University of Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Industrial Economics/ Project management	Spring / ~~Autumn~~ semester, 2020 Open / ~~Confidential~~
Author: Josefine Rygh	
Programme coordinator: Supervisor(s): Sigbjørn Landazuri Tveteraas	
Title of master's thesis: Wisdom of Crowds: A Literature Review	
Credits: 30	
Keywords: Wisdom of crowds Forecasting Predictions Collective intelligence Aggregation	Number of pages: 37 + supplemental material/other: 10 Stavanger, 14/07/2020 date/year

Title page for master's thesis
Faculty of Science and Technology

Abstract

The wisdom of crowds is an idea that could be a valuable resource if it is used, and even more valuable if it us used in the best way. There are many possibilities, and this thesis' purpose is to find out if the wisdom of crowds is beneficial for getting more accurate crowd predictions, and how the aggregation and information method, crowds' diversity, expertise and size play a role on the prediction accuracy. To get an overview of the empirical literature on the subject a traditional literature review is conducted, and the empirical literature is found using *Google Scholar* and the search words *wisdom of crowds*, *crowd predictions* and *collective intelligence* is used. There are 27 papers that is found and used for the literature review.

The results of the literature review show that there are several great ways to aggregate the individuals' predictions, and the crowds are more accurate than most of the individuals in them. Distributing information to the individuals will make the crowd less diverse but as long as the reduction in diversity is not too large, the added information can still manage to make the crowd more accurate. Expertise does not give a significant more accurate crowd prediction; it is more important to have a diverse crowd. The size of the crowd can be small when added factors are used for aggregating, like accentuating the individuals who previously predicted good, but for the arithmetic average it is better with larger crowds.

Acknowledgements

When COVID-19 entered the country and sat everything on hold, my original master thesis could no longer be executed. I had to change my line of thought and start to work on a new thesis. I did want to keep close to my original thesis, and therefore my new thesis became a literary review of the subject of *wisdom of crowds*.

It was my supervisor Sigbjørn Landazuri Tveteraas who introduced me to the subject of wisdom of crowds. I had little to no knowledge about it when he first introduced it to me, but soon understood how interesting this would be to look further into. He also supplied me with alternatives when I had to change my thesis. I am grateful for all the help and guidance he has given me.

The work with this thesis has been hard because I do not have a lot of experience in writing big reports like this. It has been tedious, stressful, and tough work. I have not always been a joy to be around, and my family have had to deal with several different moods.

I want to give a huge thanks to my parents for being there for me, serving me food, lending me the car, and supporting me through everything. I will be forever grateful.

Thank you to my co-workers who have given me a little escape each week from my thesis.

There have been plenty of times where I have thought that I would never be able to make it, but I finally made it through.

Hommersåk, 14 July 2020

Josefine Rygh

Josefine Rygh

Table of contents

List of tables

List of figures

Glossary

Asymmetric information – An unequal amount of information.

Iterative communication – The individuals receives a prediction that another individual has made. They can change it or let it be, and then it goes to another individual.

Kemeny-Young – For predictions that consist of a combination of predictions. Finds the predictions that is closest to all other predictions based on measuring the distance between the ordering of items. This is the crowd’s prediction.

The Borda count – For predictions that consist of a combination of predictions. The items are given weights according to which placement they have in the ranking. The highest ranked item gets the most points. All weights are added, and the items are sorted from highest total score to lowest. This is the crowd’s prediction.

The “Greedy count” – For predictions that consist of a combination of predictions. Sums up the number of times an item is placed in a position. The items are ordered after the amounts of times they occur in a certain position. This is the crowd’s prediction.

1 Introduction

Back in 1906 when Francis Galton went to the farmers market, he found that 787 people were able to predict the weight of an ox better than cattle experts. The 787 people gave individual predictions about what the ox's weight would be, and Galton took their predictions and averaged them. The average prediction from the crowd turned out to be 0,5 grams from the actual weight of the ox, which no individual at the fair had been able to predict (Galton, 1907).

Galton's findings led to an interesting phenomenon, which would change our way of working if his findings could be proven. Many have studied this further trying to figure out if it was luck or legitimate. There is no doubt that a crowd can be wiser than the individuals in the crowd, which will be apparent later in this thesis, but to outperform every individual is a greater task. But if you were to choose one person from a crowd, and you were to trust this individual 100%, you risk choosing one of the individuals that is worse than the crowd's prediction.

A lot of industries consist of predictions in some form. In 2015 KPMG researched the global building and construction industry and found that only 1/3 of projects are finished within the cost- and time budget (KPMG, 2015). By using predictions about cost and time from several individuals who made individual predictions for budgets, they could possibly reduce how much they go over budget, or even go under budget. This could be used for every industry struggling with going over the budget.

It could also be essential in industries where they have to prepare for a number of customers. It is difficult to know exactly how many people who will eat at a restaurant, visit an amusement park, or buy a specific product in a store in one day, unless every table, ticket and product is pre-ordered. This is even harder for longer periods like months. For companies that have been open for some time they have data available for previous days, weeks, and months, which is useful information when they prepare. If they used the wisdom of crowds, they could get predictions which could be truer than a single manager could come up with alone.

Of course, there is cost to take into consideration. By going through with the method of wisdom of crowd, they would need to engage more people into a decision were they previously only use one or a couple. It would be difficult to convince someone to take on those extra cost for something that they do not know will work 100%. Since there is a lot of studies and literature on this topic, a review like this will help accentuate what works and what method they could use to try to implement the wisdom of crowds in their decision making and if it will be worth it. It will also give other researchers an overview of what have already been studied on and give them ideas for what need to be researched.

1.1 Problem statement and research questions

The text above is what have given the reasoning behind the problem statement, as well as the following research questions. The aim of this thesis is to give an overview of some of the studies that have studied the wisdom of crowds. The thesis will look at how the studies that is found will show wisdom of crowds, and if they follow the theory that is stated to be essential. It will also look at how to best aggregate the crowd and if it is possible to get a wiser crowd with giving the individuals information. The problem statement is as follows:

To what degree do empirical studies find that the wisdom of crowds leads to more accurate predictions?

With the research questions:

- In these empirical studies, what is the role of
  - Aggregation method
  - Information
  - Expertise and diversity
  - Crowd size

  on the crowds' prediction accuracy?

The limitations that have set boundaries for the research questions is the fact that this is a master thesis, and there is a set amount of time available to work on it. The time to find and analyse literature about an unknown topic was limited.

1.2 The thesis' disposition

The thesis' disposition is as follows:

2 Theory	Enlightens the topic of wisdom of crowd and helps with understanding more of the findings that is to come
3 Methodology	Explains the method used during the work of this thesis to find the data which gives the basis for the analysis
4 Findings	Goes through the different findings, finds similarities, and differences and presents them in a systematic manner

5 Discussion	The findings are discussed and compared to the theory that have previously been presented
6 Conclusion	Concludes the thesis and gives suggestions for further research

2 Theory

2.1 What is the wisdom of crowds?

To explain the wisdom of crowds we have to picture a group of different individuals. Every individual has their own way of thinking and have different kinds and amounts of knowledge. The theory behind the wisdom of crowd is that when these individuals are collected in a group, they will be able to express a level of intelligence higher than they would have been able to individually (Lévy, Casalegno, & Amemado, n.d.).

Hong and Page (2008) used Aristoteles observations of the democracy to show that wisdom of crowds exist. Aristoteles saw that "when individuals see distinct parts of the whole, the collective appraisal can surpass that of individuals" (Hong & Page, 2008). We can use this to look at the democracy we have today. The individuals are everyone that has a right to vote. They have different opinions, interests, and priorities. They see different parts of the whole picture. Some thinks the agriculture is most important and others have transportation as their main priority. In an election every individual will follow their own opinion and vote accordingly. When all votes are added up, we get the total opinion of all the individuals. The result is a government and a Parliament that represents the collective judgment of the individuals.

A really simple example to explain the wisdom of crowds, and also show how easy it can be to use, is to picture the classic guessing game: "how many candies are in this glass jar?". Everyone can guess, and the winner is the closest one. It is then possible to take every guess and average them, which then becomes the *crowd's* guess. This guess should be closer to the true value than most of the individual guesses, or maybe even the closest one.

2.2 What is necessary for a crowd to be wise?

Surowiecki (2004) explains in his book *The Wisdom of Crowds* that there are five factors that is relevant for the wisdom of crowds to exist. This thesis will use these factors as the theoretical foundation.

2.2.1 Diversity

Diversity amongst the individuals means that they have different ways of thinking. This comes from the individuals' upbringing, education, experience, and other factors that will shape a person through life. This gives each individual a unique view of the world which is important for the crowd to be wise, because it reduces the risk of the crowd falling into destructive traps were everyone thinks the same (Surowiecki, 2004, pp. 32-44). Diversity also includes knowledge, and with different amounts and types of knowledge the crowd will be wiser (Surowiecki, 2004, pp. 32-44). When the individuals are diverse, their predictions are negatively correlated which gives a wise crowd (Hong & Page, 2008).

2.2.2 Independency

Independency is important because people are easily affected by what they see or hear other individuals do and say. When they are independent of each other they make predictions from their own private information, which reduces the chances that the individuals will make the same mistakes when predicting, which would have made the crowd unwise (Surowiecki, 2004, pp. 46-65).

2.2.3 Decentralisation

Decentralisation ensures that the power is spread out and that the individuals are encouraged to be independent. They are able to specialize and use their private knowledge to solve problems (Surowiecki, 2004, pp. 66-79).

