Testing Wisdom of the Crowd

Predicting the 2014 FIFA World Cup soccer group-stage matches

Name: **Bas Jacobs**

Student number: **322536BJ**

Degree: **Master of Economics**

Supervisor: **Yu Gao, MSc**

Table of Contents

ABSTRACT

In his book *The Wisdom of the Crowd* James Surowiecki reveals that, in the past, some groups made excellent decisions while other groups failed miserably at decision making via a large amount of examples. He explains that a group can make wise decisions collectively if certain requirements are met. He calls collective wise decisions 'wisdom of the crowd'.

In his paper *Making the Difference: Applying a Logic of Diversity* Scott Page uses the Diversity Prediction Theorem to mathematically explain that, by increasing the diversity, the crowd's opinion becomes more accurate compared to its average individual members.

To mathematically validate James Surowiecki's reasoning, an experiment was conducted among a crowd of 119 high school students, who fit the requirements for a crowd to be collectively wise. In the experiment the Prediction Diversity Theorem was used to decide if the crowd's opinion was significantly more accurate than the opinion of its average individual members. The theorem compares the crowd's opinion to that of its individual members, and in this experiment a reference crowd was used to validate that the crowd's opinion is not only accurate compared to its average individual members, but also has real-life value.

This thesis will provide evidence that wisdom of the crowd occurs when asking a crowd of 119 high school students to predict the outcome of the 2014 FIFA World Cup soccer group-stage games in Brazil. It will also show that, if you separate the crowd into a high and low knowledge group, the high knowledge group has a lower individual error on average, while the low knowledge group has a higher prediction diversity. These two findings combined show that the crowd's opinion of the low and high knowledge group do not differ significantly. Lastly, this thesis will show that the opinion of this crowd does not significantly differ from the predictions of a betting website, showing that the opinion of the crowd has real-life value.

Dedication

Many thanks go out to Yu Gao, for pointing me in the right direction and having patience, more than once. To Judith Zaagman, who helped with the grammar, so I could tell my story and to Ellen Jacobs for helping me to conduction the research with her students, lending me her subjects to become my subjects.

INTRODUCTION

For the past 25 years, I have gone to camping site ”Chateau du Montrouant” in France to spend my summer holidays. As I am not the only person who has this habit of returning, I meet the same people there each year. Among these people is a group of five girls spending their holiday together. What fascinates me is that, when, on almost any given day, I talk with these five girls separately, they have a clear picture of what they want to do: hang out at the pool, play tennis, go sightseeing; as long as they can do what they want individually without burdening the others, it is fine. The part that fascinates me is that, the moment they have to decide what to do as a group, they set their individual preferences aside, and make a decision that clearly not a single one of the individual members of the group is happy with. How five rational individuals can make such terrible group decisions is beyond me.

In his book *The Wisdom of the Crowd* James Surowiecki wrote that, if a crowd meets the requirements of diversity, independence, decentralization, and aggregation, it is most likely capable of collectively making wise decisions. Other crowds that do not meet said requirements are not capable of collectively making wise decisions. In this thesis an experiment is conducted to test if a crowd of 119 high school students that meets these requirements could be considered wise compared to its average individual members. The crowd is also split in a high and low knowledge group to see if knowledge is related to the accuracy of the predictions, and to see if there is a relationship between knowledge and the diversity between the groups. Finally, a reference crowd is used to verify if the crowd of high school students has predictive value.

The thesis is separated into three parts. The first part consists of background information on wisdom of the crowd, with chapters considering the requirements that need to be met for a crowd to become wise, and three types of problems that can be solved by using wisdom of the crowd. The second part consists of information on the reference crowd, the betting website, and the mathematical explanation behind the wisdom of the crowd: the prediction diversity theorem. The third and final part consists of the experiment and its results, with chapters concerning the hypothesis, the experiment, the data analysis, results, conclusion, and discussion and limitations of the research.

WISDOM OF THE CROWD

Surowiecki starts his book *The Wisdom of the Crowd* with the famous example of Francis Galton and the country fair. Francis Galton was a 85-year-old scientist, who wanted to prove that crowds were -of all things- certainly not wise. At a country fair in 1906,a crowd consisting of experts, such as farmers and butcher, and non-experts, such as curious bystanders, was asked to predict the weight of an ox. 800 people participated, and Galton used these predictions to provide evidence that the crowd holds no knowledge. He stated that ''the average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes'' (Galton, 1907). As this quote suggests, Galton believed that the power should not be in hands of the public, but in the hands of a select few, and he wanted to use the results of this experiment to strengthen his opinion. The results of the experiment proved exactly the opposite: the crowd's prediction, which was equal to the average prediction of its participants, was 1,197 pounds while the ox weighted 1,198. The crowd's prediction was more accurate than any of its individual members.

We can attribute this finding to luck, since, when you ask enough crowds to predict the weight of an ox, there will be crowd that gets it right. It could be that the lucky crowd was exactly the crowd Galton encountered. However, since then a lot of experiments have been conducted on this phenomenon, and came up with the same conclusion. If certain requirements are met, crowds can be very intelligent, and often even more intelligent than its individual members. We call this phenomenon 'wisdom of the crowd'.

Not all groups are wise, and there is an endless amount of examples where groups made very poor decisions. Said groups are not limited to those consisting of non-experts, as especially groups of experts can make very poor decisions. The Bay of Pigs Invasion, which was a specialized attack on Cuba, constructed by a group of men known for their knowledge and expertise on the subject, exemplifies this notion very well. However, the invasion failed miserably, and, when reflected upon the decisions made, its failure was contributed to a phenomenon called Groupthink. This is a phenomenon where people in groups come to an answer based on consensus, and try to avoid disagreement (Janis, 1972).In everyday life we encounter a lot of groupthink, especially when the subject the decision is based on holds little value. Think about questions as what to eat for dinner, or what movie to watch. We do not mind the choice that is made as long as we avoid disagreement. When groups make decisions based on groupthink, we will not find wisdom of the crowd within these groups. In the Bay of Pigs Invasion all members agreed that, when the invasion was spotted, the Cubans would be too late to adjust to it. By agreeing with each other and avoiding disagreement, they stopped looking at the plan critically, assuming that, if something was wrong with the plan, they as a group of experts would have noticed it.

So if not all groups are wise, what requirements must be met for a group to become wise, and what problems can a wise group help solve? There are four requirements or conditions that must be met before wisdom of the crowd occurs: diversity (each individual member has some private information), independence (individual opinions are not determined by the opinions of those around them), decentralization (people are able to specialize and draw on local knowledge), and aggregation (a way to transform the individual opinions to a collective opinion). Wisdom of the crowd can help to solve three types of problems: cognition problems (problems with definitive solutions, who wins this tennis match, how many beans are in this jar), coordination problems (people try to coordinate their behavior on a certain matter knowing other people will do the same), and cooperation problems (how can self-interested people work together) (Surowiecki, 2004). Each condition and problem will be discussed in its own chapter.

DIVERSITY

In the context of wisdom of the crowd, we defined diversity as individual members having some private information, on which they base their decision. To put this in the context of crowds, a diverse crowd consists of people having different information and different opinions.

The importance of diversity can be explained through a random episode of almost any police series. To solve any crime, the police uses forensic evidence and motives of the suspects. While forensic evidence always points towards suspect A, the motive will point towards suspect B. In the end, by combining both the forensic evidence and motive, the police finds out the perpetrator is suspect C. What we take from this is that limited information does not paint a full picture. The more pieces of information we gather, the clearer the picture becomes. When we look back at the Bay of Pigs-example, the group of experts had only a part of all total available information, and, since they all had the same information, they had no reason to believe the conclusion they had drawn based upon their information was incorrect. By increasing the diversity of the group, we add information to the group that can be used to come up with a better solution.

The more diverse the crowd, the more private information is added that can be used to create a solution. As Surowiecki states: a person's guess consists of two components, namely knowledge and error. When you ask a large group of diverse people to make a certain prediction and average those predictions, the errors will cancel each other and you will be left with the knowledge part (Surowiecki, 2004). Diversity decreases the error. From this we can derive that diversity on its own is not sufficient. If we ask a diverse group that lacks knowledge to make a certain estimate, we might get a more accurate answers than we would get from the individual members of the group, since we cancel the errors, the answer itself might not be accurate at all as it lacks the knowledge part. For clarification, consider the following example:

We fill a jar with 100 beans and ask three participants to estimate the amount of beans in the jar. All 3 participants estimate the amount of beans to be 100. The average individual error is 0, since all participants have no error, and (0+0+0)/3=0. The crowd's opinion is equal to the average estimate, which is (100+100+100)/3=100, where the crowd's error is equal to 0, since 100-100=0. Say two more people participate, and participant 4 estimates 90 beans, while participant 5 estimates 110 beans. As we can see, the crowd became more diverse. The average individual error became 4, since (0+0+0+10+10)/5=4. This is because the last two participants have an error of 10, as the absolute difference between their prediction and the outcome is 10. The crowd's prediction remains 100, since (100+100+100+90+110)/5=100, and the crowd's error remains 0. As we can see, the crowd's opinion is now smarter than that of its individual members, since the average individual error is higher than the crowd's error. However, diversity did not make the crowd on its own smarter, since the average individual error became 4 after the diversification, while it previously was 0. To conclude, making the crowd more diverse makes the crowd's opinion more accurate than that of its individual members. However, a more diverse crowd does not necessarily have a lower error .

