ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics (ESE)

Master Thesis Behavioural Economics

E Pluribus Unum: An Analysis of Conditions for Wisdom of Crowds

Shania Sinha (648036)

Supervisor:	Prof. Dr. Peter Wakker
Second assessor:	Prof. Dr. Robert Dur
Date final version:	11 August, 2023

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Acknowledgements

I wish to extend my profound gratitude to Professor Dr. Peter Wakker for serving as a veritable beacon of enlightenment throughout the arduous process of crafting this thesis. His invaluable insights and guidance have been instrumental in elevating this work to its current caliber. Furthermore, I am deeply appreciative of the unwavering support of my parents and friends. Their encouragement gave me the strength to navigate this journey.

Abstract

The concept of ‘wisdom of crowds’ has captured the attention of researchers, policymakers, and practitioners alike across various fields over the last couple of decades. Despite abundant empirical evidence advocating the wonders of this phenomenon, the theoretical underpinnings of the ‘wisdom of crowds’ have not been analyzed closely. This thesis looks into the history of ‘wisdom of crowds’, what it means, and when can a given crowd be considered wise. Most researchers seem to hold the opinion that the greater the diversity and independence of individuals within the crowd, the higher the accuracy of the aggregate response. Is this really the case? The thesis examines the flaws in theorems such as ‘diversity trumps ability’ and ‘diversity prediction theorem’. Due to the lack of theoretical models substantiating the importance of independence for crowd wisdom, two empirical studies are examined that show conflicting evidence regarding the effect of social influence on the performance of the crowd.

Keywords: wisdom of crowds, collective decision, information aggregation

Contents

Introduction

The concept of ‘wisdom of crowds’ has captivated researchers, policymakers, and practitioners alike across various fields in the last couple of decades. Generally, this term encompasses situations where aggregating responses or decisions of a collection of individuals to some questions have yielded answers that were surprisingly close to the true answer. The term ‘wisdom of crowds’ was popularized by science journalist James Surowiecki in his 2005 book of the same name (The wisdom of crowds, 2005). Researchers oftentimes use the terms ‘collective intelligence’, ‘collective wisdom’ and ‘crowd wisdom’ to refer to the same phenomenon.

The first example of ‘Wisdom of Crowds’ as we know it today comes from Sir Francis Galton’s seminal 20th century study known as ‘*Vox Populi*’ (1907). He collected and analyzed responses from a country fair competition, in which participants were asked to guess the weight of a displayed ox once it had been “slaughtered and dressed”. Participants wrote down their guesses on numbered cards, along with their name and address. Given that there was a small fee to enter the competition, and a possibility to win a reward by the participant who provided the closest answer, Galton assumed that the responses were unbiased and were free from any possibility of “practical joking”. He also states that he believed the responses to be uninfluenced by any conversation among any participants. On creating a distribution of the responses, he was surprised to discover that the “middle-most” response was quite close to the true answer (the “middle-most” response indicates the median of the distribution). In a follow-up to the original article, Galton (1907) clarifies in an article titled “The ballot-box” that he advocates taking the median of the distribution, even though in this example, the mean of the distribution was closer to the true weight of the ox.

This example also marks the beginning of Surowiecki's journey to understand why collective decision-making fails in certain instances and what can be done to prevent a negative outcome. One must note the discrepancy between Galton's experiment and Surowiecki's description of it – Surowiecki states that Galton took the average of the participant responses, but he actually used the median of the responses. Taking the unweighted arithmetic mean is the most common method of aggregation today in studies focused on harnessing the wisdom of crowd effect (Lorenz et al., 2011).

Regardless of the aggregation technique used, Galton's analysis demonstrated the ability of a diverse group of individuals to generate a highly accurate estimation collectively – a revelation that would set the stage for future research on harnessing the wisdom of crowds. The publication of Surowiecki's book sparked a new interest in this idea: of wisdom or intelligence emerging from decisions made by a large group of independent individuals that have not been influenced by any communication or deliberations with another member of the group. Since then, empirical research in crowdsourcing (Surowiecki, 2005; Sunstein, 2006), prediction markets (Wolfers & Zitzewitz, 2004), and financial forecasting (Nofer & Hinz, 2015; Kelley & Tetlock, 2013) have shown that the aggregated judgements of many individuals are more accurate than the judgements of individual experts (Galton, 1907; Surowiecki, 2005; Page, 2008).