2.2.4 Coordination

Humans have the ability to make choices and act based on how they think the people around them will act. The choices will be made based on what they think will be the best for the group, and with everyone in the group thinking like this, they are able to coordinate a great solution (Surowiecki, 2004, pp. 80-97).

2.2.5 Trust

Trust is a key factor because humans need to trust people and the systems around them. Humans do not like injustice. They think others should not get more for the same amount of work and they do not want to participant if they think others are not (Surowiecki, 2004, pp. 99-124).

2.3 What can negatively impact the wisdom of crowds?

The individuals in the crowd can affect the wisdom of the crowd in different ways. They can do it on purpose, but it is probably most common that they do it without understanding and without meaning to do it.

2.3.1 Not taking it seriously

Since trust is important for the crowd to be able to be wise, the individuals in the crowd can easily try to interfere with the results. By purposely predicting way off, or not taking the time to the consider the question, they could put the crowd's prediction away from the true value. They could also break the rules that have been set, like discussing the predictions with someone else.

2.3.2 Communicating and discussing

The reason the crowd is not wise is because the individuals in the crowd know too much about each other's thoughts and try to change their own thoughts accordingly (Surowiecki, 2004). When the individuals share information with each other they risk reducing the diversity, which Page (2007) points out has happened several times. When the individuals choose to go for the same solution, it is called herding (Surowiecki, 2004). The individuals do this because it is safest, and they will not risk being wrong by choosing what they think is best.

2.3.3 The madness of crowds

In the book *The Difference: how the power of diversity creates better groups, firms, schools, and societies*, Page (2007) writes about "The Madness of Crowds". He explains that people tend to follow the people around them and will base their choices on what they do and not follow their own thoughts. People can either do this spontaneously or think about it before they do it. When they take the time to think about it and still follow the crowd, Page (2007) blames it on a lack of diversity.

2.4 How to calculate and measure the wisdom of crowds

The easiest way of measuring how wise the crowd is compared to the individuals in the crowd, is to find the number of individuals in the crowd the crowd's prediction outperforms. This number is often given in percent. Another way is to measure how close the crowd's prediction is to the true value.

The reason it is possible to aggregate several predictions and get a better result than the individual predictions, is said to be because of the law of large numbers (Hong & Page, 2008). This is because when the individuals are not as accurate, the law of large numbers will make the errors cancel each other out (Hong & Page, 2008). For the errors to cancel each other out, the predictions need to be negatively correlated, which according to Hong and Page (2008) is crucial for the crowd's prediction to be the best it can be. When the predictions are negatively correlated, they are bracketing the true value, which means that the predictions are on either side of the true value (Larrick & Soll, 2006). When the predictions are then aggregated with the arithmetic average, the crowd's prediction should be close to the true value.

Many methods can be used to aggregate the individuals' predictions together. Since Galton (1907) used the arithmetic average, it can be looked at as the original way. Other methods that can be used is the geometric average, the median or the mode. It is also possible to give more weights to the individuals who is considered to predict better to try to accentuate the more knowledgeable individuals. Explanation of different aggregation methods:

Arithmetic average: Gives the average where every individual is weighted equally. When the predictions are on either side of the true value, the average will be close to the true value.

$$Arithmetic\ average = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Geometric average: Used when the variables are not independent of each other or the variables varies a lot in value (Blokhin, 2020).

$$Geometric\ average = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Median: The central variable of a set of variables ordered by value, descending or ascending (Ganti, 2019). The median is not sensitive to extreme values (Australian Bureau of Statistics, 2013).

The mode: The variable represented the most times. Can be used for both numerical and non-numerical data (Australian Bureau of Statistics, 2013).

Weighted: Can be used with any aggregation method. A method to single out the individuals who should have more positive influence on the crowd's prediction is all that is necessary.

3 Methodology

3.1 Literary review

For conducting a literature review it was decided that a traditional review should be carried out. The wisdom of crowds is a subject that can be studied in different study fields, which supplies a numerous amount of research papers. To be able to answer the problem statement of this thesis, while not being able to read through all the available papers, a traditional review was the answer (Cronin, Ryan, & Coughlan, 2008). The further subchapters will look into the criteria set for this literature review.

3.2 Selecting a review topic

Since the subject is so broad, it was necessary to scale it down. The focus was put on people who was asked specific questions and gave individual predictions. Their predictions were then aggregated together as one prediction. Papers studying only prediction markets was excluded because of the way they are driven by the prospect of winning money. Also, papers including only groups who worked together to either come up with a solution or prediction in unison was excluded, because this thesis' focus is on individuals who predicts.

3.3 Searching the literature

Google Scholar was chosen as the search engine, as it contains papers from many study fields (Google Scholar, 2020). To be able to find the papers containing the correct subject, a few key words were chosen to be the search words. The words that were used was: *wisdom of crowds*, *crowd predictions* and *collective intelligence*, and they provided more than 2 570 000 results. To set a boundary for how long to search for papers, it was decided to stop after 10 pages for *wisdom of crowds*, and after five pages for the other search words, each page containing 10 results. It was clear early on that *wisdom of crowds* was the search word that provided more papers inside the boundaries that was set, which is why it was looked into the most. To make sure that the papers contents was reliable and worth looking at, a citation boundary was set. It was set to a minimum of 10 citations for a paper published in 2010, could be less for a newer study.

3.4 Gathering, reading, and analysing the literature

At first the papers abstract, introduction and conclusion were read, to get a sense of what the papers were about. Papers that did not contain the selected topic was removed. In total 21 studies were found via *Google Scholar*, and the overview of how many papers that were chosen from each search word is listed down below, in *Table 1*. As they were found they were added to Table 2, so it would be easy to keep track of them, know how they were found and why they were chosen.

Search words	Number of papers chosen
Wisdom of crowds	17
Crowd predictions	4
Collective intelligence	0

Table 1: The number of papers found in each search word

Title	Key words	Result	Search word

Table 2: The table used to keep control over the chosen papers

Furthermore, seven more papers were provided from the supervisor of this thesis. These were papers he thought would be interesting for the thesis. In total there was 28 papers ready to be read more closely.

When each paper was read through in more detail, the interesting aspects of the paper was added to another table, like *Table 3* (see Appendix for the filled out table). This made sure that similar information from each paper was found, which made the job of finding the papers that contained similar aspects easier to allocate after when the writing for the thesis started. The table helped to make sure that nothing important was left out.

Title	Author	Journal	Method	Crowd	Aggregation	Information	Results

Table 3: The table used to keep track of the papers

The results that comes from this literature review can be used to learn about the wisdom of crowds. The focus in this thesis is not what type of study the different studies have done, but which methods they have used to aggregate. The result show that the wisdom of crowds will not just be apparent in studies and their surveys that is put together just for the sake of the study. The wisdom of crowds is also likely to be apparent in the real world, for situations where one can have similar attributions as is found in these papers. The result is therefore generalizing with great transferability.

3.5 Strengths and weaknesses

A big weakness to the traditional literature review is all of the papers that is not found and could possibly contain interesting information that would either support or contradict the findings of this thesis. When time is short it is not possible for a single person to read through over 2 million papers. A solution would be to use a systematic literature review and use a computer program to search through all the papers.

On the other hand, a traditional literature review is not complicated, and anybody could conduct one if they wanted to. It is easy to understand how the review have been carried out. The fact that a person has researched, read, and chosen out papers manually, gives the review a different personal feel than a review done with the help of a computer program.

Another weakness is that only one database has been used but considering all the papers available there for the few search words that was used, more databases would only make it harder to find the papers with the right theme.

When a person is choosing which papers to study further, that person will have personal bias. When a person has spent time thinking and reading to try and figure out what to research, they develop bias towards what kind of information they want to find. Even though one tries to not be affected by the bias, one can never be sure that the bias is completely gone from the choices that is made.

4 Findings

The papers that were found had constructed surveys specifically for the purpose of writing their paper or they used previously found data that contained predictions. The studies were conducted in different ways, but what they all had in common was that they contained data from individuals that gave an answer to a concrete question. Then these data were aggregated in different ways to conclude if the individuals showed wisdom as a crowd. *Figure 1* how many times Francis Galton and James Surowiecki are mentioned in the papers. It is interesting to see that so many different papers rely on the same theory, and that these two are important for the wisdom of crowds.

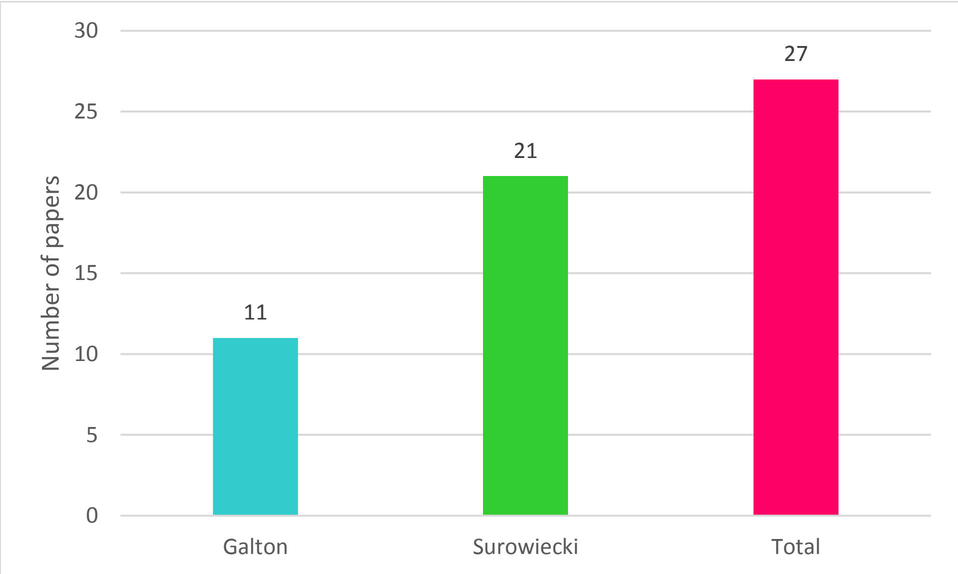

Figure 1: The number of papers the two authors are mentioned in, compared to the total number of papers