To showcase the importance of diversity in a more economic setting, we looked at research on diversity within top management teams. It showed that diversity will avoid consensus and create debate, which will enhance company performance (Simons, 1999).

In everyday economics diversity plays a vital part. When we look at the early days of the United States car industry, or any other industry for that matter, you will see a business climate with a very large amount of alternatives. For example, in the case of the car industry, there are electronic cars, gasoline powered cars, and steam powered cars. From all these alternatives, the successful ones will be used, while the unsuccessful alternatives disappear. Only by showing as much alternatives as possible we can determine which of them are successful, and which of them are not. As showed, diversity adds a lot of solutions and information. If the group can determine the value of all the provided solutions and information, it will show that diversity is vital for good decision making. (Surowiecki, 2004). In the chapter: Mathematical Explanation of Wisdom of the Crowd the exact mathematical value of diversity will be explained by using the diversity prediction theorem.

In this thesis research will be done on whether or not wisdom of the crowd occurs within a group of high school students predicting the outcomes of the 2014 FIFA World Cup in Brazil. Within this group of high school students sub-groups were created based on the knowledge of the subject, namely soccer. Interesting for research is the value of diversity and knowledge between the high and low information groups. As previously explained, each guess or prediction is based on knowledge and error. From this we assume that the high information group will have more knowledge, but also a higher error due to lower diversity, while the low information group will have less knowledge, but a lower error due to a higher diversity. It will be interesting to see if the value added by diversity will be lower or higher than the value expertise/knowledge will add.

Independence

In the context of wisdom of the crowd, independence has been defined as the individual opinions of the members of a crowd, which are not determined by the opinions of other members of said crowd. Wisdom of the crowd suggests that, if a crowd is wise, it might be smart to follow the crowd's opinion. Thus is seems like wisdom of the crowd and independence are contradictive, as independence suggests that individual opinions should not be determined by opinions of others, and wisdom of the crowd suggest that following the opinion of the crowd can be wise.

The following example explores this idea of wisdom of the crowd and independence being contradictive. There are two restaurants: restaurant A, which is filled with people dining, and restaurant B, where no people are dining. A person who knows nothing of either of the restaurants is asked to choose to dine in one of both restaurants, He or she chooses restaurant A, using the logic that the restaurant where most people are dining is the better restaurant. If all people currently eating in both restaurants each individually chose the restaurant based on which restaurant they believe is better, or to put it differently: if the individual opinions of the people are not determined by the opinions of other people dining, then the crowd would suggest that the restaurant with most people dining is the better restaurant, and it indeed can be wise to follow the crowd. In this case the crowd itself was independent, so the opinion of the crowd is based on an independent crowd and should be a wise opinion (assuming it does not lack diversity, decentralization or aggregation).

If the same example of the two restaurants is looked at, now knowing that, not only the person we previously asked to pick a place to eat used the logic that the restaurant with most customers better restaurant, but all other people who are currently dining also used this same logic. Now it can be determined that this crowd lacks independence, since the individual members of this crowd did not determine their opinion on their own, but used the opinions of the people who chose before them to determine where to eat. Only the first person who chose between the two empty restaurants made an independent decision. However, a crowd of one person clearly lacks diversity, and it could be wise not to follow it. To summarize, for a crowd to be independent it needs a decent amount of its members to make a decision not based on opinions of others.

It can be hard to determine whether an opinion is independent or not. Humans are social creatures, and take into account not only their own needs, but also the needs of others. As Surowiecki stated "the more influence we exert on each other, and the more personal contact they have with each other, the less likely it is that the group's decisions will be wise ones'' (Surowiecki, 2004). For a group to be wise, its individual members should think for themselves.

An important crowd used as reference group in this thesis is called Betting Website. As later will be explained, the opinion of this crowd is determined by certain odds that, in their turn, are determined by the crowd's own bet distribution. The lower the odd of a certain outcome, the more people apparently will bet on this outcome. We can argue that bets placed on these odds are not independent but based on anchoring. Anchoring means that, if you give people a certain number and ask for a prediction, the predictions will be related to that number. Or as Tversky and Kahneman once put it: "different starting points yield different estimates, which are biased towards the initial values.'' (Tversky, 1974)

Consider the following example: we tell a person that there is a chance a certain event A is going to happen. All this person knows of event A is that the odds of it happening are 1.1 (which means that you bet 1 on event A to happen, and if event A happens, you get 1.1 back), and the odds of event A not going to happen are 100 (you bet 1 on event A not to happen, and if event A does not happen, you get 100 back). If we ask this person if he or she thinks event A will happen or not, then there is a big chance he will say event A will happen. The reason behind this is that the payout of when it happens is clearly lower than the payout when is does not happen, suggestion the chance of event A happening is higher than of it not happening. We can say that the odds of this event help determine this person's opinion, and that the opinion is at least partly based on anchoring. The bet distribution that determined the odds –and, with that, the opinion of people on this betting website- is not simply choice based on two clear options as, in the previous example. For each match, the betting website had 20 different possible outcomes with odds related to those outcomes of at least six or higher. We assume that, in this case, anchoring has less or even no effect on how people determine their bets since, by looking at the odds, there is no clear odd that suggests that a certain outcome is most likely to happen.

Decentralization

In the context of wisdom of the crowd, we defined decentralization as the ability for people to specialize and draw on local knowledge. In a broader, more general definition, decentralization means that the power does not fully reside in one central location, making it possible for people to specialize and draw on local knowledge (Surowiecki, 2004).

A great example of decentralization related to wisdom of the crowd is Wikipedia. Wikipedia is an online platform where people can add and adjust information on anything of relevance. At the beginning, Wikipedia was ran by a small amount of admins, but over the last few years the highest number of new pages and adjustments have been made by a large amount of users that each add a small amount of information. (Kittur, 2007). In the case of Wikipedia, a large decentralized crowd creates content, with no one in charge of what content has to be created. Users with knowledge on specific subjects can create content related to those subjects,, and it can be said that the users have the ability to specialize and use their local knowledge to determine the content they create.

Decentralization creates an environment that facilitates diversity and independence. Consider the difference between Microsoft and Linux, for example: Linux is an operation system created by Linus Torvalds, who, after completion, made the operation system an open source and free to use for everyone. The idea behind the open source is that, if a problem occurs, anyone from anywhere can solve it. With Microsoft, on the other hand, people are hired and told to solve certain problems. As result of this, Linux has a very diverse crowd of users, and an unlimited amount of people who can work on a certain problem, where Microsoft is limited in its resources, since it has a limited amount of employees who, in their own turn, have a limited amount of time. As showed, the decentralization of Linux allows a diverse crowd of people to come up with a solution for any given problem that occurs, successfully making Linux the biggest competitor of Microsoft (Surowiecki, 2004).

Aggregation

In the context of wisdom of the crowd, we defined aggregation as a way to transform the individual opinions of the individual members of the crowd to a collective opinion. Where diversity, independence, and decentralization allow individual members of the crowd to create a valued opinion, aggregation is needed to combine those opinions into a representative opinion of the crowd.

In the example of Galton and the country fair, Galton himself collected all the bet slips after the event, to determine the average. In the example of Linux, a few admins determine which of all possible solutions to a certain problem will be implemented, and, meanwhile at Wikipedia, a few admins control all new pages that have been created and all adjustments that have been made. The betting website itself aggregates all the bets that are being placed and adjusts its odds accordingly.

The importance of aggregation is showcased when a situation lacks this aggregation. When we look at the example of the two restaurants, the used form of aggregation is watching through the window to determine where the crowd wants to eat. Now, say that the empty restaurant is not empty, but that the people dining inside are not visible through the window: then maybe the majority of the individuals of the crowd will think the empty looking restaurant is better. However, due to lack of aggregation, the crowd's opinion is interpreted as if the other restaurant is better.

Important to note is that aggregation should not be achieved at the expense of decentralization, diversity, or independence. For a crowd to be wise, its opinion should be aggregated without adjusting the opinions of its individual members. As you can imagine, if your boss asks you to evaluate his performance, the opinion you give might not be independent. This is why in a lot of experiments it is made very clear to participants that they stay anonymous or that their identity is only known to the experimenter to avoid their choices or answers becoming dependent.