In this thesis, the definition of wisdom of crowds will be discussed first. Different authors define wisdom of crowds differently, but we build up to a more general definition. It will be followed by defining a 'crowd' and highlighting how it differs from a 'group' in this context. Having acquired a foundational comprehension of the phrase itself, the focus of the discussion shifts to exploring the factors that have been suspected to contribute to a crowd's wisdom. Many different factors have been suggested to lead to a crowd becoming wise, but every theorist and practitioner agrees that ‘diversity’ and ‘independence’ are key to a successful aggregation result, being a result that demonstrates the wisdom of crowds effect. Diversity and independence as predictors of the crowd’s success are then discussed. Finally, we explore why this phenomenon is so popular, and whether it should be given any significance as a judgement aggregation technique.

Defining wisdom of crowds

Following the publication of Surowiecki’s book, researchers started to explore the wisdom of crowds effect but with different conditions for when a crowd is considered to be wise. Suroweicki (2005) wrote that wisdom of a crowd emerges when a group of diverse, independent individuals make separate decisions, and the aggregate of all the decisions is better than or more accurate than what could have been achieved by any individual within the group. He also stated that the aggregate decision of the individuals within this group may also be better than some experts. It is unclear who Surowiecki refers to as experts, but an expert may be either a high-performing individual within the crowd, or an external entity known to possess domain knowledge in the topic to which the questions belong.

Another description of wisdom of crowds given by Nofer & Hinz (2015) states that “a diverse and independent crowd can make more precise predictions than a few people, even when only professionals are involved”. This definition suggests that the size of the crowd is also an important determinant of a surprisingly close-to-accurate aggregated result. In the case that the only individuals that can form a crowd are “professionals” or experts, there is bound to be less diversity than if the crowd consisted of laypeople as well (Hong & Page, 1998). Nofer & Hinz suggest that despite the decrease in diversity, aggregating the decisions of a large crowd of experts will yield better results than aggregating the decisions of a small one.

Hong and Page (2012) represent the crowd's wisdom in terms of the squared error of the crowd's prediction. They present a proof that squared error of the crowd will always be less than or equal to the average squared error (which is the average of the squared errors of all individuals in a crowd).

Since most descriptions of wisdom of crowds focus on comparing the crowd's accuracy to that of the average individual member (Larrick, Mannes, & Soll, 2012), we define the wisdom of crowds effect to be the phenomenon where the aggregated knowledge of a crowd comprising diverse individuals who are independent (no interactions), is better than the average knowledge of the crowd.

What is a crowd?

Group-decision making has been a subject of interest for a long time. An early advocate of group decision making was Aristotle (circa 300 BCE), remarking in his book *Politics*:

> *…the many, who are not as individuals excellent men, nevertheless can, when they have come together, <u>be better than the few best people, not individually but collectively</u>, just as feasts to which many contribute are better than feasts provided at one person's expense. (Politics, Book 3, Chapter 11)*

In this setting, Aristotle refers to the decision taken by many minds as being superior to that taken by an individual due to pooling of information and deliberation. When talking about wisdom of crowds however, the ‘groups’ and ‘crowds’ are defined differently.

Andler (2012) frames the connectedness of a collection of individuals as a spectrum. On one extreme, the collection may be *thickly collective*, meaning that there is a flow of information among its members. In pooling all their intellectual resources, participants in such a collection may exchange ideas freely and change their minds, for instance going from having no opinion to having a firm and clear opinion. Such a collection is also subject to failures such as groupthink (Janis, 2008), social loafing (Simms & Nichols, 2014), group polarization (Myers & Lamm, 1976), escalation of commitment (Staw, 1981) and many others. In fact, the idea that individuals may fall prey to bad popular opinions within a group has been discussed for a long time, for example in Charles Mackay’s popular 1841 book ‘Extraordinary Popular Delusions and the Madness of Crowds’ (Mackay, 2012).

Surowiecki’s book plays off the title of Mackay’s book (Surowiecki, 2005) and aims to uncover when the crowd may be considered wise, and what causes it to succumb to ‘madness’ as Mackay said. A collection of individuals within which each individual simply provides her own conclusions, without involving any reasoned discussion with the other members of the collection, is a *thinly collective* one, referring to the other extreme of Andler’s spectrum of collections (Andler, 2012). An extreme case of a *thin collective* is one where the individuals comprising the collective do not communicate with each other in any capacity. If all individuals in such a collective are asked the same question, their responses can be assumed to be completely independent of each other. The ‘crowd’ Surowiecki refers to this extreme case of a *thin collective*. This ‘crowd’ is one that is not bound by a common thread, and individuals within such a collective provide responses in a decentralized manner. This is also the ‘crowd’ that researchers generally refer to in papers on or related to wisdom of crowds.