4.1 Aggregation methods

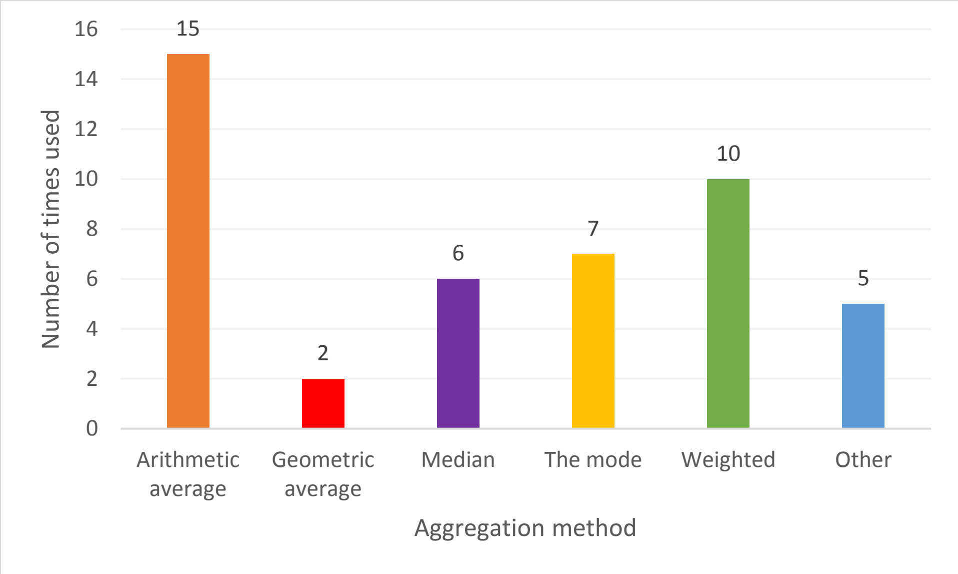

Figure 2: The number of papers that used the different aggregation methods

By looking at *Figure 2*, it is clear there are many possible ways of aggregating a crowd's predictions. The simplest one is arithmetic average, which is widely used (Ariely et al., 2000; Atanasov et al., 2017; Becker, Brackbill, & Centola, 2017; Da & Huang, 2020; Endress & Gear, 2018; Kattan, O'Rourke, Yu, & Chagin, 2016; Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Navajas, Niella, Bahrami, & Sigman, 2018; Palan, Huber, & Senninger, 2019; Poister & Thomas, 2007; Simmons, Nelson, Galak, & Frederick, 2011; von der Gracht, Hommel, Prokesch, & Wohlenberg, 2016; Vul & Pashler, 2008; Wagner & Suh, 2014; Wagner & Vinaimont, 2010). Using the arithmetic average implies that all members of the crowd are weighted equally. One would think that some members have more relevant information or expertise than others. However, with the absence of any cues on members' expertise a statistical measure of centrality (e.g., arithmetic average or median) is often preferred.

Several compare arithmetic average with other simple aggregation methods like the geometric mean (Lorenz et al., 2011; Palan et al., 2019) and median (Becker et al., 2017; Griffiths & B., 2006; Hueffer, Fonseca, Leiserowitz, & Taylor, 2013; Lorenz et al., 2011; Palan et al., 2019; Simmons et al., 2011).

While these are easy approaches to use for many aggregation problems, there are other methods that can be used which might give a better result but that require more information.

4.1.1 Arithmetic Average

Using the arithmetic average as the aggregation method can be said to be the "original" way of aggregating a crowd of individuals' predictions, because that is what Francis Galton (1907) did. Some of the papers mention him and his experiment (Becker et al., 2017; Lorenz et al., 2011; Navajas et al., 2018; Simmons et al., 2011; von der Gracht et al., 2016) as seen in *Figure 1*. They do not state any other reason for using the arithmetic average, so it is likely that their reason is because that is what Galton used.

A reason that was stated for wanting to use the arithmetic average was to test it against other aggregation methods (Atanasov et al., 2017; Becker et al., 2017; Palan et al., 2019). The benefits to the arithmetic average is that it reduces error if the predictions bracket the truth (von der Gracht et al., 2016) and it is easy to use (Ariely et al., 2000; von der Gracht et al., 2016). Disadvantages to the arithmetic average is that it might not be feasible to conduct in real life situations, like having several doctors assessing a patient and giving a prediction about their health in a year (Kattan et al., 2016). The arithmetic average is also underconfident in prediction polls, while the individuals usually is overconfident in prediction polls, which often give equal probabilities for the options in the poll (Atanasov et al., 2017).

Out of these 15 papers that uses the arithmetic average, 13 of them got good results by using the arithmetic average (Ariely et al., 2000; Becker et al., 2017; Da & Huang, 2020; Endress & Gear, 2018; Kattan et al., 2016; Navajas et al., 2018; Palan et al., 2019; Poister & Thomas, 2007; Simmons et al., 2011; von der Gracht et al., 2016; Vul & Pashler, 2008; Wagner & Suh, 2014; Wagner & Vinaimont, 2010). Lorenz et al. (2011) got predictions that was right-skewed, and as stated, the arithmetic average performs best when the predictions bracket the truth. For Atanasov et al. (2017) it did not work the best when aggregating prediction polls and comparing against prediction markets, because prediction markets consider updated predictions, the individuals' skills and corrects over- and underconfidence in the individuals, which the arithmetic average does not do for prediction polls.

4.1.2 Geometric average

The geometric average is used because the distribution of the predictions from the crowd is log-normal, and the geometric average was a better fit than the arithmetic average (Lorenz et al., 2011), and it worked well as an aggregation method. Palan et al. (2019) used the geometric average to compare different aggregation methods, and the geometric average worked well, but it was not the best.

4.1.3 Median

The median is mainly used for comparing with other aggregation methods (Becker et al., 2017; Griffiths & B., 2006; Lorenz et al., 2011; Palan et al., 2019; Simmons et al., 2011), but also used when the data is to extensive to digitalize for further calculations (Hueffer et al., 2013). Hueffer et al. (2013) had data from 54 years, and predictions from each year was compiled in a book chronologically, which made it easy to find the middle page and use the last prediction on the page as the median.

The median gives good results in showing the crowd's prediction (Becker et al., 2017; Griffiths & B., 2006; Hueffer et al., 2013; Lorenz et al., 2011; Palan et al., 2019; Simmons et al., 2011) and even outperformed the arithmetic and geometric average (Palan et al., 2019). An advantage of median is that it is not sensitive to outliers. This can be especially relevant in cases where respondents answer without any relevant knowledge or if they do not give much thought to their judgment (Galton, 1907).

4.1.4 Weighting

Weighting is used for giving individuals more influence on the crowd's prediction. Different methods are used, each trying to find the method that improves the crowd's prediction the most. The methods consist of giving weights based on the individuals confidence in their prediction (Simmons et al., 2011; von der Gracht et al., 2016), the individuals past performance (Mannes, Soll, & Larrick, 2014; von der Gracht et al., 2016), summing probability predictions (Murr, 2011), Borda count (Miller, Hemmer, Steyvers, & Lee, 2009; Miller & Steyvers, 2011) or accentuating the individuals who are expected to predict good (Atanasov et al., 2017; Budescu & Chen, 2015; Nebbione, Doran, Nadella, & Minnery, 2018; von der Gracht et al., 2016).

Using weights yielded both good and bad results. When using individuals confidence, the crowd did not appear to be wise, caused by their own personal bias and therefor being overconfident (Simmons et al., 2011). For von der Gracht et al. (2016) the individuals was not overconfident, and was wiser compared to the individuals aggregated by past performance. Even with good results, they still found no extra value in weighting the crowd, because equally weighting gave better results. Summing probability predictions yielded a good crowd performance (Murr, 2011), and the Borda count did as well (Miller et al., 2009; Miller & Steyvers, 2011).

When trying to accentuate the better individuals both past performance and confidence is assessed, as well as the latent topics of the predictions they performed well on (Nebbione et al., 2018). The individuals were ranked, and only the highest ranked was averaged as the crowd, which gave a good crowd prediction. The individuals can also be assessed by how they contribute to the crowd prediction, and only the individuals which have a positive contribution is aggregated as the crowd's prediction (Budescu & Chen, 2015). Both compared their method to weighting only on past performance, which was outperformed. This is also supported by (von der Gracht et al., 2016) which used the top 15 individuals ranked solely on past performance, which was their worst aggregation method. But Mannes et al. (2014) who also ranked the individuals based on past performance had better results when they used the top 5 individuals. The individuals can be given weights based on their past performance and how often they have updated their prediction, which is calculated with an algorithm (Atanasov et al., 2017). This algorithm outperformed continuous double auction which is more costly to use (Palan et al., 2019).

4.1.5 The mode

The mode is used when the predictions is not numbers that can easily be averaged, but when the predictions is a word. Such as which football player or team performs best (Goldstein, McAfee, & Suri, 2014; O'Leary, 2017) or which political party that will win the election (Murr, 2011, 2016).

The mode worked well as an aggregation method when the individuals predictions consists of one word, and the crowd also performed well (Goldstein et al., 2014; Murr, 2011, 2016; O'Leary, 2017), but it is not the best method for ranking the order of events, because it is rare that several individuals will agree on a ranking of 10 items (Miller et al., 2009).