Cognition Problems

The crowd's opinion cannot be used to solve all problems. Sometimes the average of the individual members is just the average. When we let 100 people run a marathon, and take their average time, then the time we get is indeed no more than the average. Surowiecki considers three types of problems that can be solved by using wisdom of the crowd. Important to note is that not all questions that categorize under those three types will be solved by wisdom of the crowd. In most of these cases, the problem does not lie within the type of question but with the crowd itself. If we ask a crowd of Feyenoord fans which Dutch soccer club has most supporters, we can expect the average answer will be Feyenoord even though the Club Position Matrix states that Ajax has most fans (Volkskrant 2012) In this case the problem lies not within the question, but within the fact that the crowd lacks diversity and is biased.

The first kind of problem that wisdom of the crowd can help solve we discuss is cognition problems. These are problems that have –or, in the future, will have- a definitive solution, such as the amount of beans in a jar, who will win any game, weight of an ox, or how many items a shop will sell. Not only problems with a definitive answer, but also problems with answers clearly better than others are cognition problems, such as where to open a new restaurant.

One of the reasons behind the fact that wisdom of the crowd can help overcome cognition problems is that individuals will over- and underestimate certain guesses. If we ask a crowd to predict the amount of beans in a jar, there will be a part of the crowd that overestimates the amount, while another part will underestimate the amount of beans in the jar, making the crowd's opinion more accurate on average.

The same logic holds in a more economic environment: in the stock market a part of the stock brokers will overestimate the future price of a stock, while others underestimate the future price of a stock, making the crowd's average accurate. Important to note is that there are individuals who outperform the crowd, giving individuals incentive to participate. However, if multiple situations are considered, it will not be the same individuals that outperform the crowd (Surowiecki, 2004).

Another reason why wisdom of the crowd helps solve cognition problems is that sometimes the majority outperforms the individual. The idea behind the game show *Who Wants to Be a Millionaire*, for instance, was simple: a contestant has to correctly answer 15 multiple choice questions with four possible answers, and he or she would win a million pounds. The contestants had three lifelines he could use: a lifeline where the contestant could ask the studio audience what they thought the correct answer was, the 50/50, where only two answers for that question remained, and ‘call a friend/expert’. Calling this experts would provide the correct answer in 65% of the time, while the studio audience picked the correct answer 91% of the time, showing that, in this case, the group outperformed the individuals. (Surowiecki, 2004).

A more interesting example is Google. Google works the way it works due to the PageRank Algorithm, which -according to Google- “capitalizes on the uniquely democratic characteristic of the web by using its vast link structure as an organizational tool. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. Google assesses a page’s importance by the votes it receives. Google looks at more than the sheer volumes of votes or links; it also analyses the page that casts the vote. Votes cast by pages that are themselves ‘’important’’ weight more heavily and help to make other pages important ’’ (Page L. a., 1998) In other words, Google uses the crowd to determine the results of your search, by using a weighted average of all votes made within the crowd. Looking at the success of Google, we may say that the Algorithm works very well.

Coordination Problems

Coordination problems are problems where people try to coordinate their behavior on a certain matter, knowing other people will do the same. What time should I leave for work? Where do we want to eat tonight ? How can we make sure people get the goods and service they want? All of these are coordination problems. Coordination problems have in common that a person should not only consider what he or she thinks is the correct answer, but also what he or she thinks other people perceive as the correct answer. If you want to know what time you need to go work with your car, then you need to take into consideration how many other people are going to use their car at that time. Heavy traffic suggests you need to leave earlier than when there is no traffic. As we can see, crowds that have coordination problems do not have independent decision making, since the decision of its individual members depends on the action or perceived actions of others. As a result, coordination problems cannot always be solved by wisdom of the crowd, but it is striking how often wisdom of the crowd actually can be used to solve these problems.

To show the value of wisdom of the crowd in coordination problems, an example can be given: in the 1990, the economist Brian Arthur tried to solve the El Farol Problem (Brian, 1994). El Farol was a local bar in Santa Fe where, when it was less than 60% full, everyone would have a great time. However, when it was full for more than 60%, people would not have any fun and could better have stayed home. As we can see, the amount of fun a person has does not only depend on his attendance, but also on the attendance of others. Arthur assumed that people had different strategies to determine the amount of people that were going to the bar, and so determine if they want to go or not. Some would go if they had fun the last time they went (so the bar was filled for less than 60%), and some assumed the same amount of people would go each week, while others used the average over the last few weeks. He showed that no strategy yielded the best result on the long term, as some strategies worked for several weeks and then fell off. However, in the 100 weeks he documented, the average amount of people in the bar was 60% of its full capacity, showing that, even while individual strategies depend on each other's behavior, the groups opinion was accurate (Surowiecki, 2004).

When we go to the supermarket to buy orange juice, most of the time the orange juice is just there in stock for you to buy. A few days before, the orange juice was boxed and shipped to your supermarket, while you nor the orange juice company knew that you were going to buy the juice. Whether or not there is orange juice in stock depends not only on your decision to buy, but also on decisions of others to buy orange juice, thus making it a coordination problem. The orange juice company assumed that a certain amount of people will buy that certain amount of orange juice. If they overstock, it will cost the supermarket money since they have to destroy the orange juice that has reached its expiration date, and if they understock it, they could have sold more orange juice and they will get disappointed customers. Even while not always the same customers buy the orange juice, they can use the average of the crowd to accurately predict the amount of orange juice to stock. Wisdom of the crowd provides a very accurate answer on this coordination problem (Surowiecki, 2004).

Where cognition problems have a definitive answer, coordination problems do not have to have one, since the correct answer depends on the behavior of others, which is not always same. Crowds can be used to predict their own behavior and to come up with a solution related to that predicted behavior.

Cooperation Problems

Coordination and cooperation problems have similar features, as in both problems people have to take the behavior of others into account. While for coordination problems people can pursue their own interest to come up with a good solution pursuing one's own interest will be at the expense of others with cooperation problems.

at the following example is that of the prisoners dilemma. In this dilemma, two criminals are caught and locked in separate interrogation rooms. The police does not have enough evidence to put them in jail for the crime they committed, and needs both criminals to tell on each other. If both prisoners do not tell on each other, they get a prison sentence of five years. However, if one of the prisoners tells on the other while the other stays silent, then the one who tells goes free while the other goes to jail for 20 years. If both prisoners tell on each other they both go to jail for 10 years. As this problems shows, pursuing their own interest goes at the expense of the other. This is a great example to show that not all cooperation problems can be solved by wisdom of the crowd. Although the crowd's optima is for both to not tell on each other and get a five-year sentence, the dominated strategy is actually to tell on each other and get a 10-year sentence.

Getting people to pay their taxes is also a cooperation problem, since a person is best off if everyone pays their taxes except himself. In that case, this person gets the benefits of taxes (police force, infrastructure) without paying for it. Wisdom of the crowd can help solve this problem by looking at the situation in which the majority of people pay their taxes. Some people will always pay their taxes, as they are selfless, while others pay their taxes due to the fear of punishment for not paying their taxes. With this knowledge, governments can calculate the amount of effort they need to put into punishment of tax evaders to make the majority pay their taxes.

Betting Websites

The Oxford Dictionary of English defines betting as follows: "the action of betting money on the outcome of a race, game, or other unpredictable event". (Stevenson, 2010) A betting website or bookmaker provides the odds for those races, games, and outcomes. (Spann, 2008)distinguishes two types of odds: odds that adjust over time provided by prediction markets, and fixed odds provided by bookmakers. Prediction markets are places where buyers and sellers trade virtual stocks that represent future market situations, namely outcomes of sporting events. The payoff or -in stock market terms- the dividend is determined by the outcome of the sport event. The price of each virtual stock represents the expected value of the outcome. Fixed betting odds are predetermined odds that determine the payout in case the event happens.

Betting websites have three ways to make money. To clarify the three possible ways, the way of determining betting odds shall first be discussed.

$$E(M) = 1 - \sum_{i=1}^{n} P_i \cdot w_i \cdot d_i$$

The expected value of a bet (M) is determined by the probability of the outcome (Pi), multiplied with the percentage of bets on the outcome (Wi), multiplied with the odds for that outcome (Di) (Levitt, 2014).

To clarify this, the following example can be given: Nadal plays against Djokovic on Roland Garros. The betting website knows the probabilities of each player to win, and also know the percentages of bets on each player. For simplification we shall assume the bet distribution between the players and the chance to win for both players are both 50%. The odds related to a 50% chance of winning or related to a bet distribution of 50% on both players is 1.85. This means that, when Nadal wins, each dollar bet on Nadal gets a 1.85 dollar return, while, when Djokovic wins, each dollar bet on Djokovic gets a 1.85 dollar return. Using the above stated formula, the expected value of a bet on this match is 0.5 (chance of Nadal winning) * 0.5 (percentage of people betting on Nadal) * 1.85 (odds if Nadal wins) + 0.5 (chance of Djokovic winning) * 0.5 (percentage of people betting on Djokovic) * 1.85 (odds Djokovic winning) = 0.925. This shows that, for every dollar bet on this event, the expected value of that bet is 92.5 cents. The difference between 1 (price to participate) and the expected value of the bet is the expected value for the betting website. Since for every dollar bet the expected payout is 92.5 cents, this leaves the expected profit for the website to be 7.5 cents. It is important to state that this is a simplified example, as in almost all situations neither the bet distribution nor the chance of each player to win is known exactly.