Can individuals in a crowd really be completely independent in a real-world setting? The popular idea of ‘six degrees of separation’ makes a case against true independence between any two individuals in the world, stating that there are, on average, at most six individuals between oneself and a stranger. Many studies over the years have also found evidence for this claim (Travers & Milgram, 1969; Watts et al., 2002). This means that although there may be no deliberation in a crowd, the individuals in the same crowd are still in communication and exchange ideas with each other through deliberation at a larger scale. Then what is the basis of wisdom of crowd? Landemore (2012) claims that this is not a valid concern since the emergence of wisdom of crowd occurs only in small clusters of participants, not the complete population. This smaller collection of individuals (crowd) also focuses on specific subtopics or elements rather than all the knowledge in the world. Although the definition of a ‘crowd’ does not say anything about any limits on the size of the crowd, one can expect the ‘crowd’ considered in an empirical study to be a small subset of the human population, within which the constraints may still apply.

In the following parts of this thesis, I shall use ‘crowd’ or ‘collective’ to refer to a collection of individuals who provide responses to the questions asked without exchanging any information with each other. Their responses are ideally completely independent and will be aggregated to measure the knowledge of the ‘crowd’. The term ‘group’ will indicate a collection of individuals in which there is exchange of information, and the individuals within it must communicate to a reach a consensus. This consensus will represent the level of knowledge the ‘group’ achieves as a whole.

When does wisdom of crowds emerge?

A crowd may not always be considered wise. In his analysis of historical cases where the crowd acted wisely and made the right decision, Surowiecki (2005) uncovered some factors that he claims were the drivers of the crowd's wisdom:

- **Diversity** : Each individual within the crowd must have private information or their own eccentric representation of known facts.
- **Independence** : Opinions of any individual in the crowd must not be influenced by the opinion of any other member of the crowd.
- **Decentralization** : There is no communication among the individuals of the crowd. Each individual makes their own conclusion.

The decentralization factor is already taken into account when considering the 'crowd' as defined above. The intuition behind having individuals with diverse opinions in the crowd is that the errors of individual judgements will either be uncorrelated or negatively correlated (Hong & Page, 1998; Page, 2008). When taking the aggregate, the errors cancel out, resulting in accurate and wise crowd decisions (Hong & Page, 1998; Page, 2008).

Hong & Page (2012) offer another condition that makes a crowd wise – each individual in the crowd must have the ability to provide answers such that their error rate is at least better than random. This means that participants in a wise crowd at least have some idea of the topics at hand. It is interesting to note that the conditions of diversity and 'better than random' prediction are also the conditions that are required to ensure that the aggregation of weak models in ensemble methods provide high prediction accuracy (Dietterich, 2000). One may even state that ensemble learning is the machine learning representation of wisdom of crowds (Sagi & Rokach, 2018). However, this discussion is out of the scope of this thesis.

In economics and finance, many researchers focus on satisfying the diversity and independence conditions in order to harness wisdom of crowds in empirical studies. Hence the following sections will discuss the meaning and importance of these factors.

Diversity as a factor for wisdom of crowds

Diversity has become an important goal to achieve within workplaces, within organizations and even within nations. Diverse crowds comprising individuals with varying backgrounds, perspectives and knowledge bring a wide range of opinions to the collective decision-making process. But what does diversity mean? Is diversity across ethnicity, race, or gender sufficient to elicit the wisdom of crowd? Scott Page, a diversity researcher, writes about this in his 1998 paper along with Lu Hong (1998):

> *What do we mean by diversity? Do we then mean race, profession, gender, or ideology? We mean all of these and yet none of these. Zenisms aside, to us diversity means differences in problem solvers' perspectives and heuristics–variations in how people encode and search for solutions to problems.*

No discussion about diversity within crowds would seem to be complete without mention of Hong and Page (Landemore, 2012; Andler, 2012). They are frequently cited researchers relating importance of diversity in contributing to a wise crowd. Over the years they have published numerous papers and books discussing the importance of diversity in crowds. This pair of researchers have also attempted to provide theoretical models for collective wisdom.

