high or too low?**
4. **Based on this new perspective, make a second, alternative estimate on the funding each project has raised.**

This approach is designed to deepen your understanding and explore different angles in assessing each project. Your ability to reconsider and provide an alternative viewpoint is crucial for a comprehensive analysis of the data.

When you are ready to offer these new insights, please proceed to the next page. Thank you!

All three questioners were in the same structure besides the fact that the temporal separation received two questioners, the initial one and another with a 3-week delay and the forced group was just asked the same questions again.

The second guess questions were made in the exact same way as the first.