A closer look at the ways for betting website to make money needs to be taken. The first way for betting websites to make money is if the website knows the probabilities of the outcome, but not the bet distribution, like in a game of roulette. Roulette is a gambling game in which players predict on which slot on a rotating disk a small ball win end. There are 37 possible slot outcomes between the numbers 0 and 36. When predicting the correct outcome, the payout is 36. So correctly predicting on which number the small balls ends grants you 36 times your stake. Since the odds of correctly predicting an outcome are equal to 1 in 37, the expected value of playing roulette are 0.97, since 1/37 (chance of predicting the correct number) * 36 (odds when correctly prediction the correct number) = 0.97. This means that, for every dollar bet on predicting the correct number at roulette, you get an expected value of 97 cents. The difference between 1 (price to participate) and the expected value of the bet is the expected profit for the betting website, which is 3 cents in this case.

The second way for betting websites to make money is if the website does not know the probabilities of the outcome, but knows the distribution in which its bettors are going to bet. Considering the previous example, the tennis match Nadal versus Djokovic, we then assume that the betting websites exactly knows the percentages of bets placed on Nadal and Djokovic before the bets are actually placed. It might sounds unlikely that betting website can predict the distribution of bets made on each player, but previous results, current form, and underground they play on can be used to accurately predict how people are going to bet, for instance. In our previous example we used that both players get 50% of the bets for simplification. The betting website will set the odds in the way that the percentages of bets multiplied with the odds have an expected value lower than 1. As showed above, this is because the difference between 1 and the expected value of the bet is the expected profit for the betting website. In the 50% example 1.85 was chosen, as 0.5 *1.85 equals 0.925. This shows that, for each betted dollar, the expected return for tor the betting website is 7.75 cents.

The third way for betting websites to make money is if the website knows the probabilities of the outcomes and the distribution in which its bettors are going to bet. As previously mentioned, the expected value is determined by the probability of the outcome multiplied with the percentage of bets on the outcome, multiplied with the odds for that outcome. If both the percentage of bets on the outcome and the probability of the outcome are known, the websites can adjust the odds to change the expected value. This is valuable since the difference between 1 (price to participate) and the expected value of the bet are the expected returns for the betting websites. Adjusting the expected value for of each bet can increase or decrease the expected returns for the betting website. Assuming the betting website knows that for a certain event, 80% of bets is placed on outcome Y to happen and 20% on outcome Y not to happen. It can decrease the expected value of outcome Y by adjusting the odds and gaining a larger average return, called the vig. For this method to work, the betting website does not need to know the probabilities of the outcome exactly, but it just needs to know it better than the bettors do (Levitt, 2014).

Important to note is that, in the long run, betting websites only need to know either the distribution of bets or the probabilities of the outcomes to make profit. This is clarified by the following example: Murray plays against Federer on Rolland Garros. The bet distribution is as follows: 40% of the bets are placed on Murray to win, and 60% of the bets are placed on Federer to win. The odds for Murray to win are, for example, 2.3 (since the expected value will be 0.4 * 2.3 = < 1), while the odds for Federer to win are, for example, 1.5 (since the expected value will be 0.6 *1.5 = <1). Regardless of the chances for both players to win, in the long run the betting website will make a profit. To clarify: if Murray has a X chance to win, and Federer has a 1-X chance to win, the expected value will be X (chance of Murray to win) * 0.4 (amount of bets played on Murray) * 2.3 (odds for Murray to win) + (1-X) (chance of Federer to win) * 0.6 (amount of bets placed on Federer) * 1.5 (odds for Federer to win) < 1. We can substitute the distribution of bets to X and 1-X and the probabilities of the outcome to 0.4 and 0.6, and the formula still holds to show that only one of the two factors needs to be known. To summarize: as long as one of the two factors is known, the odds can be adjusted to the known factor and the expected value can be set in a way that it is lower than 1, creating an expected profit for the betting website.

In most cases the websites only approximately know approximate chances of each player to win and the approximate bet distribution. The less that is known of both factors, the bigger the vig (difference between the price to participate and expected value of the bet) created by the betting websites will be. In roulette the exact chances of winning are known, so the websites can afford to create a low vig (as previously shown only 3 cents) since they know that, in the long run, they will make a profit. The vig for sport events is larger since the chance of each player winning is not exactly known. They create a larger vig to ensure themselves of a profit in case they incorrectly predicted the bet distribution or the chance of either player winning.

The fact that betting websites can predict outcomes better than bettors, can be explained by certain biases. For example, the favorite-longshot bias and home bias. Those biases show that bettors over- or underestimate certain effects. For example, the favorite-longshot bias, that shows favorite bets (bets with low odds) outperform longshot bets (bets with high odds), but that bettors still prefer longshots over favorites. In certain markets the vice versa effect is found. The home bias is that people overestimate the chances of the home team (Levitt, 2014) (Cain, 2000) (Vlastakis, 2009) (Andrikogiannopoulou, 2011). As previously stated, betting website do not need to correctly predict the outcomes: they just have to predict them better than their bettors. This can be illustrated by the favorite-longshot bias. For example the odds of X to happen are 1.5, while the odds for X not to happen are 300. However, knowing that players have a favorite-longshot bias, betting websites can place the odds for X to happen on 2 knowing bettors favorite the longshot, and put the odds of X not to happen on 100, knowing bettors will still bet on it since they favor the longshot. This shows that, if betting websites can properly predict the betting distributions of certain events, they can adjust the odds to decrease the expected value of the bet and increase the betting website's return.

Betting odds on websites change overtime, which is due to the fact that betting websites do not know the chances of the possible outcomes, but know the bet distribution. The determination of the odds works the same as the price of stock in the stock market: the higher the demand, the higher the price of the stock, or, in betting terms, the more bettors bet on a certain outcome, the lower the odds and possible payout of that outcome becomes. Betting websites adjust their odds to make sure they always make a profit. The possible payout of each outcome will always be lower than the total amount of money placed on all outcomes. the simplified example of Nadal versus Djokovic can illustrate this, with both having odds of 1.85. If, for example, 60% of the bets are placed on Nadal, and Nadal wins, the payout the website will have to make will be 1.11 since 0.6 * 1.85 = 1.11.Tthis means that the website would have a loss of 0.11. However, if the website lowers the odds for Nadal to win for example to 1.5 when they notice the bet distribution becomes 60/40 in Nadal's favor, the website will only have a payout of 0.9 and will have a profit of 0.1. Using this method it is irrelevant for betting websites who wins as they will make a small profit for each bet placed since the total amount of bets placed on all outcomes will be larger than the possible payout of the individual outcomes.

The betting website is used as a reference group in our experiment. We use it to test if the results of the experiment hold any real-life value. For this we need to verify that the results of the betting website indeed represent the opinion of its bettors. The results of the betting website are the predictions of the website compared to the actual results, whereas the predictions of the website are the odds transformed to predictions. As previously explained, betting websites work partly like stock markets. To determine if the results reflect the opinion of its bettors, we look at the market efficiency. If it is efficient, supply and demand meets one another to determine its price (odds). If it is not efficient, the predictions by the betting website do not have to represent the actual opinion of its bettors. We look at two factors that influence the market efficiency: arbitrage opportunities, and predictive power of the website compared to its bettors. Arbitrage opportunities are possibilities of placing non-risk bets that always yield a positive expected value. For example: if Federer plays against Nadal, and the odds on one website for Federer winning are 2.1, and on another website the odds for Nadal winning are 2.1, than it is possible for bettors to distribute their bets in a way they create a risk-free bet that always yields a positive expected value. (Vlastakis, 2009)provided evidence that highly profitable arbitrage opportunities are rare but possible, becoming rarer in more recent periods studied. When only considering the online bookmakers, the arbitrage opportunities become even rarer. As previously stated, betting websites can predict the outcomes better than the average bettors. However, they can only adjust the odds by a small margin, because otherwise bettors might recognize and exploit mispricing by the bookmakers (Levitt, 2014).Considering these findings, we assume that the results of the betting website accurately represent the opinion of its bettors, and that it can be used to test if the results of the experiment hold any real-life value.

Mathematical Explanation of Wisdom of the Crowd

The Beans-in-a-Jar Experiment, Google, and estimating the weight of an ox at the country fair are all examples where wisdom of the crowd occurs. Those examples show that the opinion of the crowd can be more accurate than that of its individual members. Some of those crowd's opinions were even more accurate than any opinion of the individual members of the crowd.