A slightly different kind of the mode that works for predictions consisting of a combination of predictions, like the Traveling Salesman Problem (Yi, Steyvers, Lee, & Dry, 2012) and ranking the order of occurred events (Miller et al., 2009), is the "Greedy count" where the most occurring placement for each prediction is aggregated as the crowds prediction. But even this is not the best method for aggregating the ordering of occurred events, which must be because the individuals' orderings differ to much (Miller et al., 2009).

4.1.6 Other

There are other, more specific methods which is not used as much. The Kemeny-Young method, were the prediction from an individual that is closest to all other individuals' predictions is the crowd's prediction, works well for combinatorial problems (Miller et al., 2009; Yi et al., 2012). The

Kemeny-Young is the best method for aggregating the ordering of occurred events (Miller et al., 2009).

Palan et al. (2019) compared the simple aggregation methods to continuous double auction and call auction, where they found that the continuous double auction outperformed all other methods. They stated that continuous double auction is expensive to use, considering the time and resources needed to construct a correct way to conduct it. Therefor they explain that it is important to consider each case if such extensive measures are necessary to get a good crowd prediction, or if simpler methods can be sufficient.

4.2 Information methods

Providing information to individuals or information sharing among individuals in a group can make the group less diverse. For instance, by hearing how another individual is thinking might start to shift one's thoughts towards the other individual's opinions. As stated in subchapter *2.2.1 Diversity*, the theory behind the wisdom of crowd emphasises that all individuals should be as diverse as possible to get a good group prediction, and that the sharing of thoughts will worsen the group prediction. However, studies show that different amounts of information can do the opposite and make the crowd wiser, which the next subchapters will look further into.

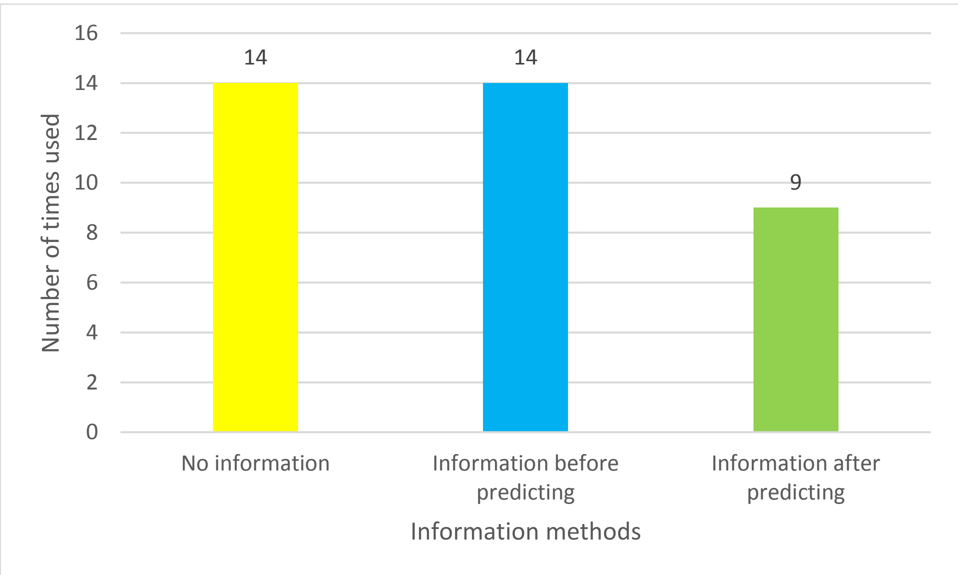

Figure 3: The number of papers that used the different types of information methods

4.2.1 No information

Surowiecki (2004) emphasised the importance of not using additional information to preserve the crowd's diversity. This approach is used by several where they want to study the crowd the way the theory says it should be (Ariely et al., 2000; Miller et al., 2009; Poister & Thomas, 2007; Vul & Pashler, 2008; Wagner & Suh, 2014; Wagner & Vinaimont, 2010; Yi et al., 2012). Another approach is to compare groups with no added information to groups with added information (Becker et al., 2017; Endress & Gear, 2018; Lorenz et al., 2011; Miller & Steyvers, 2011), this will be looked at in the next subchapters.

The crowds was over 70% better than the individuals who was aggregated together (Miller et al., 2009; Poister & Thomas, 2007; Wagner & Suh, 2014; Yi et al., 2012), and those papers who did not present the wisdom of crowds in the same manner, they also had good results (Ariely et al., 2000; Wagner & Vinaimont, 2010). Their findings supports the theory about not using information. Vul and Pashler (2008) tested if individuals would be more accurate if they could give two predictions for the same question and average their answers. The individuals were not informed about giving a second prediction. One group did it right after, while the other group did it after three weeks. The second prediction from both groups was poorer than the first, which shows that they were not able to gain any new knowledge in between giving the predictions. Even so, the average of an individual's two predictions was better than either of them, and it was best for the group who predicted a second time after three weeks. This shows that it might be beneficial to think twice.

4.2.2 Information available before giving a prediction

Several studies do not set any boundaries for what information the individuals can use. This involves being able to research information or even having access to previous predictions from individuals and the crowd. The reason for using no boundaries is because the individuals are encouraged to update their predictions as often as they want, which often ends up being every time they get new information (Nebbione et al., 2018; von der Gracht et al., 2016) or because they know what prediction they are making and can look up information to help them predict (Atanasov et al., 2017; Budescu & Chen, 2015; Da & Huang, 2020; Goldstein et al., 2014; Hueffer et al., 2013; O'Leary, 2017). The individuals also showed great wisdom of crowds, when they in addition to having no boundary for how much information they could consume, they were grouped together and could discuss amongst themselves before giving individual predictions (Atanasov et al., 2017).

Comparing crowds who get access to information and crowds who do not get access to information is a great way to find out if information has any effect on the wisdom of crowds. When using an iterative method to share information, where the individuals received the predictions from another individual before giving their own prediction, the crowd turned out to be wiser (Miller & Steyvers, 2011). In this study they were ranking the order of events, and the first half of the questions was presented with a random ordering of the events. The second half of the questions was given with the ordering from another individual, and they were then informed that it was already ordered by at least one individual, and they could change it as they wanted to. This means that every individual received an ordering, but those who were communicated iteratively did better, which was what Miller and Steyvers (2011) wanted to find out.

When Da and Huang (2020) studied individuals who had no boundaries before making predictions, which included information about the crowds previous predictions and even individuals' predictions, they found that the crowd was showing signs of herding. After removing the information about other individuals' predictions as well as the crowd's prediction, the crowd got more diverse and wiser. As a result, this study's findings support Surowiecki's view that information sharing can reduce diversity and lead to worse predictions.

When studying football fans, it is apparent that personal bias is what ruins the crowd's possibility to be wise. Simmons et al. (2011) gave information about a point spread being biased against favourites, but still they found that both the crowd with this information and the one without still betted on favourites. So, when they received information, they were not able to use it, they just followed their own bias that said the favourite would win. This suggests that the wisdom of crowds can be vulnerable to the kind of systematic biases of which the research of Kahneman and Tversky (1974) have focused on.

4.2.3 Information available after giving a prediction

An interesting way of finding out if information influences the crowd's wisdom is to give them information after they have predicted once and see if they get wiser. An easy way of doing this is to share the arithmetic average of the crowd's response from the former prediction, and then let them predict again (Becker et al., 2017; Endress & Gear, 2018; Lorenz et al., 2011).

When Becker et al. (2017) used the arithmetic average they compared it with giving a group the prediction from one of the individuals in the group and with a group that was not given any information. For the group with no information, they had no significant change in their predictions after predicting a second time. The group that received the arithmetic average the average error decreased about 10%, but for the other group that received the prediction from a random individual had an average error decrease of 43%. The decrease was when the random individual's prediction was towards the true value, when it was away from the true value, there was an increase in the average error of 19%.

Lorenz et al. (2011) compared the groups with arithmetic average to groups that received information about all the individuals' predictions and a group with no added information. They found that information worsens the crowd's prediction, and the crowd became less reliable.

Making the individuals come together in groups to discuss is really going against the theory of keeping individualism and diversity (Surowiecki, 2004). Navajas et al. (2018) did this, where the individuals predicted before and after group discussion, and once as a group. The individual predictions after the group discussion was more accurate than both the individual and group predictions. It is also possible to use asymmetric information, where every individual receives different amounts of factual information, which helped the crowd give a better prediction (Palan et al., 2019).

These studies show that information influence crowd predictions, but that the type of influence depends both on what type of information is shared and who receives the information. Specifically, administrating information about the crowd prediction (Becker et al., 2017) or about complete information about participants prediction that allow inferring crowd prediction (Lorenz et al., 2011) have ambiguous effects on accuracy of crowd prediction. However, providing only information about another individual prediction (Becker et al., 2017) or prediction of a small sub-set of the total crowd (Navajas et al., 2018) appear to improve crowd prediction. An explanation why the effect differs based on the amount of information about other participants predictions can be linked to the diversity effect. When participants only receive information about a limited set of other participants the diversity in the group is retained; the information might induce individuals with some of the most extreme predictions to modify their judgements, but overall still retain a 'healthy' variation in predictions. However, when participants receive information about the entire group's predictions it is more likely that some type of regression towards the mean effect will prevail.

4.3 The crowd

The crowd has certain attributes that needs to be present. The number of individuals have to be *large* (Hong & Page, 2008), the individuals should be diverse and have some knowledge (Surowiecki, 2004). It is interesting to try and find out how important these attributes are, and if they can be used at various levels. How small or large the crowd can be is also a useful attribute, because a lot of resources can be saved if only a few individuals is necessary.