Hong and Page (2004) provided proof that a group of 'diverse problem solvers' can outperform a group of 'high-ability problem solvers'. The authors offer results from a computational experiment and a mathematical model to compare the performance of diverse 'problem-solvers' and the 'best-performing problem-solvers'. Through the result of the simulations, they establish that the 'diverse group' consistently achieves performance superior to the 'high-ability group', which they term the "Diversity trumps Ability" theorem. The mathematical and computational models that Hong and Page described, however, have been criticized by mathematicians (Thompson, 2014; Houlou-Garcia, 2017; Hedtke, 2014).

To form the group of 'diverse problem-solvers', the authors simply select agents (also called problem-solvers) randomly from a pool. This unfortunately does not guarantee that the selected collective of agents is truly diverse. If the population itself consists of a majority of agents that are homogeneous, then random selection of some number of agents does not guarantee a truly diverse crowd, neither does it line up with the definition of diversity Hong and Page provided in their 1988 study (Hong & Page, 1998; Houlou-Garcia, 2017). The authors confuse the correlation between the composition of the 'diverse group' and its success as causation, when in reality there is no evidence to show that the population from which these agents have been sampled is diverse to begin with.

The way Hong and Page select the best-performing ('high-ability') agent may also cause concern. They define the best-performing agent by considering the quality of their heuristic and taking the average of all the scores obtained by the agent. Houlou-Garcia (2017) explains the problem here by taking the example of two agents who have been asked to provide five potential solutions. Agent A achieves scores 10%, 15%, 5%, 100%, and 8%, meaning the first solution solves 10% of the problem, the second solution solves 15% of the problem, and so on. The average of Agent A's score is 27.6%. Agent B achieves scores 70%, 65%, 75%, 80%, and 85%. The average of Agent B's score is 75%. By the logic described in the study, Agent B is the best-performing agent. Houlou-Garcia suggests that the performance of an agent should not be determined by the average of the scores, but with maximum score of all solutions evaluated. Between A and B, A succeeded in finding the optimal solution, solving 100% of the problem. This means A is the true 'high-ability' agent. Even this, however, does not seem like a solution that provides a true measure of expertise. In this simplistic case, a more complete solution might be to look at the average accuracy of the agents, but also the variability in their responses. I propose to use the coefficient of variation (COV) measure to find the best-performing agent in this case. For Agent A, the average score is 27.6% and the standard deviation of the scores is 37.2%. This means that the COV of Agent A is 1.348 or 134.8%. Similarly, the average score of Agent B is 75% and the standard deviation of the scores is 7.07%. The COV of Agent B is 0.0943 or 9.43%, making Agent B the better pick for a high-performing agent. We can also consider another agent's scores, Agent C achieves scores 5%, 7%, 6%, 10%, 8%. The mean of Agent C's scores is 7.2% and standard deviation is 1.92%, making the COV of Agent C 0.2667 of 26.67%. We can see that Agent B still outperforms Agent C.

In the context of this paper by Hing and Page, Thompson (2014) writes that the better performance of the 'diverse group' compared to the 'high-ability group' may be credited to the use of randomization rather than any true notion of diversity among the agents. The idea that randomization can improve algorithms has also been well proven, and that may be the explanation for the results Hong and Page obtained (Thompson, 2014). Thompson (2014) conducted another simulation where they compared the success of a randomly-selected collection of agents to that of a maximally diverse collection of agents, using the same definition of 'maximum diversity' between any two agents as stated by the Hong and Page (2004). In all cases, Thompson (2014) found that the 'maximally diverse' collective performed worse than the median performance of the randomly grouped agents.

In an article called 'Some Microfoundations of Collective Wisdom' (Hong & Page, 2012), Hong and Page attempt to provide statistical models to prove that diversity is an important condition to achieve collective wisdom. This time around the authors also claim that every individual in a crowd or collective must be knowledgeable enough to have an error rate that is at least better than random. They present the 'Diversity Prediction Theorem':

$$SqE(c) = SqE(\vec{s}\,) - PDiv(\vec{s}\,)$$

Consider a crowd of size n, $SqE(c)$ the squared error of the collective (aggregate of the individuals), $SqE(\vec{s}\,)$ is the average squared error of all the individuals in the crowd, and $PDiv(\vec{s}\,)$ is the predictive diversity of a vector of signals $\vec{s} = (s_1, s_2, s_3, \ldots, s_n)$, indicated by the variance of the signals generated by the individuals in the crowd. The collective prediction c is defined as the average of all the generated signals (responses) of the individuals in the crowd. Although seemingly novel, it is actually a well-known result known as the König-Huygens formula, dating back to the seventeenth century (Houlou-Garcia, 2017).