If you live in a big city and want to know if it is going to rain, you can turn on the TV and watch the weather forecast, which will most likely provide you with an accurate answer. You can also look outside and see if the people passing by have an umbrella with them, which will most likely provide the same answer as watching the weather forecast. This is another example of a wise crowd. However, not all crowds are wise. In America there was a plank road boom between 1847 and 1853. Due to the poor conditions of the roads, people sought economically beneficial solutions to improve the quality of the roads. When the weather was poor, a lot of small villages were not accessible due to the fact the roads to the villages were flooded and could not be used. This made trading with those villages impossible when poor weather occurred. By putting wooden planks on the low quality roads, the villages were accessible even with poor weather, which made trading with these villages always a possibility, and would drastically improve their economics. The plank roads were relative cheap, so affordable for small communities, making them seem like the perfect solution. All over America plank roads emerged and expended rapidly since its benefits could easily be seen, and people assumed that, if other places have them, they should be beneficial, otherwise those others places would not have taken them in the first place. However, to be economically beneficial, the roads needed to last way longer than they did. Sometimes the roads needed to be replaced within a few years, while the prospect was that they would last at least twice as long. Making them now economically not viable, the roads disappeared as fast as they came (Majewski, 1993).

As shown in the previous examples, it is sometimes wise to follow the crowd, while other times it might not be. In those two examples a clear answer can be given, and in the case of the plank roads the answer was clearly that it was very unwise.

As previously stated by Surowiecki, certain requirements must be met before a crowd can become wise. Often those requirements are partly met, which raises the question if in those situations a crowd is wise or not. It is clear that some of those questions have no clear yes or no answer, while the crowd's opinion might help solving problems in those situations. This raised the question if there could be a mathematical method to conclude if the wisdom of the crowd occurs. In 2007 Page created a mathematical theorem that explains the phenomenon of wisdom of the crowd. He called it the Diversity Prediction Theorem. The theorem shows that diversity within a crowd explains the difference between the average individual error (the average individual opinion) and the collective error (the opinion of the crowd). The Diversity Prediction theorem states:

$$collective\ error = average\ individual\ error - prediction\ diversity$$

The theorem shows that the collective error is always equal or lower than the average individual error, making the opinion of the crowd always equal to or better than that of its average individual member. This is due to the diversity between the members of the crowd. The following example helps illustrate the Diversity Prediction Theorem. Two imaginary Beans-in-a-Jar Experiments, both containing six beans, are used as an example to prove the theorem.

	Alexander	Peter	Margo	Crowd's opinion (collective error)	Average individual error	Outcome
Predictions amount of beans in the jar 1	3	6	9	6		6
Squared error jar 1	9	0	9	0	6	
Predictions amount of beans in the jar 2	9	6	3	6		6
Squared error jar 2	9	0	9	0	6	

Important to state is that the Diversity Prediction Theorem is based on squared errors. This avoids positive and negative errors cancelling one another, which would occur when a person is equally likely to overestimate or underestimate an outcome. This person would, in that case, make zero errors in total. In our example, this is represented by both Alexander and Margo. For the first jar Alexander underestimated the amount of beans by 3, where for the second jar he overestimated the amount of beans by 3. For Margo this is vice versa. If not the squared error but just the error was used, the error for Alexander at the first jar would be -3, since 3-6=-3. For the second jar the error would be 3, since 6-3=3. Alexander his total error would have been -

3+3=0. When using squared errors Alexander's error for the first jar is $(3-6)^2=9$. His error for the second jar is $(6-3)^2=9$. Now Alexander has a total error of 9+9=18.

The crowd's opinion is formed by adding all predictions and dividing them by the amount of predictions made. The amount of predictions in our example is 3.

The Diversity Prediction Theorem shows that the average squared individual error for the first jar is 6, while the collective error is 0. The difference is explained by the prediction diversity. The prediction diversity shows the value of diversity is equal to the variance of the opinions of the people in the crowd. The variance is equal to the sum of squared differences between the predictions and the outcome divided by the amount of predictions. In our example this is $((6-3)^2+(6-6)^2+(9-6)^2)/3=6$. The theorem shows that the higher the diversity, the higher the variance is, and the wiser the crowd becomes compared to its average individual members.

To show the value of the Diversity Prediction Theorem, we provided a real life example. For American Football fans one of the most important days of the year is draft day. On draft day American football teams can pick former college football players to play for their respective teams. Since the teams pick in turns, it is important for teams to know what other teams want to pick. To clarify, the team that picks first can select any player it wants, the second team that picks can select any player it wants except the previously picked player, the third team can select any player except the two players previously selected, and so on. For this reason analyst or experts are working day and night before draft day to most accurately predict the draft. (Page S. E., 2007) used those predictions of the experts to provide proof of wisdom of the crowd applied on a crowd of experts. This is represented in the following table 1.

Experts' Predictions of 2005 NFL Draft

Player\Expert	Wright	Adler	Fanball	SNews	Zimm	Prisco	Judge	Crowd
Smith	1	1	1	1	1	1	1	1.0
Brown	2	2	4	2	2	5	2	2.7
Edwards	3	3	2	7	3	2	3	3.3
Benson	4	4	13	4	8	4	8	5.9
Williams	8	5	5	5	4	13	4	6.4
Jones	16	9	6	8	6	6	9	8.1
Williamson	13	14	12	12	13	7	7	9.7
Rolle	6	6	8	10	9	8	6	7.9
Rogers	9	8	9	9	16	9	9	9.9
Williams	7	7	7	6	7	12	12	8.0
Ware	11	15	14	24	11	11	13	13.9
Merriman	12	11	3	11	12	10	11	10.1
Sq Error	158	89	210	235	112	82	75	34.4

table 1

Table 1 shows that the crowd's error is even lower than that of the best expert, with a collective error of 34.4, average individual error of 137.3, and a prediction diversity of 102.9.

The Diversity Prediction Theorem shows that the collective error (the crowd's opinion) is lower than the average individual error (opinion of the average individual of the crowd) due to prediction diversity. From this we can deduce certain limitations.

The first limitation is that the diversity prediction theorem only states something about the wisdom of the crowd compared to its own members, since its value is based on the average error of its members and the diversity between its members. The more diverse the crowd, the higher the prediction diversity and the higher the difference between the collective error and the average individual error. This only shows that the crowd's opinion is wiser than the opinion of its average member, but gives no real value on the opinion compared, for example, with the correct prediction. If we would ask elementary school kids to predict the NFL draft, we probably get a very high prediction diversity since there is a high chance the school kids would give very different answers. However, the average individual error will probably be very high, and the collective error, while lower than the average individual error, will probably still be very high. If we compare the collective error of a crowd of kids to the average individual error of our experts, there is a very high chances that experts predict way more accurately than the crowd of kids. When we want to check if the crowd's opinion holds any real-life value, it has to be compared to the opinion of other crowds or experts.

The second limitation is that, since the model is based on error, events must be measurable in terms of error for the model to work. This is easy when we ask people the amount of beans in a jar, the NFL draft order, or the outcome of a horse race. Difficulties arise when it becomes harder to calculate an error, for example for the question if 9/11 or the challenger crash could have been avoided.

The third limitation is that the Diversity Prediction Theorem contains the term average individual error. Because of this, it is important that members of the crowd try to make correct predictions. Giving extremely wrong predictions (on purpose) has a big effect on the average individual error. For example, if we go back to the Beans-in-a-Jar Experiment, if one of the participants jokes and gives an answer 1.000.000 times higher than the actual outcome, no matter what the other participants answer, the crowd's opinion will be most like inaccurate. From this we can conclude that people should have the right incentive or get the right incentive to participate. Experts in the NFL draft are judged on their picks, so already have the right incentive, but for the Beans-in-a-Jar example a financial incentive for the correct prediction could help avoid intentional extremely incorrect answers. Rules can be made to exclude those intentional extremely incorrect answers: however, those rules need to be carefully watched to make sure they only exclude the intentional extremely incorrect answers and do not decrease the prediction diversity.

The fourth limitation of the model is that it gives a more accurate prediction than the average individual prediction. However, it does not have to have a higher chance of giving the exact correct prediction compared to each individual of the crowd. Most of the time it even has a lower chance of giving the exact correct prediction. To clarify, when we look at the NFL draft of 2005, we see that the crowd's prediction for Williamson is 9.7. We know for sure that Williamson will not drafted as 9.7^{th} since that is not possible. So even while the prediction of the crowd is more accurate than that of most experts, since only Prisco and judge had a better prediction, it is still incorrect.

The fifth and last limitation has most influence on this thesis. The Prediction Diversity Theorem compares the crowd's opinion to the opinion of the average member of the crowd. In formula terms: the Prediction Diversity Theorem compares the error of the crowd to the average individual error. The possibility exists that the average individual error does not properly represent the opinion of the average member of the crowd.