4.3.1 The number of individuals in the crowd

The number of individuals that is aggregated together varies from 2 (Mannes et al., 2014) to 61 653 (Goldstein et al., 2014). Different numbers of individuals are also used in the same study. The reason is to see how few and how many individuals one can use and still have the crowd be smarter than the individuals (Goldstein et al., 2014; Kattan et al., 2016; Miller & Steyvers, 2011; Navajas et al., 2018; Wagner & Suh, 2014; Yi et al., 2012).

Kattan et al. (2016) found that five individuals were the lowest they could go, but for each added individual the crowd got wiser. The increase in accuracy got less for each individual they added until they had 22 individuals, then the accuracy levelled out. The top five individuals based on past performance can be a "very robust judgment strategy" according to Mannes et al. (2014). The top 10 individuals based on performance is also found to give a better crowd prediction than the whole crowd (Goldstein et al., 2014), while the top 15 individuals based on past performance gives the worst crowd prediction (von der Gracht et al., 2016). Miller and Steyvers (2011) found that aggregating more than 15 did not make the crowd wiser, and when they gave the individuals information while predicting they did not get wiser after 10 individuals was aggregated. In Navajas et al. (2018) they had the individuals in groups of five where they came up with a group prediction, and averaging only four groups was about 50% more accurate than the average of 1 400 individual predictions. They also found that after the individuals predicted a second time individually, they could average five individuals that had been a part of different groups, and they outperformed all 5180 individuals. But it is also possible to get great results with larger crowds (Griffiths & B., 2006; Miller et al., 2009; Murr, 2016; von der Gracht et al., 2016).

4.3.2 Diversity

A factor that is said to lessen the diversity is social influence (Surowiecki, 2004). When giving the individuals information about the average prediction of the crowd in between predictions, Becker et al. (2017) found that the added information reduced the diversity of the crowd, but their accuracy still improved. This makes sense since when individuals, on average, predict more accurately, the dispersion of their predictions necessarily reduces. Lorenz et al. (2011) also provided the average prediction of the crowd for four rounds of predictions and discovered that the information reduced both diversity and accuracy. The results for Da and Huang (2020) was similar, because their individuals could revise information before giving their predictions, and their predictions were poor and not diverse. As pointed out earlier, exposition to information can make people more like-minded in their assessments. In contrast, after (Da & Huang, 2020) removed some of the information, the predictions became more accurate and more diverse.

4.3.3 Experts or lay people

Since knowledge is supposed to improve prediction accuracy (Surowiecki, 2004), it is of interest to look at how lay people compare to experts, who necessarily should be more knowledgeable. Three different studies that compare the two groups have different results. Poister and Thomas (2007) used employees from different departments, where the highest department predicted one and two questions better than the other departments lower in the organizational hierarchy. The crowd predicting on the World Cup 2014 "was statistically significantly better than three out of five experts" (O'Leary, 2017). Endress and Gear (2018) used groups of lay people and groups of experts where one group of five experts was only 1,6% points better than a group of 21 lay people. One single expert was worse than the group of lay people, but in total all the experts aggregated together was wiser than all the lay people aggregated together.

4.3.4 Learning by doing

When individuals predict the same question several times, there is a possibility that they will learn, and their predictions will gradually improve. A few studies have found support for this view. Wagner and Suh (2014) had the individuals predict the weight of a cup containing different substances. There was a possibility that the individuals would use the density of the substance to predict the next substances weights. However, they found that with the unknown substances the individuals just guessed and were unable to transfer knowledge from previous exercises with known substances. Palan et al. (2019) found that the individuals' predictions improved over time and that there were small proofs of learning. In Simmons et al. (2011) there was great opportunities for a positive learning effect, because the individuals predicted over several weeks. But the individuals' bias interfered, and they continued to predict similarly, even when they saw that they lost.

4.4 Aggregation and information

As we have seen from the findings in different studies, aggregating with added information can change the crowd's wisdom in different ways, depending on what kind of aggregation method is used and what kind of information they get. In the following we focus on how the aggregation method can influence the wisdom of crowds.

4.4.1 Arithmetic average

When the crowd is aggregated with the arithmetic average and receives information the crowd's predictions become more accurate when the information the individuals receive is the crowd's average prediction (Becker et al., 2017; Endress & Gear, 2018), factual asymmetric information (Palan et al., 2019) or discussing with others (Navajas et al., 2018). The problem with information is that the individuals might put too much weight on the information at the expense of their own knowledge, which will make the crowd less diverse (Surowiecki, 2004). This was the case for Da and Huang (2020) when the individuals had access to information about others' predictions, the crowd's predictions and previous predictions, they found herding behaviour amongst the crowd's result. Herding is bad since it undermines diversity and its benefits. They removed the information about the others' predictions and also the crowd's prediction, which resulted in a 60% improvement in accuracy of the crowd's prediction compared to the situation with all the information available.

4.4.2 Median

According to Becker et al. (2017) the median can produce a wiser crowd when the individuals receives information about the crowd's average prediction or the prediction from a random individual in the group. In contrast, Lorenz et al. (2011) found that information about the crowd's average prediction or every individual's prediction can make the crowd less wise. Finally, Palan et al. (2019) found that asymmetric information made the crowd wiser when aggregated with the median.

4.4.3 Weighting

The studies that use different types of weighting schemes do not add information (Miller et al., 2009; Murr, 2011) or, conversely, have no boundaries for how much information they can consume (Atanasov et al., 2017; Budescu & Chen, 2015; Nebbione et al., 2018; von der Gracht et al., 2016). Miller and Steyvers (2011) used iterative communication when aggregating with the Borda count. The individuals were ordering the different sets of events, and half of the sets was done with the iterative communication. The Borda count turned out better for the sets of events with the iterative communication.

5 Discussion

5.1 Aggregation methods

The same aggregation methods do not work for every situation where you want to aggregate, but these papers have shown that several methods work for the wisdom of crowds. The arithmetic average was able to outperform the median (Becker et al., 2017) and be outperformed by the median (Palan et al., 2019). The reasons for this could be that in the former each individual received the same information about the other predictions which would lead the individuals towards similar predictions, while in the latter the individuals received asymmetric information about the true value and their further predictions would probably be more different because they have different information. Another essential point is the fact that Becker et al. (2017) aggregated 40 individuals while Palan et al. (2019) aggregated 144 individuals. Since the median works well with predictions in the outer region, that is a probable cause to why the median is better for Palan et al. (2019); the predictions from the individuals vary because they have different information and are a large group of people.

The arithmetic average performed good, and was able to outperform the weighted arithmetic average based on confidence and based on performance (von der Gracht et al., 2016). But the more advanced weighting methods that was based on positive contribution to the crowd's aggregate (Budescu & Chen, 2015) and past performance together with update frequencies of predictions (Atanasov et al., 2017) was slightly superior to the arithmetic average.

Weighting methods showed that they perform well, but they are also more costly. They perform well because they make it possible to find the *better* individuals, and the worst predictions which would make the crowd's prediction less accurate, is removed. This is a great advantage of the weighted methods. But there is also more data to be collected and processed, which will need more time and money to be able to generate the system that will collect the data and further use them for calculating. People who can make the system and use it is also necessary, and costly. This makes it important to assess each situation if the method is necessary or if a simpler one can give good enough results.

The variety of the mode that was used for a combination of predictions, the "Greedy count", was in one instance able to outperform the Kemeny-Young method, but in another case was not, both times with very small differences. When the "Greedy count" was worse than Kemeny-Young, the individuals were ranking the order of 10 events, but when the "Greedy count" outperformed the Kemeny-Young, the individuals was working with combinatorial problems, with 30, 60, and 90 different nodes. So maybe the “Greedy count” needs more events to order, and Kemeny-Young needs less.

It is not easy to compare the different aggregation methods when it is not possible to use all of them in the same types of studies. It is important to consider each case were wisdom of crowds is going to be used and figure out if the resources are available to use one of the weighted methods. This will be preferable if the cost is not too much. If the choice is to use the simpler methods, the cost of using the geometric average and the median alongside the arithmetic average is very low, so it is not necessary to choose between them. The mode and its variation the “Greedy count” is preferably used for aggregating words because they cannot be calculated, and Kemeny-Young works good for combinatorial problems and ranking the order of events.