Presented in the above format, the reader is immediately led to believe that the collective squared error decreases as predictive diversity increases ceteris paribus, proving the claim that the more diverse the crowd is, the smarter or wiser the crowd becomes, compared to the average individual within the crowd. However, this interpretation would only hold if predictive diversity and average squared error were independent. In the way these terms are defined by the authors, average squared error $SqE(\vec{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - \theta)^2$ and predictive diversity $PDiv(\vec{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - c)^2$ are dependent on the same variable: the generated signals (responses) of all the individuals. In this case, it is possible that the predictive diversity increases and the average squared error also increases, leading to an increase in the squared error of the collective. This disproves the Hong and Page's claims that increase in diversity lowers the collective squared error (Hong & Page, 2012).

These examples of misuse of mathematics are particularly disheartening, as diversity is an important topic and has significant impact in informing decision making. Incorrect or incomplete proofs being touted as mathematical truths has a significant negative impact on the field. Many researchers have cited Hong and Page's mathematical work to provide a mathematical basis to their own research in the field of diversity (Andler, 2012; Landemore, 2012; Görzen & Laux, 2019), without truly understanding whether the evidence really proved what Hong and Page claimed with their models and theorems.

Independence as a factor for wisdom of crowds

The success of wisdom of crowds also hinges on the independence of the individuals comprising the crowd (Surowiecki, 2005; Lorenz et al., 2011). Independence decreases when individuals are influenced by the opinions or decisions of their neighbors or other members of the crowd. True independence is unlikely to be achieved in a real-world setting since individuals are interconnected with each other in one way or another through extensive social networks.

Lorenz et al. (2011) tested the effects of social influence on the performance of the crowd on estimation tasks that tested real-world knowledge regarding geographical facts and crime statistics. Twelve experimental sessions were carried out, each consisting of 12 participants. The experimenters deliberately chose questions for which the individual participants were unlikely to know the exact answers but were not completely unknowledgeable about it either. Participants were then split into three groups. In the first treatment group, the "aggregated information" condition, participants could base their answers to all questions except the first on the average of all 12 estimates from the previous round. In the second treatment group, the "full information" condition, participants received all 12 separate estimates from the previous round. In the control group, the "no information" condition, participants did not receive any estimates from the past rounds. It is unclear what estimates the very first 12-participant round received, since there are no disclosed historical estimates that can be provided. The authors found that despite providing bare information of the estimates of others, the aggregated estimate of participants in each round deviated significantly from the true answer. The unweighted arithmetic mean of the estimates of participants was closer to the truth than the individual first estimates in only 21.3% of the cases. The social influence seemingly eroded the independence among participants and lead to lesser diversity in the total responses.

Meanwhile, another study by Becker et al. (2017) showed that in decentralized networks, social influence can actually improve the accuracy of the collective, even if individual beliefs become similar. In a centralized network, however, where a few individuals have strong influences, collective estimates are more likely to be incorrect. The authors define a decentralized network as one where every individual has the same number of connections. For example, in a decentralized network of 10 individuals, each individual may be connected to exactly 2 other members of the network. The authors observed that in the decentralized network, the social influence dramatically decreased the diversity in responses, although individual accuracy increased in each consecutive round. The average error of each group's median estimate also went down in the last round of a single experimental session. The results of the study effectively challenge the notion that social influence undermines the wisdom of crowds.

Conclusion

In the preceding sections, we explored the definition of wisdom of crowds and discussed factors believed to contribute to it. Although Surowiecki and other researchers assert the significance of diversity and independence for crowd wisdom, the analysis of these factors reveals a less conclusive foundation than is expected. Mathematical proofs that were used to give diversity research, in the context of wisdom of crowds, a sense of scientific underpinning have been proven to be incorrectly interpreted to the point of being false (Houlou-Garcia, 2017). Similarly, the empirical studies conducted to uncover the effects of social influence on the accuracy of the crowd presents conflicting results. In one case, minimal cue from previous participants significantly decreased crowd accuracy (Lorenz et al., 2017), while in the other case social influence actually significantly increased individual and hence crowd accuracy (Becker et al., 2017). Thus, the effect and importance of diversity and independence as the forces driving crowd wisdom are still unclear.