The following imaginary Beans-in-a-Jar Experiment helps to clarify this limitation. In the example the amount of beans is 6.

	Alexander	Peter	Margo	Crowd's opinion (collective error)	Average individual error	Outcome
Predictions amount of beans in the jar 3	6	6	9	7		6
Error jar 3	0	0	9	1	3	

Looking at the difference between the collective error of 1, and the average individual error of 3, it can be concluded that, if we want to obtain a lower error, we should follow the opinion of the crowd over its average individual members. Nevertheless, by ranking the errors it can be seen that if the crowd is followed, resulting in an error of 1, there would have been a 66% chance to get an even lower error by just following a random member of the crowd. This is because the error of Alexander and Peter is 0, and lower than that of the crowd. Depending on the question, it can be determined if wisdom of the crowd occurs. Only looking at the Prediction Diversity Theorem, one should always want to follow the crowd, since the crowd's opinion is always more accurate than that of its average individuals. However, if the question for example would be: "does following the crowd grant a higher probability of getting a lower error compared to following a random individual in the crowd", arguments can be found that suggest wisdom of the crowd does not occur in this example.

This limitation can be overcome by looking at rank of the collective error compared to the individual errors. However, the rank of the collective error compared to the individual error does not paint the full picture. The collective error could have a lower rank than most of the individual errors, but the absolute difference between the errors might be insignificant. By testing both the rank and the absolute difference, it can be concluded if the crowd's opinion holds real-life value compared to its individual members.

Hypothesis

As previously stated, Surowiecki claimed that, whenever the crowd satisfies the four conditions that characterize a wise crowd (diversity of opinion, independence, decentralization and aggregation), a crowd's judgement should likely be accurate. In this paper an environment is created where those conditions are satisfied.

In the chapter The Mathematical Explanation of Wisdom of the Crowd the Diversity Prediction Theorem is explained. Since the created environment satisfies the four conditions, wisdom of the crowd should be present in our experiment, and from this the first hypothesis is created.

Hypothesis 1: wisdom of the crowd is present when a crowd is asked to predict the outcome of the 2014 FIFA World Cup scores.

In the experiment a high information and a low information group were created. Since the high information group had more knowledge of soccer compared to the low information group, we predict that the high information group has a lower average individual error than the low information group. From this the second hypothesis is created.

Hypothesis 2:the average individual error of the high information group is significantly lower than that of the low information group.

We predicted that the individual members of the high information group predict more accurately than the individual members of the low information group. This creates the assumption that the prediction diversity of the high information group should be lower than that of the low information group. This is based on the fact that more accurate predictions mean lower variance, which entails a lower prediction diversity. From this the third hypothesis is created.

Hypothesis 3: the prediction diversity of the low information group is higher than the prediction diversity of the high information group.

The value of the opinion of the crowd tested in our created environment should be compared to a crowd that holds real-life value. From this we can conclude if the opinion of the crowd from our created environment holds any real-life value. The possibility exists that the opinion of the crowd might have a higher predictive value than that of its average members without holding real-life value. By comparing this to a real life opinion we can test this. From this we derive our fourth hypothesis.

Hypothesis 4: the predictive value of the crowd should not deviate significantly from the control crowd, to verify it has predictive value.

Experiment

To test the previously stated hypotheses, an experiment is conducted that meets said previously stated conditions. A questionnaire was conducted among 124 students of the Minkema College in Woerden. The students are all between the age of 15 and 18 and are following a HAVO or VWO education. The first part of the questionnaire consisted of a survey, which in its own turn consisted of 15 soccer related questions. The survey's questions are related to the top 4 countries on the EUFA ranking: the Spanish Primaire division, the German Bundesliga, the English Premier League, and the Italian Serie A, plus the Dutch Eredivisie, and the survey also has a few questions related to the Champions League and the Europe League.

The survey was designed to divide the subjects into groups based on low and high soccer knowledge. In a trial run conducted among 10 subjects of whom the level of soccer knowledge was known beforehand, the subjects were compared to one another. The subjects with high soccer knowledge scored above 12 correct answers, while the subjects with low soccer knowledge scored below 3 correct answers. This gives no reason to assume that the outcome of the survey does not reflects actual knowledge on the subject.

The second part of the questionnaire consist of a prediction sheet of the 48 first-round games of the 2014 FIFA World Cup in Brazil. Subjects were asked to fill in their exact prediction of the 48 first-round games of the World Cup. The subjects were told that the subject with the most correct predictions would win 30 euros, which is used as incentive to fill in the questionnaire seriously. Whenever the survey or prediction sheet showed the subject did not fill in the questionnaire seriously, the subject was removed. In total 5 subjects were removed from the research, showing predictions with results 300% higher than the average result of the other subjects, such as estimated outcomes between 6-6 and 10-10 for all 48 matches. Some also responded with made up answers on the survey which showed they did not take it seriously.

correct answers	subjects
0	16
1	10
2	11
3	13
4	12
5	10
6	4
7	6
8	2
9	2
10	2
11	1
12	5
13	7
14	9
15	9

Groups

The 119 subjects could be divided in 16 groups according to the answers given in the first part, ranking from 0 to 15 correct answers. Since a lot of groups consist of a low amount of subjects, the following three groups were created. The first group consists of all 119 students, and this group will be referred to as the crowd. The second group consist of 37 students who gave 0,1 or 2 correct answers, and this group will be referred to as the low knowledge group. The third group consist of 30 students who gave 12, 13, 14, or 15 correct answers: this group will be referred to as the high knowledge group. From the website www.bwin.com the outcomes and matching quotes to those outcomes were taken and reformed into predictions. When referring to the group betting website, those predictions are being referred to.

DATA ANALYSIS.

To test for wisdom of the crowd, we want to use the Prediction Diversity Theorem. To do so, the predicted outcomes and actual outcomes should be compared to one another to create an outcome in the value of an error. When we look at the NLF draft outcomes and predictions, they are transformed into errors by comparing the predictions and outcomes and square the difference. When we want to do the same for the 2014 World Cup predictions and outcomes, the predictions and outcomes should be transformed to rankings first. The following formula was created to transforms them into an error term based on rankings.

$$((predicted\ goals\ home\ team - predicted\ goals\ away\ team) - (actual\ goals\ home\ team - actual\ goals\ away\ team))^{2}$$

The clarify, whenever a player predicts a draw, the ranking becomes 0 since 0-0, 1-1, 2-2 are all equal 0. Whenever a player predicts a win for the home team, the ranking becomes positive since 1-0, 2-1, 3-2 are all >0. Whenever a player predicts a win for the away team the ranking becomes negative since 0-1, 1-2, 2-3 are all <0.

When we subtract the actual ranks from the predicted ranks, we get the error, which we then can square to avoid the sum of errors to cancel one another out. Now we can use the Diversity Prediction Theorem to prove there is wisdom of the crowd.

Using this method has certain limitations.

- More the one outcome can hold the same rank, for example 0-0 and 2-2. The predictions 3-3 and actual outcome 1-1 are seen as the same in this model since they both hold rank 0. This is incorrect since a person who predicts 1-1 with the outcome 1-1 should have a lower error than a person who predicts 3-3 with the outcome 1-1.
- Since the error is based on the differences between ranks, the error for prediction rank 1 and outcome rank 3 is equal to prediction rank 1 and outcome rank -1. In outcome terms, a 1-0 prediction with the outcome 3-0 created the same error as a 1-0 prediction with the outcome 0-1. This means that a person who predicts correctly which team wins, but incorrectly predicts the goal difference has the same error as a person who incorrectly predicts which team wins with the same incorrect goal difference. Logically those errors should differ.

A new formula that can be used if the previously two limitations occur has been created. We tested if the results from the new formula and the formula based on rankings differ. If they do not differ, we can assume the two limitations do not occur in our data and we can use the formula based on the rankings. If the predictions of the new method differ from the formula based on rankings then we will use the new formula to analyze the data. To compare and analyze the predicted results with the actual results, a formula has been created that formulates an error term. To overcome the above stated limitations, the formula needs to meet certain requirements.

- The error term for prediction 3-1 and outcome 1-3 should be equal to prediction 1-3 and outcome 3-1
- The error term must increase with an increasing goal difference between predicted outcome and actual outcome.
- The error term of a predicted win with outcome draw with a goal difference Z > the error term of a predicted win with outcome win with a goal difference Z
- The error term of a predicted win with outcome draw with a goal difference Z < the error term with a predicted win with the outcome lose with a goal difference Z

The formula is still based on the formula used to compare predictions and actual ranks during the NFL draft in 2004. The following formula is created:

$$0{,}5 * (|predicted\ goals\ home\ team - \ actual\ goals\ home\ team|)^2 + 0{,}5 * \left(\left|\begin{matrix} predicted\ goals\ away\ team - actual\ goals\ away \\ team \end{matrix}\right|\right)^2 + Y$$

.