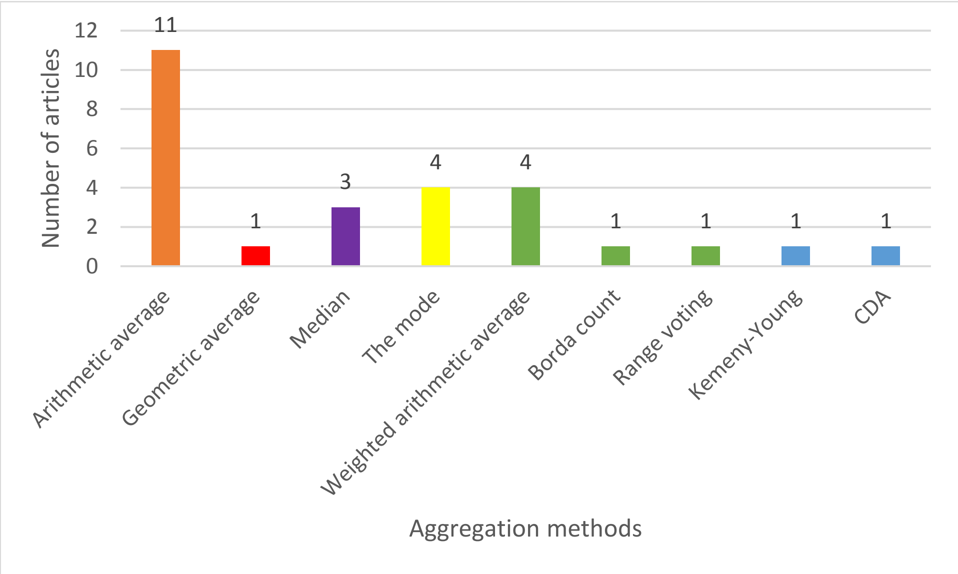

Figure 4: An overview of how many papers who had the best result with the different aggregation methods

5.2 Information methods

The different studies have shown that supplying information can actually improve the crowd’s predictions in some instances. Some of the methods used for allowing the usage of information seems to be not as versatile, as there are no boundaries for how much information, where or how they can find the information. While this is a method that works for predictions about the future, where the individuals can update their predictions as much as they want, it could maybe ruin the predictions because the individuals will not use their own knowledge to make the predictions, they will base their predictions on the information they can find because they feel it is better than their own information. This probably also counts for risky predictions like earnings forecast where the information caused herding, because the individuals did not trust their own opinions (Da & Huang, 2020).

Letting the individuals discuss with each other before they make a prediction have also shown that it helps the wisdom of crowds (Atanasov et al., 2017; Navajas et al., 2018). The individuals can share any thoughts they have, the crowd becomes less diverse, but still the crowd is more accurate. The reason this works could be that the individuals are still somewhat independent of each other and keeps some of the diversity in the group. The individuals manages to find the *perfect* middle ground between their own knowledge and the information they hear.

The simple way of sharing the crowd's average prediction with the individuals after they have predicted is an easy way of giving the individuals some controlled information. Becker et al. (2017) found the crowd to be less diverse after the average prediction was shared, but the crowd's prediction was more accurate. Lorenz et al. (2011) also experienced the crowd to be less diverse, but the crowd's prediction did not improve. The reason for this could be that the former used groups of 40 individuals, while the latter only had 12 in one group. Since the law of large numbers is essential (Hong & Page, 2008), it sounds plausible that the numbers have something to do with it. Because a larger crowd of individuals will according to the theory be able to cancel out the predictions, so the crowd's prediction gets closer to the true value. On the other hand, Da and Huang (2020) also found the crowd to be less diverse when they had information available, which included the crowd's previous prediction, and the crowd became both more diverse and wiser after the information was removed.

The arithmetic average does both improve and worsen in studies where information is used, and the type of information varies greatly. The same is found for the median and different weighting methods. Because the studies do not use the same information methods, it is difficult to compare and find a certain way that is the best for each aggregation method, and certainly not one that is universal.

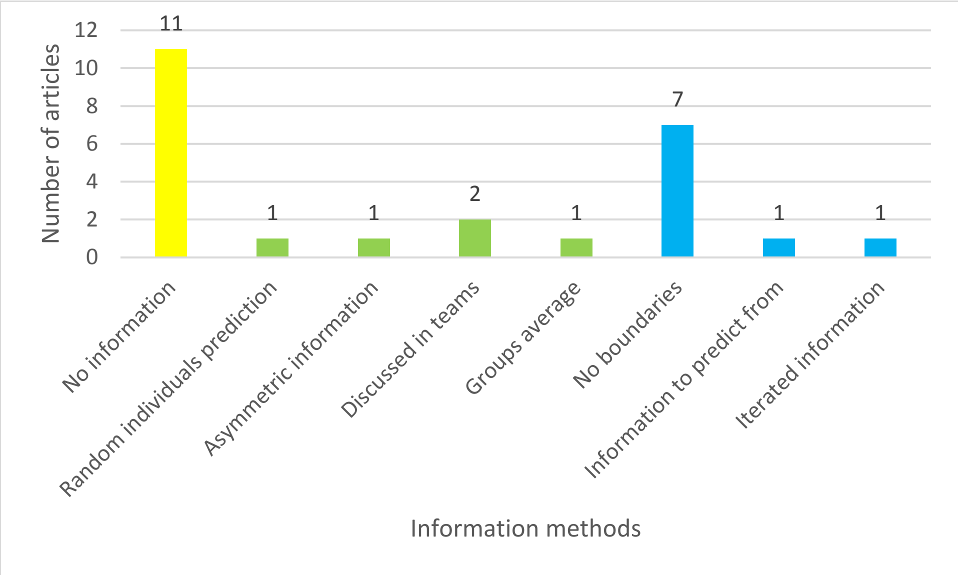

Figure 5: An overview of how many papers who had the best result with the different information methods

5.3 The crowd

While the law of large numbers seems to be important when using the arithmetic average, other studies have shown that the crowd can be wise even with lower numbers of individuals. Great results have been found by using as little as five individuals (Kattan et al., 2016; Mannes et al., 2014; Navajas et al., 2018), which is a feasible number to use. Navajas et al. (2018) used five individuals that had been a part of different group discussions and found that they outperformed 5180 individuals. When aggregating the top individuals ranked based on their past performance both top five (Mannes et al., 2014) and top 10 is found to be wise (Goldstein et al., 2014), while the top 15 is mediocre (von der Gracht et al., 2016). These studies show that when other factors are a part of the aggregation the crowd does not have to be large. While for the arithmetic average it is more important, as Kattan et al. (2016) got their best result by aggregating 22 individuals.

The studies that compare experts and lay people (Endress & Gear, 2018; O'Leary, 2017; Poister & Thomas, 2007), found that neither groups are significantly better than the other. The studies show that there is a slight benefit of using experts, but looking from the costly side of things, resources could be spared by using lay people instead.

6 Conclusion

The purpose of this study has been to find out what the empirical literature can say about the wisdom of crowds, try to get an overview of what they have studied and what they have found. The problem statement was as follows:

To what degree do empirical studies find that the wisdom of crowds leads to more accurate predictions?

The studies have shown that the wisdom of crowds is existent in different studies, with varieties in diversity, expertise, and methods used for information and aggregating. Most of the studies had situations with predicting that was specifically constructed for the study, and a few had taken data from a real situation.

The crowds perform well in all studies except for one, where the crowd's accuracy is almost at zero, and the blame is put on the individuals being too hanged up on their personal bias. Finding the right individuals will always be a challenge, but with only 1 out of 27 having extremely poor results because of bias is a good result. Based on the studies it is possible to say that wisdom of crowds leads to more accurate crowd predictions, but it does depend on the aggregation method, information method, expertise, diversity, and crowd size.

The research questions:

- In these empirical studies, what is the role of
    - Aggregation method
    - Information
    - Expertise and diversity
    - Crowd size

  on the crowds' prediction accuracy

All of the different aggregation methods that is found in the empirical studies are valid methods in giving the crowds prediction. It is not possible to say how each method affects the crowds' prediction accuracy specifically, because some methods outperform others and vice versa in the different studies. The arithmetic average has shown what a stable method it is, and it should give good results if weighted methods are not preferred to use.

These empirical studies show great proof that giving the individuals information can help the crowds' prediction become more accurate, in contrast to the theory. While there are also situations when the information worsens the prediction accuracy, in line with the theory. It seems like it all comes down to what kind of question the individuals get and what kind of people they are. If the individuals do not believe in themselves and think that they can come up with a good enough prediction because they find the question difficult, it is reasonable to think that they would predict closer to whatever information they get.

The advantage in using experts instead of lay people is found to be slim. There are bigger advantages to be found in a diverse crowd. While the crowd is found to be less diverse when information is distributed, the studies have both shown that the predictions can still be more accurate when the crowd is less diverse and be less accurate when the crowd is less diverse. As long as there is focus on having a crowd that is as diverse as possible, it is fair to think that if the amount the diversity is reduced is small, the crowd is still diverse enough to be accurate.

The crowd is able to be wise with different amounts of individuals, but it seems like the arithmetic average needs more individuals than more complex methods to get a more accurate crowd prediction. The more complex methods have more factors, like past performance, and the aggregation of the individuals' predictions are not as dependent on many predictions.

6.1 Future research

Since this thesis only looked at a small part of the empirical literature that is available, there are more literature out there which could substantiate the findings in this thesis or contradict them. But based on the findings here, there are more research that should be conducted, especially more real-life ones. Research consisting of real businesses where the crowd prediction can be tested against real-life results. Using individuals that work with the specific question in mind daily, mixed with individuals from other departments in the business and some individuals that work in another business from the same industry could be helpful in the case of how important expertise is. Future research should also include following the same crowd over some time and look at how their predictions change. Without using the same composition of individuals and changing the crowd size systematically it is difficult to tell which composition is the best. Also using the different aggregation and information methods with the same crowd is necessary to get a clearer result of how to best aggregate, and what factor matters the most when aggregating. A lot of interesting discoveries can be found, which can lead to more cost-efficient industries and better results with the help of the wisdom of crowds.

References

Ariely, D., Au, W. T., Bender, R. H., Budesco, D. V., Dietz, C. B., Gu, H., . . . Zauberman, G. (2000). The Effects of Averaging Subjective Probability Estimates Between and Within Judges. *Journal of Experimental Psychology: Applied, 6,* 130-147.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., . . . Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science, 63,* 691-706.

Australian Bureau of Statistics. (2013). Statistical Language - Measures of Central Tendency. Retrieved from https://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+measures+of+central+tendency

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,* E5070-E5076.

Blokhin, A. (2020). When to Apply the Geometric Mean: Key Examples. Retrieved from https://www.investopedia.com/ask/answers/060115/what-are-some-examples-applications-geometric-mean.asp

Budescu, D. V., & Chen, E. (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science, 61,* 267-280.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British journal of nursing, 17*(1), 38-43.