But the question must be asked – is chasing after the wisdom of crowd a considerable tradeoff for chasing the experts? The original appeal of the phenomenon was that a collection of unassuming individuals may also reveal surprisingly accurate predictions at times. While the wisdom of crowds effect may be easily achievable for computer agents using ensemble learning methods in machine learning, human decision-making and judgement elicitation are too complex for simple statistical aggregations to provide a 'good' response at all times.

Despite the lack of strong evidence for determinants of wisdom of crowds, the appeal of this method of judgement aggregation is understandable. The wisdom of crowds phenomenon allows the researcher to spend less effort in identifying expertise to get surprisingly good solutions to questions posed within experiments or polls, given that the questions have definitive answers. The success of the wisdom of crowds method may also rely on the nature of the question being asked. As Hackman & Morris (1975) suggest, there may be no single theory that can encompass and deal with the complexity of factors that affect group task effectiveness. Given this consideration, it becomes beneficial to explore the interplay between the technique to achieve the wisdom of crowds effect and the nature of the problem posed.

References

Andler, D. (2012). What Has Collective Wisdom to Do with Wisdom? In H. Landemore, & J. Elster, *Collective Wisdom: Principles and Mechanisms* (pp. 72-94). Cambridge: Cambridge University Press.

Aristotle, A., Jowett, B., & Davis H. W., C. (1920). *Aristotle's Politics*. At the Clarendon Press.

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the national academy of sciences*, *114*(26), E5070-E5076.

Dietterich, T. G. (2000, June). Ensemble methods in machine learning. *In International workshop on multiple classifier systems* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.

Galton, F. (1907). The ballot-box. *Nature*, *75*(1952), 509-510.

Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 450-451.

Görzen, T., & Laux, F. (2019). *Extracting the wisdom from the crowd: A comparison of approaches to aggregating collective intelligence.*

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in experimental social psychology*, 8, 45-99.

Hedtke, J. (2014, December). *Mathematicians refute oft-cited 'diversity trumps ability' study*. From https://www.thecollegefix.com/mathematicians-refute-oft-cited-diversity-trumps-ability-study/

Hong, L., & Page, S. (1998). *Diversity and Optimality*. Santa Fe Institute.

Hong, L., & Page, S. (2012). Some Microfoundations of Collective Wisdom. In H.

Landemore, & J. Elster, *Collective Wisdom: Principles and Mechanisms* (pp. 56-71). Cambridge: Cambridge University Press.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385-16389.

Houlou-Garcia, A. (2017). Collective wisdom, diversity and misuse of mathematics. *Revue francaise de science politique*, *67*(5), 899-917.

Janis, I. L. (2008). Groupthink. *IEEE Engineering Management Review*, *36*(1), 36.

Kelley, E., & Tetlock, P. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, *68*(3), 1229-1265.

Landemore, H. (2012). Collective Wisdom: Old and New. In H. Landemore, & J. Elster, *Collective Wisdom: Principles and Mechanisms* (pp. 1-20). Cambridge: Cambridge University Press.

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger, *Social Judgment and Decision Making* (pp. 227-242). Psychology Press.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, *108*(22), 9020-9025.

Mackay, C. (2012). *Extraordinary popular delusions and the madness of crowds*. Simon and Schuster.

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological bulletin*, *83*(4), 602.

Nofer, M., & Hinz, O. (2015). Are Crowds on the Internet Wiser than Experts?–The Case of a Stock Prediction Community. In M. Nofer, *The Value of Social Media for*

Predicting Stock Returns: Preconditions, Instruments and Performance Analysis, (pp. 27-61).

Page, S. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4).

Simms, A., & Nichols, T. (2014). Social loafing: A review of the literature. *Journal of Management Policy and Practice*, *15*(1), 58.

Staw, B. M. (1981). The escalation of commitment to a course of action. *Academy of management Review*, *6*(4), 577-587.

Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford University Press.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Thompson, A. (2014). Does diversity trump ability? *Notices of the AMS*, *61*(9), 1024-1030.

Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, *32*(4), 425-433.

Watts, D. J., Dodds, P. S., & Newman, M. E. (2002). Identity and search in social networks. *Science*, *296*(5571), 1302-1305.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of economic perspectives*, *18*(2), 107-126.