Y is the penalty for incorrectly predicting the outcome in terms of win/draw/lose, where Y is 0 whenever the predicted result in terms of win/draw/lose is correct, for example if a prediction is 2-2 and the actual result is 3-3. Y will be 2 whenever the predicted results in terms of win/draw/lose are incorrect, but only minimally incorrect, for example the prediction is a draw and the actual result is a win or lose or vice versa. Whenever the prediction is a win and the actual result is a lose or vice versa then Y will become 4.

Important to note is that, with the ranked based formula, the prediction diversity is completely based on the variance, and the prediction diversity increases with the variance. With the new formula this is not the case. The prediction diversity is still partly based on the variance, since a part of the error term is determined by the goal differences. The larger the goal differences, the larger the variance becomes, which increases the prediction diversity. Another part of the prediction diversity is based on the penalty: it could be possible that the penalty could decrease the prediction diversity. For example, 60% of the individuals predicted a draw, but 40% predicted a clear win for a certain team, then there is a possibility that the crowd's opinion is a win for that certain team. If in this scenario the game ends in a draw, then prediction diversity decreased due to the penalty (since the crowd gets a penalty of 2 while the average individual error only gets a penalty of 0.4 * 2 = 0.8)

Since the predicted outcomes of the participants are clean predictions like 1-1, 2-1, 3-2 the formula can be used to create an error term for each participant. However for the group predictions those results are not clean numbers, for example 1.643 – 1.235. To determine which predictions count as win/draw/loses the following formula is created.

$$predicted\ goal\ difference > 1 = home\ team\ win$$

$$predicted\ goal\ difference < -1 = away\ team\ win$$

$$predicted\ goals\ home\ team > 1.5 * predicted\ goals\ away\ team = home\ team\ win$$

$$predicted\ goals\ away\ team > 1.5 * predicted\ goals\ home\ team = away\ team\ win$$

If none of the following formulas is satisfied, the outcome is seen as a draw.

The reasoning behind the formula is as follows: for a team to win, it should score significantly more goals than the other team. On individual predictions this is clear, since the predicted winning team always scores at least 1 goal more than the opposed team, which is clearly significant. However, we can argue that, for example, the group-predicted score of 1.643 – 1.235 does not significantly differ from each other. We ca not use an absolute difference lower than 1 to determine if a team scores significantly more than the other team. We illustrate this by using an absolute difference of 0.6. A prediction of 1.8 -1.2 might suggest that the team that scores 1.8 scores significantly more than the team that scores 1.2 times since, if we round the numbers, it suggests a 2-1 score. However, that same absolute difference suggests less of a winner when we look at a prediction of 2.3 and 1.7 since, if we round the numbers, it suggests a 2-2 draw. This immediately shows that we also cannot round numbers to clarify which team wins, since a prediction of 2.6-2.4 would suggest a 3-2 prediction, while we can argue that the group does not significantly predict a winner in this case. Since most goal predictions range between 1 and 3, goals the following formula was created:

$$predicted\ goals\ home\ team > 1.5 * predicted\ goals\ away\ team = home\ team\ win$$

and

$$predicted\ goals\ away\ team > 1.5 * predicted\ goals\ home\ team = away\ team\ win$$

This transforms scores like 1.81-1.20 into wins for the 1.8, while it transforms scores like 2.3-1.7 into draws, which fits our data.

The control crowd used in hypothesis 4 is the group betting website. The formula used to transform the betting odds into probabilities is (1/ odd) / (the sum of all 1/every odd). For example if the odds of a certain event happening are 2.3, and the odds of that certain event not happening are 1.5, then the chance of that event happening according to the odds is (1/2.3) / (1/2.3 + 1/1.5) = 39.5 %. When the probabilities are multiplied with the outcomes related to those odds, the sum of those outcomes is the predicted result by the betting website. The same rules apply for the betting websites as for the groups to determine if they predict a win/draw/lose.

To clarify, the odds of the game Brazil vs Croatia have been taken and transformed into the predicted results of 2.301539 – 0.701223, which is equal to a predicted victory for Brazil.

Brazil vs. Croatia		Betting odds			1/betting odd		Percentage /100		Home	Away
1	0	5,8			0,1724138		0,125178		0,125178	0
2	0	6			0,1666667		0,121006		0,242011	0
2	1	9			0,1111111		0,08067		0,161341	0,08067
3	0	8			0,125		0,090754		0,272262	0
3	1	13			0,0769231		0,055849		0,167546	0,055849
3	2	31			0,0322581		0,02342		0,070261	0,046841
4	0	12			0,0833333		0,060503		0,242011	0
4	1	21			0,047619		0,034573		0,138292	0,034573
4	2	56			0,0178571		0,012965		0,05186	0,02593
4	3	161			0,0062112		0,00451		0,018038	0,013529
5	0	21			0,047619		0,034573		0,172865	0
5	1	37			0,027027		0,019623		0,098113	0,019623
5	2	101			0,009901		0,007188		0,035942	0,014377
6	0	41			0,0243902		0,017708		0,106249	0
6	1	76			0,0131579		0,009553		0,057318	0,009553
6	2	226			0,0044248		0,003213		0,019275	0,006425
7	0	91			0,010989		0,007978		0,055849	0
0	1	161			0,0062112		0,00451		0	0,00451
8	0	226			0,0044248		0,003213		0,0257	0
0	0	14			0,0714286		0,05186		0	0
1	1	9			0,1111111		0,08067		0,08067	0,08067
2	2	21			0,047619		0,034573		0,069146	0,069146
3	3	81			0,0123457		0,008963		0,02689	0,02689
0	1	20			0,05		0,036302		0	0,036302
0	2	51			0,0196078		0,014236		0	0,028472
1	2	27			0,037037		0,02689		0,02689	0,05378
0	3	151			0,0066225		0,004808		0	0,014424
1	3	81			0,0123457		0,008963		0,008963	0,02689
2	3	71			0,0140845		0,010226		0,020452	0,030677
1	4	276			0,0036232		0,002631		0,002631	0,010522

2	4	251			0,0039841		0,002893		0,005785	0,01157
					1,3773476		1		**2,301539**	**0,701223**
					sum of 1/odds		total percentage/100		goals	

Table 2

As we know, odds are determined as follows: the higher the odds, the less likely the event occurs and the less people have bet on the odd. The more people bet on a certain result, the lower the odd gets. By taking the sum of 1 divided by each odd, we get the weight of each individual odd. If we look at table 2, the odd 5.8 for the predicted result 1-0 has a weight of 12.5178%. This means that, from all the bets placed, 12.5178 % thought 1-0 was going to be the result. This suggests that 12.5178% thinks Brazil will score once, which is equal to 0.125178. The odd 9 for the prediction 2-1 has a weight of 8.067%. This means that 8.067% of the people think Brazil will score twice and Croatia will score once, which equals 0.161341 goals for Brazil and 0.08067 goals for Croatia. If we transform all the odds into weights, and multiply the weight with their prediction and take the sum of the results, we get the predicted results of the betting website, in this case 2.301539 goals scored by Brazil and 0.701223 goals scored by Croatia.

RESULTS

As stated in chapter X: Data Analysis, to determine if we can use the ranking based formula we must compare the results of the ranking based formula and the formula used to overcome the two limitations that might occur using the ranking based formula.

To determine if the results of the formulas differ, we use a Wilcoxon sign ranked test to compare the groups -crowd, low knowledge group, high knowledge group, and betting website- for both formulas. If the difference or lack of difference between groups is equal for both formulas, we can conclude that the results for both formulas are equal, and we can use the ranked based formula; if not, we can only use the new formula to overcome the limitations of the ranked based formula.

The ranked based formula yields the following results based on a significance level of 0.05:

Ranked based formula	Difference between	sig	Reject or retain
	Crowd – low knowledge	0.664	retain
	Crowd – high knowledge	0.505	retain
	Crowd – betting website	0.459	retain
	Low knowledge – high knowledge	0.703	retain
	Low knowledge – betting website	0.452	retain
	High knowledge – betting website	0.039	reject

Table 3

The new formula to overcome its limitations yields the following results based on a significance level of 0.05:

New formula	Difference between	sig	Reject or retain
	Crowd – low knowledge	0.975	retain
	Crowd – high knowledge	0.808	retain
	Crowd – betting website	0.634	retain
	Low knowledge – high knowledge	0.800	retain
	Low knowledge – betting website	0.434	retain
	High knowledge – betting website	0.112	retain

Table 4

As the table 3 and 4 above show, while the ranked based formula and new formula show a lot of similarities, there is a significant difference between the groups high knowledge-betting website for the ranked based formula while that difference is not significant for the new formula. Since both formulas do not yield the same results, the new formula will be used to analyze the data.