Da, Z., & Huang, X. (2020). Harnessing the Wisdom of Crowds. *Management science, 66,* 1847-1867.

Endress, T., & Gear, T. (2018, January). "Deliberated Intuition for Groups": An Explanatory Model for Crowd Predictions in the Domain of Stock-Price Forecasting. *Proceedings of the 51st Hawaii International Conference on System Sciences 2018,* 4094-4101.

Galton, F. (1907, March 7). Vox Populi. *Nature, 75,* 450-451.

Ganti, A. (2019). Median. Retrieved from https://www.investopedia.com/terms/m/median.asp

Goldstein, D. G., McAfee, R. P., & Suri, S. (2014, June). The Wisdom of Smaller, Smarter Crowds. *Proceedings of the fifteenth ACM conference on Economics and computation,* 471-488.

Google Scholar. (2020). About. Retrieved from https://scholar.google.com/intl/en/scholar/about.html

Griffiths, T. L., & B., T. J. (2006). Optimal Predictions in Everyday Cognition. *Psychology science, 17,* 767-773.

Hong, L., & Page, S. E. (2008, Mai 12). Some Microfoundations of Collective Wisdom.

Hueffer, K., Fonseca, M. A., Leiserowitz, A., & Taylor, K. M. (2013, March). The Wisdom of Crowds: Predicting a Weather and Climate-Related Event. *Judgment and Decision Making, 8,* 91-105.

Kattan, M. W., O'Rourke, C., Yu, C., & Chagin, K. (2016). The Wisdom of Crowds of Doctors: Their Average Predictions Outperform Their Individual Ones. *Medical Decision Making, 36,* 536-540.

KPMG. (2015). *Kun ett av fire byggeprosjekter leveres til avtalt tid.* Retrieved from home.kpmg/no/nb/home/nyheter-og-innsikt/2015/04/gcs2015.html

Larrick, R. P., & Soll, J. B. J. M. s. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *52*(1), 111-127.

Lévy, P., Casalegno, F., & Amemado, D. K. (n.d.). Collective Intelligence. Retrieved from https://www.collectiveintelligence.info/

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,* 9020-9025.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of personality social psychology, 107*(2), 276.

Miller, B., Hemmer, P., Steyvers, M., & Lee, M. D. (2009). The Wisdom of Crowds in Rank Ordering Problems. *9th International Conference on Cognitive Modeling*.

Miller, B., & Steyvers, M. (2011). The Wisdom of Crowds with Communication. *Proceedings of the Annual Meeting of the Cognitive Science Society, 33*(33), 1292-1297.

Murr, A. E. (2011). "Wisdom of Crowds"? A Decentralised Election Forecasting Model That Uses Citizens' Local Expectations. *Electoral Studies, 30,* 771-783.

Murr, A. E. (2016). The Wisdom of Crowds: What do Citizens Forecast for the 2015 British General Election. *Electoral Studies, 41,* 283-288.

Navajas, J., Niella, T., Garbulsky, Gerry, Bahrami, B., & Sigman, M. (2018). Aggregated Knowledge from a Small Number of Debates Outperforms the Wisdom of Large Crowds. *Nature Human Behaviour, 2,* 126-132.

Nebbione, G., Doran, D., Nadella, S., & Minnery, B. (2018). Deep Neural Ranking for Crowdsourced Geopolitical Event Forecasting. *International Conference on Machine Learning for Networking,* 257-269.

O'Leary, D. E. (2017). Crowd Performance in Prediction of the World Cup 2014. *European Journal of Operational Research, 260,* 715-724.

Page, S. E. (2007). The Difference: how the power of diversity creates better groups, firms, schools, and societies. In *The Difference: how the power of diversity creates better groups, firms, schools, and societies* (pp. 212-213). New Jersey: Princeton University Press.

Palan, S., Huber, J., & Senninger, L. (2019). Aggregation Mechanisms for Crowd Predictions. *Experimental Economics,* 1-27.

Poister, T. H., & Thomas, J. C. (2007). The Wisdom of Crowds: Learning from Administrators' Predictions of Citizen Perceptions. *Public Administration Review, 67,* 279-289.

Simmons, J. P., Nelson, L. D., Galak, J., & Frederick, S. (2011). Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds. *Journal of Consumer Research, 38,* 1-15.

Surowiecki, J. (2004). *The Wisdom of Crowds*. USA: Doubleday; Anchor.

von der Gracht, H. A., Hommel, U., Prokesch, T., & Wohlenberg, H. (2016). Testing Weighting Approaches for Forecasting in a Group Wisdom Support System Environment. *Journal of Business Research, 69,* 4081-4094.

Vul, E., & Pashler, H. (2008). Measuring the Crowd Within Probabilistic Representations Within Individuals. *Psychologial Science, 19,* 645-647.

Wagner, C., & Suh, A. (2014, January). The Wisdom of Crowds: Impact of Collective Size and Expertise Transfer on Collective Performance. *2014 47th Hawaii International Conference on System Sciences,* 594-603.

Wagner, C., & Vinaimont, T. (2010). Evaluating the Wisdom of Crowds. *Proceedings of Issues in Information Systems, 11,* 724-732.

Yi, S. K. M., Steyvers, M., Lee, M. D., & Dry, M. J. (2012). The Wisdom of the Crowd in Combinatorial Problems. *Cognitive Science, 36,* 452-470.

Appendix

Table of the papers reviewed for this thesis, next page.

	Title	Author	Journal	Method	Crowd	Aggregation	Information	Results
1.	Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds	Joseph P. Simmons Leif D. Nelson Jeff Galak Shane Frederick	Journal of Consumer Research (2010)	Four groups, four slightly different ways of betting on point spreads Participated for 17 weeks, and there where a minimum of 11 games per week They could win/lose small amount of money (max 21$ each week)	Knowledgeable fans	Calculated the percentage of people who wagered and the percentage of money that was wagered on the favourite, the underdog, and no preference More weight on the people who wagered more money on the game Mean and median predictions	One group got information about the point spread being biased Everyone could use their experience from the previous weeks	The groups betting on point spreads was not wise The group estimating the winner was wise The group betting on point spreads and estimating the winner was not wise, but wiser than only point spread betters
2.	Network dynamics of social influence in the wisdom of crowds	Joshua Becker Devon Brackbill Damon Centola	Proceedings of the national academy of sciences (2017)	1360 individuals were placed into networks, decentralized, centralized and control In each network they were placed in groups of 40, and had to answer a set of questions, and answered each question 3 times		Mean and median error DeGroot model	In decentralized and centralized they got information about the groups average answer before conducting the second and third prediction	In all groups the crowd got wiser after being able to revise their answer, but it was higher in those groups who got information
3.	Deep Neural Ranking for Crowdsourced	Giuseppe Nebbione Derek Doran	International Conference on Machine	Geopolitical forecasting Forecasters		The weighted arithmetic mean	Forecasters can update their predictions	Better predictions when the weight is put on the

	Geopolitical Event Forecasting	Srikanth Nadella Brandon Minnery	Learning for Networking (2018)	submitted a prediction and a confidence score Forecasts could be updated at any time, the most recent one was used to calculate		The weights were given based on each forecasters' ranking Brier score	several times as they get new information	individuals ranking than when it is put on their past performance
4.	How Social Influence Can Undermine the Wisdom of Crowd Effect	Jan Lorenz Heiko Rauhut Frank Schweitzer Dirk Helbing	Proceedings of the national academy of sciences (2011)	144 individuals were questioned in groups of 12 They sat alone with a computer and answered 6 questions, each question they answered 5 times	Students from a university in Zurich	Arithmetic mean Divided each estimate on the true value, took the logarithms of that number and then calculated the geometric mean	For two of the questions they got information about the 12 individuals average answer before giving their estimations For two questions they got information about every individuals' earlier predictions	Social influence effect was present The individuals became more confident when they got full information
5.	The Wisdom of Crowds: Predicting a Weather and Climate-Related Event	Karsten Hueffer Miguel A. Fonseca Anthony Leiserowitz Karen M. Taylor	Judgment and Decision Making (2013)	Used data from a betting game about when the ice on the river Tanana River will break Data from 54 consecutive years Each participant bets on the date	People all over Alaska participates	The median of all bets given in a year was used as the crowds' prediction	The participants can communicate, use data from the earlier years, use information about the weather and climate changes	The median is a good predictor

				and time the ice will break			over the last years	
6.	The Wisdom of Crowds: Impact of Collective Size and Expertise Transfer on Collective Performance	Christian Wagner Ayoung Suh	Hawaii International Conference on System Sciences (2014)	500 random people were collected from an online panel service They were told to do it without help, and answers were removed if they were perfect	The people were diverse in gender, age, education, and work	Arithmetic mean Calculated the collective error (CE) and individual error (IE), and measured the wisdom of the crowd as =IE/CE	When doing several predictions of the same kind (i.e. the weight of a cup of different substances) they could use their knowledge from the previous prediction to do the next	The crowd were not able to use experience for the difficult tasks The size of the crowd matter, it cannot be to large
7.	Aggregation Mechanisms for Crowd Predictions	Stefan Palan Jürgen Huber Larissa Senninger	Experimental Economics (2019)	144 individuals, divided into groups of 8 They are estimating the value of 4 jars filled with coins They are then given virtual money and virtual jars of money to buy and sell with and from each other	Students collected from a participant pool for students	Arithmetic average, geometric average and median values of individuals estimate Mean, median, closing prices and the closing bid-ask midpoint of a continuous double auction Uniform settlement price from a sealed bid-ask call auction	Four levels of information were given out after first estimation, two received each one Each level had the information from the prior level, but the highest level did not have all information	The individuals had lower estimation error when they had higher information level Their forecasts improved when they gained more experience Median gave the best aggregation of the simple methods Continuous double auction is the best aggregation method overall