For hypothesis 1 we want to verify that wisdom of the crowd is present when a crowd is asked to predict the outcome of the 2014 World Cup scores. Before we can answer this, we must determine when wisdom of the crowd is present. Two requirements must be met: the collective error should be significantly lower than the average individual error, and the collective error should be among the lowest 50% of the errors since otherwise asking a random member of the crowd yields a higher than 50% chance to get a more accurate prediction than following the crowd's opinion.

We can conclude that wisdom of the crowd is present when a crowd is asked the predict the outcome of the 2014 World Cup scores, since there is a significant difference between the average individual error and the collective error and the collective error is among the lowest 50% of errors. As the table below shows, for all groups the Wilcoxon sign ranked test has evidence that there is a significant difference between the average individual error and group error based on a significance level of 0.05.

New formula	Difference between	sig	Reject or retain
	Crowd – crowd error	0.000	reject
	High knowledge- high knowledge error	0.002	reject
	Low knowledge – low knowledge error	0.000	reject

Table 5

When we look at the rankings of errors, in table 5 we see that, when we look at the crowd compared to the individual errors, only two individuals have a lower error. When we look at the high and low knowledge groups, only three individuals have a lower error. This means that asking a random individual only yields a lower error than following the crowd 2 out of 119 times ,or a lower error than following the high or low information group3 out of 199 times. We conclude that hypothesis 1 is satisfied.

For hypothesis 2 we want to verify that the average individual error of the high information group is significantly lower than that of the low information group. Two requirements must be met to verify hypothesis 2: the average individual error of the high information group must be lower than that of the low information group, and the difference must be significant.

The average individual error of the high information group is equal to 169.9, and the average individual error of the low information group is equal to 198.2. The Wilcoxon sign ranked test shows that there is a significant difference between the errors of the high information and the low information group based on a significance level of 0.05.

New formula	Difference between	sig	Reject or retain
	High information error- low information error	0.002	reject

Table 6

We conclude that hypothesis 2 is satisfied.

For hypothesis 3 we want to verify that the prediction diversity of the low information group is higher than the prediction diversity of the high information group. First we determine the prediction diversity of the both groups. Since in the new formula the prediction diversity is not completely based on the variance, we subtract the collective error from the average individual errors to determine the prediction diversity. For the low information group the predictive diversity is 60.75, while for the high information group the predictive diversity is 32.88. The Wilcoxon sign ranked test shows there is a significant difference between the predictive diversity of the low information and the high information group based on a significance level of 0.05.

New formula	Difference between	sig	Reject or retain
	Diversity low information- diversity high information	0.010	reject

Table 7

We conclude hypothesis 3 is satisfied.

For hypothesis 4 we want to verify that the predictive value of the crowd does not deviate significantly from the control crowd, to verify it has predictive value. To do this, we compare the betting website to the other groups. The Wilcoxon sign ranked test shows there is no significant difference between the group betting website and the other groups.

New formula	Difference between	sig	Reject or retain
	Crowd – betting website	0.634	retain
	Low knowledge – betting website	0.434	retain
	High knowledge – betting website	0.112	retain

Table 8

We conclude hypothesis 4 is satisfied.

Conclusion

First and foremost, we wanted to test if wisdom of the crowd is present in the crowd of our experiment. Since hypothesis 1 is satisfied, we conclude it is in fact present. This means that, when it comes to predicting the group-stage games of the FIFA World Cup in Brazil, the crowd gives a more accurate prediction than its average individual members. The same holds for the low and high knowledge group. In betting sense on predicting the correct outcome this holds little value, since 'more accurate' is still wrong. However, possibly a more accurate prediction holds value when predicting the amount of goals scored. Further research could look into the possibilities of the value of wisdom of the crowd in betting situations.

We know that a prediction is based on information and error. Hypothesis 2 provided us with evidence that knowledge on this particular subject is equal to more information. The average members of the high information group had significantly more accurate predictions and lower errors than the average members of the low information group. Nevertheless, hypothesis 3 provided us with evidence that the crowd's diversity of the low information group was higher than the crowd's diversity of the high information group, making the predictions of the low information group more diverse, which adds value to its prediction. Overall we saw that, while the average predictions of the high information group were more accurate than of the low information group, the group's opinions did not significantly differ. Showing that, in this particular situation, the value of having extra information was equal to the value of having more diversity. This finding shows that -not only in theory, but also in practice- diversity holds value: as much value to overcome the lack of knowledge between the two groups.

When we look at hypothesis 4 we can conclude two things. First, the betting website outperformed the group. This is expected, since, if the group outperformed the betting website, it suggests that the crowd's opinion could be used to generate profit on gambling. The betting websites could be seen as a crowd of its individual bettors, as the odds represent the opinion of its bettors. The outperformance of the betting website compared to the other groups confirms the fact that betting websites make more accurate predictions. We cannot determine if this value is generated from a higher amount of information or a lower error, since we lack this knowledge. The second conclusion is that the betting website did not significantly outperform the other groups. Showing that the opinion of the crowd holds value compared to the betting website provides evidence that the crowd's opinion holds real-life value.

Discussion and Limitations

As previously mentioned, the Prediction Diversity Theorem has certain limitations. While we tried to overcome these limitations, we made some consensus. First, the predication diversity is not completely based on variance anymore, which might decrease its mathematical value. Secondly, the Prediction Diversity Theorem is used to predict if the crowd is more accurate than its average individual member. Nevertheless,, in betting situations 'more accurate' holds little value when you only get paid if you are correct, and not when you are 'less incorrect'. The more accurate prediction could have value when it comes to predicting the amount of goals scored in matches. Currently the betting websites provide the option and odds to bet on a certain amount of goals, like over/under 1.5 goals or over/under 2.5 goals. In hindsight that information could have had added value, this information was not collected during the research.

One might argue that the results of the World Cup group-stage games are related to each other. Since all teams only play three games, a team that wins the first game might change their tactics for the second game. In other words: a team that wins game one might play game two with the intention not to lose it, while a team that loses game one will play game two with the intention to win in order to go on to the next round. An example of this was Spain, which lost game one versus the Netherlands, and got countered in game two and lost versus Chili. In our experiment all groups had to make prediction on all the games at the same time. If the predictions are done after each match, the results might be more accurate.

As we partly previously explained, the FIFA World Cup is a competition on its own, and without further research we cannot assume that, if wisdom of the crowd occurs for prediction World Cup group-stage games, it also occurs when prediction regular games in regular competitions. The same reasoning applies for the high and low knowledge group. While we provided evidence that showed that a higher amount of knowledge created a lower average individual error, we cannot conclude this will hold for all sports.

At last, but might most interesting, the empirical evidence that the opinion of the high and low information groups did not differ due to higher/lower amount of information and diversity. Further research could be conducted related to the subject. Questions arise like: does increasing the amount of information always has a trade-off with diversity? Is it possible to increase the diversity without decreasing the information?. If groups can be formed were increased diversity does not go at the expense of information or vice versa could provide additional power to wisdom of the crowd. Making crowds more wise, and the topic more interesting.

References

46 procent fan van Ajax, FC Twente populairder dan PSV. (2012, mei 17). *Volkskrant*.

Andrikogiannopoulou, A. P. (2011). Market efficiency and behavioral biases in the sport betting market. *HEC Geneva, Working paper*.

Brian, A. (1994). Inductive Reasoning and bounded rationality. *American economic review*.

Cain, M. L. (2000). the favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy*, 25-36.

Galton, F. (1907). he Wisdom of Crowds. *Nature*, 450-451.

Janis, I. L. (1972). *Victims of Groupthink.* New York: Houghton Mifflin.

Kittur, A. C. (2007). Power of the few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. *Proceedings of Alt*.

Levitt, S. D. (2014). Why are gambeling markets organized so differently from financial markets. *The economical Journal*, 223–246.

Majewski, J. B. (1993). Responding to Relative Decline:The Plank Road Boom of Antebellum New York. *The Journal of Economic History*, 106-122.

Page, L. a. (1998). The PageRank Citation Ranking: Bringing Order to the Web. *Technical Report. Stanford InfoLab*.

Page, S. E. (2007). Making the Difference: Applying a Logic of Diversity. *Academy of Management Perspectives*, 6-20.

Simons, T. P. (1999). Making Use of Difference: Diversity, Debate, and Decision Comprehensiveness in Top Management Teams. *Academy of management journal*, 662-673.

Spann, M. S. (2008). Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets,Betting Odds and Tipsters. *Journal of Forcasting*, 55-72.

Stevenson, A. (2010). *Oxford Dictionary of English.* Oxford, New York.

Surowiecki, J. (2004). *The wisdom of the crowds.* New Jork: Anchor Books.

Tversky, A. K. (1974). Judgment under Uncertainty: Heuristics and Biases. *Sience*, 1124-1131.

Vlastakis, N. D. (2009). How efficient is the European Football Betting market? Evidence from Arbitrage and Trading Strategies. *Journal of Forecasting*, 426-444.