8.	Harnessing the Wisdom of Crowds	Zhi Da Xing Huang	Management science (2019)	Data was collected from 3 previous years Estimize.com is a webpage were one can forecast earnings Tested removing almost all information after these 3 years	2516 users from the webpage Estimize.com Financial analysts, working professionals and students	Arithmetic average	Users can access as much information as they like about previous earnings or other users forecasts before making their own forecast Tried removing information so they had to make “blind” forecasts	The “blind” wisdom is more accurate 60% of the time, because it is more diverse The more information they view the more weight they put on the information versus their private information
9.	The Wisdom of Smaller, Smarter Crowds	Daniel G. Goldstein R. Preston McAfee Siddharth Suri	Proceedings of the fifteenth ACM conference on Economics and computation (2014)	Data from a season of fantasy soccer Focus on the managers’ choice of team captain of each week	Data from 100 000 random participants were chosen Everyone can play fantasy soccer	The most popular choice by the individuals is the choice of the crowd	There are no limitations with how much information one can consume	By aggregating those with higher prediction skills, it is possible to find smaller, smarter crowds
10.	Distilling the Wisdom of Crowds:Prediction Markets vs. Prediction Polls	Pavel Atanasov Phillip Rescober Eric Stone Samuel A. Swift Emile Servan-Schreiber Philip Tetlock Lyle Ungar Barbara Mellers	Management Science (2017)	2400 individuals made geopolitical predictions, some in prediction markets, some in prediction polls and individuals and some in teams	The individuals had a bachelor’s degree or higher and most of them where men	Simple mean Brier score Aggregation algorithm, with weight on prior accuracy and prediction update frequencies The teams’ prediction was the median	In teams they could discuss with each other as much as they wanted, but they made their own predictions	Prediction markets was better with simple aggregation Team prediction score was better with the advanced aggregation

11.	«Wisdom of Crowds»? A Decentralised Election Forecasting Model That Uses Citizens' Local Expectations	Andreas Erwin Murr	Electoral Studies (2011)	Pre-election survey for the 2010 British Election	13334 people from all over Great Britain	Percent of people who predicted the same winner The sum of all probabilities toward each choice (more weight for those votes with a higher probability)	No boundaries	Groups are better than individuals Same result for both aggregation methods
12.	The Wisdom of Crowds: Learning from Administrators' Predictions of Citizen Perceptions	Theodore H. Poister John Clayton Thomas	Public Administration Review (2007)	Mail survey conducted by employees compared to survey conducted by citizens	Employees at the Georgia Department of Transportation, from different groups of managers	Mean		They predicted right 36% of the time They were better when predicting about higher-priority service The group of executives was more accurate than the others
13.	The Wisdom of Crowds of Doctors: Their Average Predictions Outperform Their Individual Ones	Micheal W. Kattan Colin O'Rourke Changhong Yu Kevin Chagin	Medical Decision Making (2016)	24 clinicians in an advisory board meeting were asked to predict individually Compared with a statistical model	Selected experts in the field in question	Concordance index		The statistical model had better predictions, but the clinician was close with a group of only 5 clinicians
14.	Crowd Performance in Prediction of the World Cup 2014	Daniel E. O'Leary	European Journal of Operational Research (2017)	Data from Yahoo's World Cup Pick'em compared to different data like	Over 16000 participants in Yahoo's Pick'em	Majority of the crowd predicts the same option The option with the most predictions	The crowd could not see the expert's prediction, but	The crowd was better than experts, other crowdsourced data, and

				Yahoo's World Cup experts		Brier score	the experts could see the crowds'	stochastic prediction models
15.	Evaluating the Wisdom of Crowds	Christian Wagner Tom Vinaimont	Proceedings of Issues in Information Systems (2010)	30 people predicted the value of a variable Simulation comparing experts to crowds of various sizes		Mean QIC=Individual error (IE) /Collective error (CE)		The crowd predicted close to the true value Large crowds are better than expert
16.	Optimal Predictions in Everyday Cognition	Thomas L. Griffiths Joshua B. Tenenbaum	Psychological science (2006)	People were predicting duration/extent of everyday phenomena compared with a Bayesian model	350 under-graduates	Median based	Each question included a number which was the basis for their prediction One of five possible numbers for each question	People estimated close to the model Shows how peoples knowledge is in line with the statistics of the world
17.	The Wisdom of Crowds in Rank Ordering Problems	Brent Miller Pernille Hemmer Mark Steyvers Michael D. Lee	9th International Conference on Cognitive Modeling (2009)	Rank the order of events and magnitude of properties Compared to a Thurstonian model	78 under-graduates from the University of California	Mode - the most frequent occurring sequence "Greedy count" – the most occurring placement Kemeny-Young – the ranking that is closest Borda count – weights given to the rankings (10 to 1) and summed, the final ranking is from	No communication was allowed	Mode was the worst aggregation method, Kemeny_young was the best The Thurstonian model is very close

						highest to lowest number		
18.	«Deliberated Intuition for Groups»: An Explanatory Model for Crowd Predictions in the Domain of Stock-Price Forecasting	Tobias Endress Tony Gear	(2018)	Different groups predicted stock price	Professionals and non-professionals	Mean	E-Delphi, they got information about the groups' prediction	Groups of experts were best at predicting The e-Delphi group of lay people were close
19.	The Wisdom of Crowds: What do Citizens Forecast for the 2015 British General Election	Andreas E. Murr	Electoral Studies (2016)	Data from predictions before elections between 1964 and 2015	Citizens from all the mainland constituencies	The one with most predictions		The citizens as a group were wise with each election
20.	The Wisdom of Crowds with Communication	Brent J. Miller Mark Steyvers	Proceedings of the Annual Meeting of the Cognitive Science Society (2011)	Rank the order of events and magnitude of properties Compared to a Thurstonian model	172 undergraduate students from the University of California	Borda count, highest weight given to the first item in a list, and lower weight for each spot Eventually all weight is summed up and ranked from highest to lowest T is calculated as distance from the truth	Half of the questions was given with the last individuals ranking, as a form of communication	The questions with communication are better
21.	Identifying Expertise to Extract the Wisdom of Crowds	David V. Budescu Eva Chen	Management Science (2015)	Binary predictions of the likelihood of current events Forecasts of the real GDP rate in	Volunteer forecasters who chooses to forecast at any time	Each judges' contribution is calculated relative to the other judges' contributions =C_j		This aggregation model outperforms all other aggregation methods which is

				Europe with the ECB Survey of Professional Forecasters	Professionals from the financial industry	All of the positive C_j give the weighted mean of the crowd		tested with the same predictions Smaller crowds are as wise or even wiser
22.	Aggregated Knowledge from a Small Number of Debates Outperforms the Wisdom of Large Crowds	Joaquin Navajas Tamara Niella Gerry Garbulsky Bahador Bahrami Mariano Sigman	Nature Human Behaviour (2018)	Estimates about general knowledge quantities Answered individually then in groups of five, then individually again	5180 diverse people	Average	When answering in groups they had to agree on one answer so they had to discuss	The groups were wiser than the individuals
23.	The Wisdom of the Crowds in Combinatorial Problems	Sheng Kung Michael Yi Mark Steyvers Michael D. Lee Matthew J. Dry	Cognitive Science (2012)	MTSP and TSP	101 individuals	Maximizes local agreements Maximizes the overall agreement		Both aggregation methods is able to come up with a solution that is as good or better as the individuals'
24.	The Effects of Averaging Subjective Probability Estimates Between and Within Judges	Dan Ariely Randall H. Bender Christiane B. Dietz Hongbin Gu Thomas S Wallsten Wing Tung Au David V. Budescu Gal Zauberman	Journal of Experimental Psychology: Applied (2000)	J&W – Chose between two alternatives and told their confidence percentage W&G – Deciding if a skeleton was of a man with their confidence percentage New Experiment – Chose what was	60 under-graduates from Sweden 64 people from the University of North Carolina	Averaging over different number of individuals with the estimates and the estimates turned into log-odds Mean estimate as a function of the objective stimulus probability Mean		Closer to the truth when averaged over more individuals

				true of two alternatives and gave their confidence percentage				
25.	Testing Weighting Approaches for Forecasting in a Group Wisdom Support System Environment	Heiko A. von der Gracht Ulrich Hommel Tobias Prokesch Holger Wohlenberg	Journal of Business Research (2016)	Online forecasting competition about macroeconomics variables Submitting low, high and best estimate for each variable	28000 LinkedIn users with a finance or economics background	A computer system	They received information about the others forecast, true values, consensus forecast	Equally weighted triangular forecast gives the best results
26.	Measuring the Crowd Within Probabilistic Representations Within Individuals	Edward Vul Harold Pashler	Psychological Science (2008)	Individuals were asked questions about the world Half were asked to give a second answer right after completing, the other half after 3 weeks	428 people from an internet based subject pool	The two guesses were averaged into one answer All guesses were averaged	No one were told they were to give a second answer	The average was more accurate than both the first and second answer Better when the second answer came 3 weeks later
27.	The Wisdom of Select Crowds	Albert E. Mannes Jack B. Soll Richard P. Larrick	Journal of Personality and Social Psychology (2014)	Data from other studies, both experimental data and economic data		Arithmetic average on the top 5 individuals based on past performance		Top 5 individuals was better than the best member and the whole crowd